LECTURE 5

Database Resource Management

LECTURE NOTES
CHAPTER 5 from Management Information Systems – O'Brian&Marakas book

FOUNDATION DATA CONCEPTS

Why Study Data Resource Management?

Data is a vital organizational resource, which needs to be managed like other important business assets. Most organizations could not survive or succeed without quality data about their internal operations and external environment. Managers need to practice ***data resource management***.

Definition

This is a managerial activity that applies information systems technology like *database management* and other management tools to the task of managing an organization's data resources to meet the information needs of business users.

Foundation Data Concepts:

A hierarchy of several levels of data has been devised that differentiates between different groupings, or elements, of data. Data are organized in a **data storage hierarchy** of increasingly complex levels: characters, fields, records, files, and databases. See Fig.5.2, p.181.

Data are logically organized into:

1. ***Characters*** - A ***character*** is the most basic logical data element from the user's point of view. It consists of a single alphabetic, numeric, or other symbol. A, B, C, 1, 2, 3, #, $, % are all examples of single characters. The terms bit and byte refer to the physical storage elements provided by the hardware.
2. ***Fields*** - A ***field*** or data item consists of a grouping of characters. **A field** is a collection of characters, i.e. a unit of data. A data field represents an ***attribute*** (a characteristic or quality) of some ***entity*** (object, person, place, or event). Example: student's name, etc.
3. ***Record*** - Related fields of data are grouped to form a ***record.*** A **record** is a collection of related fields. An example of a record would be your first and last name, address and Social Security number. Thus, a record represents a collection of *attributes* that describe an *entity*. *Fixed-length* records contain a fixed number of fixed-length data fields. *Variable-length* records contain a variable number of fields and field lengths.
4. ***File*** - A group of related records is known as a data ***file***, or *table*. A **file** is a collection of related records. An example of a file is data collected on everyone employed in the same department of a company, including all names, addresses, and Social Security numbers. Files are frequently classified by the application for which they are primarily used, such as a *payroll file* or an *inventory file,* or the type of data they contain, such as a *document file* or a *graphical image file*. Files are also classified by their permanence, for example, a *master file* versus a *transaction file.* A transaction file would contain records of all transactions occurring during a period, whereas a *master file* contains all the permanent records. A *history file* is an obsolete transaction or master file retained for backup purposes or for long-term historical storage called *archival storage*.
5. ***Database*** - A database is an integrated collection of logically related records or *data elements*. A **database** is an organized collection of integrated files. A company database might include files on all past and current employees in all departments. A ***database*** consolidates records previously stored in separate files into a common pool of data records that provides data for many applications. The data stored in a database are independent of the application programs using them and of the type of secondary storage devices on which they are stored.
   Example of p.182, fig. 5.3

TYPES OF DATABASES

Continuing developments in information technology and its business applications have resulted in the evolution of several major types of databases. Several major conceptual categories of databases that may be found in computer-using organizations include:

See Fig.5.4, p.183

1. ***Operational Databases -*** These databases store detailed data needed to support the operations of the entire organization. They are also called *subject area databases* (SADB), *transaction databases*, and production databases. Examples are customer databases, personnel databases, inventory databases, and other databases containing data generated by business operations.
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 is done to improve database performance and security.
3. ***External Databases -*** Access to external, privately owned online databases or data banks is available for a fee to end users and organizations from commercial online services, and with or without charge from many sources on the Internet, especially the Web. There are statistical, bibliographic and full text databases.
4. **Hypermedia Databases**– consist of hyperlinked pages of multimedia data, e.g. text, graphics, video clips, etc.- Fig. 5.6, p.186. 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 a home page and other hyperlinked pages of multimedia or mixed media (text, sound, etc.). Thus a hypermedia database consists of Web page content that is described by HTML code or XML labels + image files, video files and audio. The Web server software acts a DBMS to manage the transfer of hypermedia files.

DW AND DM

1. **Data warehouse** - definition: fig. 5.7, p.186
   This is a large database that stores data that have been extracted from the various operational, external, and other databases of an organization.
   A data warehouse stores data from current and previous years that has been extracted from the various operational and management databases of an organization. It becomes a central source of data, which has been screened, edited, standardized, and integrated so it can be used by managers and other end user professionals throughout an organization.
   Data warehouses may be subdivided into ***data marts***, which hold specific subsets of data from the warehouse that focus on specific aspects of a company, such as a department or a business process -
2. ***Data Mining*** - A major use of data warehouse databases is ***data mining***. In data mining, the data in a data warehouse are processed to identify key factors and trends in historical patterns of business activity that can be used to help managers make decisions about strategic changes in business operations to gain competitive advantages in the marketplace. Fig. 5.9 – DM extracts business knowledge from the stored in different locations organizational data.

Data Mining Uses

- Perform “market-basket analysis” to identify new product bundles.
- Find root causes to quality or manufacturing problems.
- Prevent customer attrition and acquire new customers.
- Cross-sell to existing customers.
- Profile customers with more accuracy.

APPROACHES FOR DATA MANAGEMENT

There are two main approaches for data management

1. Traditional file processing
2. Database management

Characteristics of traditional file processing

A ***file -processing system*** is a collection of application programs that perform services for the end-user such as production of reports. Each program defines and manages its own data.

File-based systems were developed in response to the needs for more efficient data access. They were an early attempt to computerize the manual filling systems.

Characteristics of the traditional approach – see Fig.5.10, p.190:

- Any application deals with own files.
- Any user writes own programs to process data.
- The application program comprises data definitions.

Advantage: an easy to apply approach.

One more example: consider the selling and renting of properties.

1. The Sales Department deals with Property, Owner, and Renter objects. It is responsible for selling and renting of properties. Whenever a client approaches the Sales Department to marketing his or her property for rent, a form that gives the details of the property is completed.
2. The Contracts Department deals with Lease, Property, and Renter objects. It is responsible for handling the lease agreements associated with properties for rent. Whenever a client agrees to rent a property, a proper form giving the renter and property details is filled.

With the assistance of the Data Processing Department two information systems to handle the renting of property and the lease agreements are developed. Any information system consists of different files that store the necessary data.

Obviously data are similar for both applications. It can be seen quite clearly that there is a significant amount of duplication of data in these departments. This is generally true for all file-based systems.

The disadvantages of the file-based approach are:

1. Uncontrolled Duplication of Data
   The file-based approach leads to uncontrolled duplication of data that is called ***data redundancy***. Data redundancy means to duplicate data across different files.
2. Violated Data Integrity due to Data Redundancy
   Data integrity is the possibility at any time to deals with accurate and consistent data nevertheless their file allocation.
   The data are no longer consistent. Because of duplication:
   - updates may be carried out over some files only;
   - it is also possible that data are entered incorrectly.
3. Program - Data Dependence
   The physical structure and storage of the data files and records are defined in the application code. So, changes are difficult, very time consuming and subject to errors.
4. Separation and Isolation of Data
   When the data are placed in different files, it is more difficult to extract data and to produce reports. Users themselves handle relationships among data scattered across different files.
5. Difficult multi-users access to the same data
6. Incompatible File Formats

The structures of files are dependent on the application programming language. The direct incompatibility of files generated by different programming languages makes them difficult to process jointly.

7. Fixed Queries
   File-based systems are dependent upon the application programmer. Any query or report that is required has to be written by a programmer. So, the type of query or report that could be produced is usually fixed.

Characteristics of database management approach

The development of databases and database management software is the foundation of modern methods of managing organizational data.

Database Management Approach - is a method whereby data records and objects are consolidated into a single repository called a database.

Therefore:

- The database can be accessed by many different application programs.
- There is a software interface between users and databases
- Data definition is stored once, separately from application programs

See Fig.5.11, p.190

Database Management System - (DBMS) serves as a software interface between users and databases. Thus, database management involves the use of database management software to control how databases are created, interrogated, and maintained to provide information needed by end users and their organizations.

The database management approach involves three basic activities:

- Updating and maintaining common databases to reflect new business transactions and other events requiring changes to an organization's records.
- Providing information needed for each end user's application by using application programs that share the data in common databases.
- Providing an inquiry/response and reporting capability through a DBMS package so that end users can easily interrogate databases, generate reports, and receive quick responses to their ad hoc requests for information.

DBMS – this is software that controls the creation, maintenance, and use of databases of an organization and its end users.

Examples: Oracle, DB2, MySQL, etc

Common DBMS components and functions

The three main functions of a DBMS are:

1. Create new databases and database applications
2. Maintain the quality of data
3. Use the databases to provide the necessary information to the end users

The four major uses of a DBMS include:

1. Database Development
2. Database Interrogation/ Querying
3. Database Maintenance
4. Application Development

***Database Interrogation*:**

The ***database interrogation*** capability is a major benefit of a database management system. End users can use a DBMS by asking for information from a database using a ***query language*** or a ***report generator***.

Features of a query language:

- Users receive an immediate response in the form of video displays or printed reports.
- No difficult programming is required.
- Users can obtain immediate responses to ad hoc data requests.

Features of a report generator:

- Users receive an immediate response in the form of video displays or printed reports.
- No difficult programming is required.
- Users can specify a report format for information they want presented as a report.

SQL, or ***Structured Query Language***, is a query language found in many database management packages. It is used to obtaining immediate responses to ad hoc inquiries.

Basic Form of an SQL query is:
SELECT..... FROM.....WHERE
SELECT:.....….. list the data fields you want retrieved
FROM:.....list the fields or tables from which the data must be retrieved
WHERE.....specify conditions that limit the search to only those data records in which you are interested

QBE, or ***Query by Example***, is another form of query language found in some database management packages. The QBE method displays boxes for each of the data fields in one or more files. The end user simply "points-and-clicks" to indicate which information they want

Graphical and Natural Queries - Most end user database management packages offer GUI (graphical user interface) point-and-click methods to query a database. These methods are easy to use and are translated by the software into SQL commands. Other packages are available that use *natural language* query statements similar to conversational English.

Database Maintenance:

Managers need accurate information in order to make effective decisions. The more accurate, relevant, and timely the information, the better-informed management will be when making decisions. Thus, the databases of an organization need to be updated continually to reflect new business transactions and other events. This ***database maintenance*** process is accomplished by transaction processing programs and other end user application packages with the support of the DBMS. So database maintenance involves:

- Updating a database continually to reflect new business transactions and other events
- Updating a database to correct data and ensure accuracy of the data

Application Development:

DBMS packages play a major role in ***application development***. Application development is made easier by *data manipulation language* (DML) statements, which can be included in application programs to let the DBMS perform the necessary data handling activities. Programmers can also use the internal programming language provided by many DBMS packages or a built-in application generator to develop complex application programs. End users, systems analysts, and other application developers can use the 4GL programming language and built-in software development tools provided by many DBMS packages to develop custom application programs.

The database resource management approach provides business managers and professionals with several important benefits such as:

- Reduce the duplication of data
- Integrate data so that multiple programs and users can access them.
- Programs are not dependent on the format of the data and the type of secondary storage hardware being used.
- Users are provided with an inquiry/response and reporting capability that allows them to easily obtain information they need without having to write computer programs.
- Computer programming is simplified, because programs are not dependent on either the logical format of the data or their physical storage location.
- Integrity and security of the data stored in databases can be increased, since access to data and

modification of the database are controlled by database management system software, a data dictionary, and a database administrator function.

The limitations of database management arise from:

- Its increased technological complexity.
- Development of a large database and installing a DBMS can be difficult and expensive.
- More hardware capability is required, since storage requirements for the organization's data, overhead control data, and the DBMS programs are greater.
- Longer processing times may result from high-volume transaction processing applications since an extra layer of software (the DBMS) exists between application programs and the operating system.
- If an organization relies on centralized databases, their vulnerability to errors, fraud, and failures are increased.
- If an organization relies on distributed databases problems of inconsistency of data can arise.

TECHNICAL ISSUES OF DATABASE MANAGEMENT

Database Management:

In all information systems, data resources must be organized and structured in some logical manner so that they can be accessed easily, processed efficiently, retrieved quickly, and managed effectively.

Data structures and access methods ranging from simple to complex have been devised to effectively organize and access data stored by information systems.

Database Structures:

The relationships among the many individual records in databases are based on one of several logical data structures or models.

DBMS are designed to provide end users with quick, easy access to information stored in databases.

There are five fundamental database structures:

1. ***Hierarchical Structure***:

Early mainframe DBMS packages used the ***hierarchical structure***, in which:

- Relationships between records form a hierarchy or treelike structure.
- Records are dependent and arranged in multilevel structures, consisting of one *root* record and any number of subordinate levels.
- Relationships among the records are *one-to-many*, since each data element is related only to one element above it.
- Data element or record at the highest level of the hierarchy is called the root element. Moving progressively downward from the root and along the branches of the tree until the desired record is located can access any data element.

2. ***Network Structure:***

The network structure:

- Can represent more complex logical relationships, and is still used by many mainframe DBMS packages.
- Allows *many-to-many* relationships among records. That is, the network model can access a data element by following one of several paths, because any data element or record can be related to any number of other data elements.

3. Relational Structure:

The relational structure:

- Most popular of the three database structures.
- Used by most microcomputer DBMS packages, as well as many minicomputer and mainframe systems.
- Data elements within the database are stored in the form of simple ***tables***. Tables are related if they contain common fields.

- DBMS packages based on the relational model can link data elements from various tables to provide information to users.

***4. Multidimensional Structure*:**

The multidimensional database model:

- Is a variation of the relational model that uses multidimensional structures to store data and relationships between data.
- A major benefit of multidimensional databases is that they are a compact and easy-to- understand way to visualize and manipulate data elements that have many interrelationships.
- Multidimensional databases have become the most popular database structure for the analytical databases that support *online analytical processing* (OLAP) applications.

5. *Object-Oriented Structure*

The object-oriented structure:

- Is considered to be one of the key technologies of a new generation of multimedia web-based applications.
- In an object-oriented structure, an ***object*** consists of data values describing the attributes of an entity plus the operations that can be performed upon the data. This *encapsulation* capability allows the object-oriented model to better handle more complex types of data (graphics, voice, text) than other database structures.
- Supports *inheritance*, that is, new objects can be automatically created by replicating some or all of the characteristics of one or more *parent* objects.
- -Object capabilities and inheritance have made *object-oriented database management* systems (OODBMS) popular in computer-aided design (CAD) applications. Designers can develop product designs, store them as objects in an object-oriented database, and replicate and modify them to create new product designs.

  Multimedia web-based applications for the Internet and corporate Intranets and extranets have become a major application area for object technology.

Evaluation of Database Structures

- Hierarchical data structure is best for structured, routine types of transaction processing.
- Network data structure is best when many-to-many relationships are needed.
- Relational data structure is best when ad hoc reporting is required.

Database Development:

Database management packages allow end users to develop their own databases. Large organizations with client/server or mainframe-based systems usually place control of enterprise database development with ***database administrators*** (DBAs) and other database specialists. This improves the integrity and security of organizational databases. In database development a *data definition language* (DDL) is used to develop and specify the data contents, relationships, and structures of each database, and to modify these database specifications when necessary. Such information is catalogued and stored in a database of data definitions and specifications called a *data dictionary*, which is maintained by the DBA.

The Data Dictionary - A data dictionary is a computer-based catalogue or directory containing ***metadata****,* that is, data about data. A data dictionary includes a software component to manage a database of data definitions about the structure, data elements, and other characteristics of an organization's databases.

The database administrator uses the data dictionaries to query and report on the status of any aspect of the firm's metadata. The administrator can then make changes to the definitions of selected data elements. Some *active (*versus *passive)* data dictionaries automatically enforce standard data element definitions whenever end users and application programs use a DBMS to access an organization's database.

Developing small personal databases is relatively easy using microcomputer DBMS packages. However, developing a large database can be a complex task. In many companies, developing and managing corporate databases is the primary responsibility of the database administrator and database design analysts. They work with end users and systems analysts to determine:

- What data definitions should be included in the database
- What structure or relationships should exist among the data elements.

So, enterprise-wide database development is usually controlled by database administrators (DBA).
Database development consists of several stages

1. Data planning
2. Requirements specification
3. Conceptual design
4. Logical design
5. Physical design

Database development must start with a top-down ***data planning*** process.
Database administrators work with corporate and end user management according to the following steps:

- Develop an *enterprise model* to define the basic business processes of the enterprise.
- Define the information needs of end users in a business process.
- Identify the key data elements that are needed to perform their specific business activities.
- Develop *entity relationship diagrams (*ERDs*)* that model the relationships among the many entities involved in the business process.

The user views become the basis for the ***data modelling*** steps where the relationships between data elements are identified.
Each data model defines the logical relationships among the data elements needed to support a basic business process.
These data models then serve as logical frameworks (*schemas* and *subschemas*) on which to base the ***physical design*** of databases and the development of application programs to support the business processes of the organization.
Basic concepts:

- ***Data Models -*** Represent a *logical view* of the data and relationships of the data.
- ***Schema -*** Is an overall logical view of the relationships between data in a database.
- ***Subschema -*** Is a logical view of data relationships needed to support specific end user application programs that will access that database.
- ***Physical or Internal View -*** Looks at how data is physically arranged, stored, and accessed on the magnetic disks and other secondary storage devices of a computer system.

SUMMARY

1. Data resource management is a managerial activity that applies information technology and software tools to the task of managing an organization's data resources.
2. The database management approach consolidates data needed by different applications into several common databases and provides an easy-to-use ad hoc reporting capability.
3. Database management systems are software packages that simplify the creation, use, and maintenance of databases.
4. Several types of databases are used by business organizations including operational, distributed, and external databases.
5. Data warehouses are a central source of data from other databases that have been cleaned, transformed, and cataloged for business analysis and decision support applications.
6. Data must be organized in some logical manner on physical storage devices so that they can be efficiently processed. For this reason, data are commonly organized into logical data elements such as characters, fields, records, files and databases.
7. Database structures such as the hierarchical, network, relational, and object-oriented models are used to organize the relationships among the data records stored in databases.
8. The development of databases can be easily accomplished using microcomputer database management packages for small end-user applications