

Managing Data Resources

Organization Obstacles to a Database Environment

- Implementing a database requires widespread organizational change in
 - ◆ The role of information (and information manager)
 - ◆ The allocation of power at senior levels
 - ◆ The ownership and sharing of information
 - ◆ Patterns of organizational agreement
- A **Database Management System (DBMS)** challenges the existing power arrangements in an organization and for that reason often generates political resistance.

Cost-Benefits Considerations

- The costs of moving to a database environment are tangible, up-front and large in short term.
- It soon becomes apparent to senior management that a database system is a huge investment.
- The benefits of the DBMS are often intangible, back loaded and long term.

Traditional File Environment

- An effective Information System provides users with timely, accurate and relevant information. This information is stored in computer files.
- When the files are properly arranged and maintained, users can easily access and retrieve the information they need.

File Organization Terms and Concepts

- **Bit** – Smallest unit of data a computer can handle
- **Byte** – A single character
- **Field** – A grouping of words or a complete number such as person name or age
- **Record** – A group of related fields
- **File** – A group of records of the same type
- **Database** – A group of related files

File Organization Terms and Concepts

- **Entity** – A person, place, thing or event about which information must be kept
- **Attribute** – A piece of information describing a particular entity
- **Key Field** – A field in a record that uniquely identifies instance of that records so that it can be retrieved, updated or sorted

File Organization and Access Method

- File Organization
 - ◆ Sequential File Organization
 - ◆ Direct or Random File Organization
- Access Method
 - ◆ Indexed Sequential Access Method (ISAM)
 - ◆ Direct Access Method

Sequential File Organization

- A method of storing data records in which the records must be retrieved in the same physical sequence in which they are stored.

Direct or Random File Organization

- A method of storing data records in a file so that they can be accessed in any sequence without regard to their physical order on storage media.

Indexed Sequential Access Method (ISAM)

- A file access method to directly access records organized sequentially using an index of key fields.
- Index: a table or list that relates record keys to physical locations on direct access files.

Direct Access Method

- A method of accessing records by mathematically transforming the key fields into the specific addresses for the records.
- Transform algorithm: a mathematical formula used to translate a record's key field directly into the record's physical storage location.

Traditional File Environment

- Traditional file environment (Flat File Organization, Data File Approach) is a way of collecting and maintaining data in an organization that leads to each functional area or division creating and maintaining its own data files and programs.

Problems with Traditional File Environment

- Data Redundancy
- Program-Data Dependence
- Lack of Flexibility
- Poor security
- Lack of data-sharing and availability

Data Redundancy

- The presence of duplicate data in multiple files.

Program-Data Dependence

- The tight relationship between the data stored in files and the software programs that update and maintain those files.
- Any change in data organization or format requires a change in all the programs associated with the files.

Lack of Flexibility

- A traditional file system can deliver routine scheduled reports after extensive programming efforts, but it cannot deliver ad-hoc reports and respond to unanticipated information requirements in a timely fashion.

Poor Security

- Because of little control or management of data, access to and dissemination of information are virtually out of control.

Lack of Data-Sharing and Availability

- The lack of control over access to data in this confused environment does not make it easy for people to obtain information.
- Because pieces of information in different files and different parts of the organization cannot be related to one another, it is virtually impossible for information to be shared or accessed in a timely manner.

The Database Environment

- A database is a collection of data organized to serve many applications efficiently by centralizing the data and minimizing redundant data at the same by storing and managing data so that they appear to be in one location.

Database Management System (DBMS)

- Creates and maintains databases
- Eliminates requirement for data definition statements in program
- Acts as interface between application programs and physical data files
- Separates logical and physical views of data

DBMS Components

- Data Definition Language (DDL)
 - ◆ a formal language used by programmer to specify the content and structure of the database.
- Data Manipulating Language (DML)
 - ◆ a language (e.g. SQL) associated with a database management system that is employed by an end-users and programmers to manipulate data in the database.
- Data Dictionary (DD)
 - ◆ an automated or manual tool for storing and organizing information about the data (such as usage, physical representation, ownership, authorization, groupings) maintained in the database.

Structured Query Language (SQL)

- The standard data manipulating language for relational database management systems.

Logical and Physical View of Data

- Logical View
 - ◆ The representation of data as they would appear to an application programmer or end user.
- Physical View
 - ◆ The representation of data as they would actually organized on physical storage media.

Principal Logical Database Model

- Hierarchical Data Model
- Network Data Model
- Relational Data Model

Hierarchical Data Model

- Hierarchical data model represents data in a tree-like structure. A parent record (segment) can have more than one child, but a child can have only one parent (one-to-many relationship).
 - ◆ Example: IBM's IMS
- Pointers are used to link the segments together.

Network Data Model

- The Network data model is a variation of the hierarchical database model. It represents data logically as many-to-many relationship. (E.g. student course relationship)

Relational Data Model

- Relational data model treat data as if they were stored in tables called relations. It can relates data stored in one table to data in another as long as two tables share a common data element.
 - ◆ Examples: DB2, Oracle, MS SQL Server
- Three Basic Operations in a Relational Database
 - ◆ **Select:** Creates subset of rows that meet specific criteria
 - ◆ **Join:** Combines relational tables to provide users with more information
 - ◆ **Project:** Creates subset of columns in a table

Advantages and Disadvantages of the Three Database Models

- The major advantage of the hierarchical and network database models is processing efficiency.
- Hierarchical and network models have several disadvantages: all the access paths, directories and indices must be specified in advance. Once specified they are not easily changed without a major programming effort – less flexibility.
- Hierarchical and network database models are programming intensive, time consuming.

Advantages and Disadvantages of the Three Database Models

- The strength of RDBMS are great flexibility in regard to ad hoc queries, power to combine different information from different sources, simplicity of design and maintenance, and the ability to add new data and records without disturbing existing programs and applications.
- RDBMS are relatively slower because they typically require many accesses to the data stored on disk to carry out the select, join and project commands.
- Large relational database may be designed to have some data redundancy to make retrieval more efficient. The same data element may be stored in multiple tables.

Creating a Database

- To create a database, one must go through
 - ◆ Conceptual or Logical Design
 - ◆ Physical Design

Conceptual or Logical Design

- An abstract model of database from a business perspective.
- Requires a detailed description of the business information needs of actual end users of the database.
- **Entity-Relationship Diagram:** Methodology for documenting databases illustrating relationships between database entities.
- **Normalization:** Process of creating small stable data structures from complex groups of data

Physical Design

- Shows how the database is accurate arranged on direct access storage device
- Performed by database specialists

Database Trends

- Recent database trends include the growth of distributed databases and the emergence of object-oriented databases, hypermedia databases, multidimensional data analysis, data warehousing and linking databases to the web.

Centralized Database

- Used by single central processor or multiple processors in Client/Server network

Distributed Database

- Stored in more than one physical location
- Partitioned database
 - ◆ Each remote processor has the necessary data to serve its own local needs
- Duplicated database
 - ◆ Data duplicated in all remote locations

Object-Oriented Databases

- Stores data and procedures as objects that can be retrieved and shared automatically.

Hypermedia Database

- Organizes data as network of nodes linked in pattern specified by user.
- User can choose his/her own path to move from node to node.
- Each node can be displayed on a screen which contains text, graphic, sound, video or executable programs and links to other nodes
- E.g. Web sites

Multidimensional Data Analysis / On-line Analytical Processing (OLAP)

- Sometimes managers need to analyze data in ways that cannot be represented by traditional database models. Such analysis requires multidimensional view of data.
- To provide this type of information, organization can use either a specialized multidimensional database or a tool that creates multidimensional views of data in relational databases.

Data Warehouse

- A data warehouse is a database with reporting and query tools that stores current and historical data extracted from operational systems and consolidated for management reporting and analysis.
- The data originate from many core operational system and external sources and are copied into the data warehouse database as often as needed – hourly, daily, weekly or monthly.
- The data are standardized and consolidated so that they can be used across the enterprise for management analysis and decision making.
- It includes a range of ad hoc and standardized query tools, analytical tools and graphical reporting facilities.

Benefits of Data Warehouse

- Improved and easy accessibility to information.
- Ability to model and remodel the data.
- Access data as often as need without affecting the performance of the operational systems.

Data Mart

- A data mart is a small data warehouse containing only a portion of the organizational data for a specified function or population of users.
- It contains summarized or highly focused portion of data for a specified function or group of users.

Data Mining

- Tools for analyzing large pools of data.
- Find hidden patterns and infer rules to predict trends.

Linking Databases to the Web

- Advantages using Web to access organization's internal database
 - ◆ Web browser is easy to use.
 - ◆ No change to Web interface required.
 - ◆ Cost less to add a Web interface in front of a legacy system than to redesign and rebuild a system.

Management Requirement for Database System

- Without management support and understanding, database efforts fail.
- The critical elements in a database environment are:
 - ◆ Data Administration
 - ◆ Data Planning and Modeling Methodology
 - ◆ Database Technology and Management
 - ◆ Users

Data Administration

- Data administration is a special organizational function for managing the organizational data, resources, concerned with information policy, data planning, maintenance of data dictionaries, and data quality standards.
 - ◆ Information policy specifies its rules for sharing, disseminating, acquiring, standardizing, classifying and inventorying information throughout the organization. It lays out specific procedures and accountabilities, specifying which organizational units share information, where information can be distributed, and who has the responsibility for updating and maintaining the information.

Data Planning and Modeling Methodology

- Organization requires enterprise-wide planning for data.
- Enterprise analysis, which addresses the information requirements of the entire organization, is needed to develop databases.

Database Technology & Management

- Database Technology allows firms to maintain large database with detailed personal information that pose a threat to individual privacy.
- Database administration
 - ◆ Defines and organizes database structure and contents
 - ◆ Develops security procedures to safeguard the database
 - ◆ Develop database documentation
 - ◆ Maintain the database management software

Users

- To optimize access for non-specialists, more resources must be devoted to training end users.
- Professional systems workers must be retrained in the DBMS language, DBMS application development procedures, and new software practices.