Data Base

1. Data and Information

1.1. Data

Data is a collection of facts and figures related to an object. An object can be a person, event or anything about which data is collected. Data may consist of text, numbers, images, sound and videos. Data can be processed to produce useful information.

Data is very important for an organization. It can be used by the managers to perform effective and successful operations of management. It provides a view of past activities related to the rise and fall of an organization. It also enables the organization to make better decision for future activities.

Example

The data of a student may consist of Roll Number, Student Name, Father Name and Marks of different subjects. The purpose of collecting this data is to maintain the record of the student during the study period.

1.2. Information

The processed data is called **information**. Information is meaningful, useful and organized. It is used for making decisions. Data is used as input for processing and information is the output of this processing.

Example

The data of the students can be processed to produce useful information such as total marks and grade of the student. It can also be processed to find the number of passed and failed students etc.

---

2. File, Record and Field

2.1. Field

A combination of one or more characters is called field. It is the smallest unit of data that can be accessed by the user. The name of each field in a record is unique. The data type of a field indicates

the type of data that can be stored in the field. Each field contains one specific piece of information. A field size defines the maximum number of characters that can be stored in a field.

The EmployeeID, Name, HireDate, JobTitle and Phone are examples of fields that represent an employee.

2.2. Record

A collection of related fields used as a single unit is called a record. For example, an employee's record includes a set of fields that contains EmployerID, Name, HireDate, JobTitle and Phone etc.

2.3. File

A collection of related records used as a single unit is called file. File is also known as data set. Files are stored on different storage media such as hard disk, USB flash drive or optical disc etc. The Employee file may contain the records of hundreds of employees. Each employee's record consists of same fields but each field contains different data.

---

3. File Types

3.1. File types from usage point of view (MTB)

The types of files from usage point of view are as follows:

3.1.1. Master File

Master files are used to store the information that remains constant for a long period of time. For example, a college maintains a master file of all students. The file may contain the fields such Student Name, Father Name, Address, Phone and Email etc. The master files are the latest updated files. These files are updated when any change in their contents are required. These files are never empty since they are created.

3.1.2. Transaction File

A type of file that stores the input data before processing is called transaction file. It may be temporary file and may exist until the master file is updated. It may also be used to maintain a permanent record of the data about a transaction. For example, a transaction file can be used to store the fee deposited by the student.

3.1.3 Backup File

A type of file that is used to take backup of important data is called backup file. It is a permanent file. It is used to store an additional copy of data. The data can be recovered from this file if the original file is lost or damaged. Backup files are mostly created by using specific software.

3.2. File types from functional point of view (PD)

The files are given proper names from functional point of view. It consists of file name and file extension. The name and extension of a file is separated by dot. The extension is normally assigned by the software in which the file is created.

The types of files from functional point of view are as follows:

3.2.1 Program File

A type of file that contains the software instructions is called program file. A program file has an extension of .com or .exe. These files contain the instructions that can be directly executed by the computer.

3.2.2. Data File

A type of file that contains data is called data file. Data files are created by the software being used. Different software store data in the data files using different formats. A data files is generally opened in the same software in which it is created. It can also be opened in different software that supports the format of that data file.

Some examples of data files and software in which they are created are as follows:

Software	File Type	File Extension
Notepad	Text File	.txt
Word Processor	Document	.doc,.rtf
Spreadsheet	Worksheet	.xls,.wks
Database	Data File	.dat,.dbf,.mdb
Image Processor	Image File	.if,.jpg,.eps,.git,.bmp
Audio Software	Audio File	.wav,.mid
Video Software	Video File	.avi,.mpg

3.3. Files types on storage media (SDI)

A technique for physically arranging the records of a file on secondary storage devices is called file organization. The files are organized on storage media in the following ways:

3.3.1. Sequential Files

The records in sequential file organization are stored in sequence. A sequence means the records are stored one after the other. The records can be retrieved only in the sequence in which they were stored. The primary storage media for sequential files is magnetic tape.

The major disadvantage of sequential access is that it is very slow. It requires more processing time. If the last record is to be retrieved, all preceding records are read before reaching the last record.

3.3.2. Direct or Random Files

The records in direct file organization are not stored in a particular sequence. A key value of a record is used to determine the location to store the record. Each record is accessed directly without going through the preceding records.

This file organization is suitable for storing data on disk. Direct file organization is much faster than sequential file organization for finding a specific record. A problem that may occur in this type of files is called synonym. The problem occurs if the same address is calculated to store two or more records.

3.3.3. Indexed Sequential Files

In indexed sequential file organization, records are stored in ascending or descending order. The order is based on a value called key. Additionally, indexed file organization maintains an index in a file.

An index consists of key values and the corresponding disk address for each record in the file. Index refers to the place on a disk where a record is stored. The index file is updated whenever a record is added or deleted from the file.

The records in indexed file organization can be accessed in sequential access as well as random access or direct access. The records in this file type require more space on storage media. This method is slower than direct file organization as it requires to perform an index search.

---

4. File Processing System

The file processing system was used by different organizations to store and manage data. The typical file processing system uses different set of files for each department in an organization. The records in one file may not be related to the records stored in other files.

4.1. Problems in Traditional File Approach (RIDFIS)

The business organizations faced the following problems in traditional file approach:

4.1.1. Data Redundancy

Data redundancy means duplication of data in multiple files. The redundancy of data causes wastage of storage. In file processing system, the same data may be duplicated in several files.

Suppose that two files are used in a college. The Students file contains the data such as RollNo, Name, Address, Phone and other details of the students. The Library file contains the same data of the students who borrow books from library along with the information about the book. The data of one student appears in two files.

4.1.2. Data Inconsistency

Data redundancy in file system results in data inconsistency problem. Inconsistency means that two files may contain different data of the same entity. For example, the address of a student must be updated in all files if any change occurs. It is possible that it is changed in Students file but not in Library file. The data becomes inconsistent in this situation.

4.1.3. Program Data Dependency

Program data dependency is a relationship between data in files and program required to update and maintain the files. Application programs are developed according to a particular file format in file processing system. If the format of file is changed, the application program also needs to be changed accordingly. For example, if there is a change in the length of postal code, it requires change in the program. The changes may be costly to implement.

4.1.4. Lack of Flexibility

The file process in Page 7 of 26 yields $\text{for } e = \frac{\text{in}}{\text{out}}$ for time to collect the data from different files and write programs to produce the desired information.

4.1.5. Integrity Problems

Integrity means reliability and accuracy of data. The stored data must satisfy certain types of consistency constraints. For example, RollNo and Marks of the students should be numeric value. It

is very difficult to apply these constraints on files in file processing system.

4.1.5. Security Problems

File processing system does not provide adequate security on data. In some situations, it is required to provide different types of access to data for different users. For example, a data entry operator should only be allowed to enter data. The chairman of the organization should be able to access or delete the data completely. Such types of security options are not available in file processing system.

---

5. Data Base

5.1. Database

A database is a collection of logically related data sets or files. Each file may contain different type of information and are used for specific purposes. The files may be organized in different ways to meet different processing and retrieval requirements of the users.

Example
A bank may have separate files for the clients as follows:

- Saving Accounts
- Current Accounts
- Automobile Loan
- Personal Loan
- Clients Information etc.

The client database of bank will consist of the records from all of the above files. The data of any client can be added, retrieved or updated easily using database programs.

5.2. Facilities of Database System

A database system normally provides the following facilities to the user:

- Adding new files to database
- Inserting new data in existing files
- Retrieving data from existing files
- Updating data in existing files
- Deleting data from existing files
- Removing existing files from database

---

6. Database system and components of database systems

6.1. Database System

A database system is a collection of data as well as programs required to manage that data. A database system is a computerized record-keeping system. The main purpose of this system is to maintain data and provide it to the user when it is required.

6.2. Components of Database Systems (DHSP)

The four major components of database system are as follows:

6.2.1. Data

Data is the most important component of database system. Data is a collection of facts stored in the database. The basic purpose of a database system is to store, maintain and process data for the user.

6.2.2. Hardware

The physical components of a computer system are called hardware. The hardware is used to perform different tasks such as input, output, storage and processing.

Some important hardware components are as follows:

- Secondary Storage
- I/O Devices
- Processors
- Main Memory

6.2.3. Software

Software is a collection of programs used by computer within a database system. The most important software is DBMS itself. It uses three types of software to enable the database system work properly. Different types of software used in a database system are as follows:

6.2.3.1. Operating System

It manages all hardware components. It also enables all other software to run on the computer.

6.2.3.2. DBMS Software

It is the database management system that is used to create and manage a database system.

6.2.3.3. Application Programs

These are used to access and process the data stored in the database.

6.2.4. Personnel

The people related to the database system are called personnel. Different types of persons in a database system are as follows:

6.2.4.1. Database Administrator

The database administrator is a person who is responsible to manage the whole database system.

6.2.4.2. Application Programmer

The application programmer is a person who is writes the application programs to access data from database.

6.2.4.3. End Users

The end users are the persons who use application programs to perform different operations on database. They include clerks, managers and directors etc.

---

7. Objectives of the databases (III)

The main objectives of using databases are as follows:

7.1. Data Integration

The data in file system is stored in separate files. It is very difficult to access data stored in separate and independent files. An important objective of databases is to solve this problem. The data in database may be located at different computers physically but it is connected through data communication links. In this way, data appears centralized logically.

7.2. Data Integrity

Data integrity means the reliability and accuracy of data. Integrity rules are designed to keep the data consistent and correct. These rules act like a check on the incoming data. Enforcing data integrity ensures the quality of data in the database. D3MS provides several methods to enforce integrity of the data in a database. For example, an integrity rule can make sure that the same Employee ID is not assigned to multiple employees.

7.3. Data Independence

Database approach provides the facility of data independence. It means that the data and the application programs are separate from each other. The user can change data storage structures and operations without changing the application programs. The user can also modify programs without reorganization of data.

---

8. Database models

A set of rules and standards that define how the database organizes data is called database model. It also defines how users view the organization of data.

8.1. Types of Database Models (HNR)

There are three types of database models. These models are as follows:

8.1.1. Hierarchical Model

The hierarchical model arranges records in hierarchy like an organizational chart. Each record type in this model is called a node or segment. A node represents a particular entity. The top-most node is called root. Each node is a subordinate of the node that is at the next higher level. A higher level node is called parent and lower level node is called child. A parent node can have one or many child nodes. A child node can have only one parent node. This kind of structure is often called inverted tree.

8.1.2. Network Model

The network model is similar to hierarchical model. The difference is that child node can have more than one parent nodes. The child nodes are represented by arrows in network model. It requires more complex diagram to represent a database. It also provides more flexibility than hierarchical model.

8.1.3. Relational Model

Relational model is the most commonly used database model. It is more flexible than hierarchical and network database models. The relational model consists of simple relations. A relation is a term used for table. A relation represents a particular entity. It is used to store information about the entity. The relationships are based on the data of the entities. The relationship between entities is represented by the following diagram:

---

9. Database Management System (DBMS)

9.1. Database Management System

A database management system (DBMS) is a collection of programs that are used to create, maintain and access database in a convenient and efficient manner. DBMS uses database manager software to control the overall structure of a database. DBMS is normally used by different organizations for different purposes.

9.2. Objectives of Database Management System (SAEI)

Some important objectives of database management system are as follows:

9.2.1. Shareability

An ability to share data is a fundamental objective of database management system.
Shareability means that data is shared by different people at the same time. Data is stored at a central place. Different users can access this data from different locations. It reduces storage cost and provides data consistency.

9.2.2. Availability

Availability means that the users must be able to access data and DBMS easily. The data should be available when and where it is required.

9.2.3. Evolvability

Evolvability means that DBMS should provide the facility to charge the database due to increase in user requirements or change in the technology. The charge in database may occur in two ways. It may change by contents or by structure. For example, the structure of the database may be changed if the operation of an organization is expanded.

9.2.4. Integrity

Database integrity ensures that data entered in the database is accurate and consistent. Database is normally shared among different users. Some measures must be maintained to ensure database integrity.

9.3. Features of database management system (DUQRSB)

Some important features of database management system are as follows:

9.3.1. Data Dictionary

Data dictionary is a file that stores all data definitions for a database. It may also monitor the data that is entered. It ensures that data is according to the data definition rules. The rules include field names, field sizes and data types etc. Data dictionary is also used for data access authorization for database users. Data dictionary is also called repository.

9.3.2. Utilities

DBMS utilities are programs that are used to maintain database by processing data, records and files. Some utilities are used for backup and recovery procedures of databases.

9.3.3. Query Language

A query language is used to perform different operations on the databases. The most popular query language is SQL. It stands for Structure Query Language. The statements of SQL are written in simple English-like sentences. Some important commands of SQL are SELECT, DELETE, CREATE, MODIFY, UPDATE and INSERT etc. SQL can be used for the following purposes:

- Create table structures
- Enter data in tables
- Retrieve data from tables
- Update data in tables
- Delete data from tables etc.

9.3.4. Report Generator

A report generator is a program that is used to produce reports. It retrieves data from database and displays it in different formats. The user can use report generator to format page number, dates, titles and column headings etc. The users can produce useful and attractive reports by using report generator. Report generator is also called report writer.

9.3.5. Access Security

Access security refers to the protection of database from unauthorized access. The database management system provides several procedures to maintain data security. The security is maintained by allowing access to the database through the use of passwords. Different users of database system have different levels of access rights to the database.

9.3.6. Backup and Recovery

Database management system provides the facility for backup and recovery. Backup facility is used to store an additional copy of data. The data can be recovered from backup file if original file is lost or damaged.

9.4. Disadvantages of DataBase

Some important disadvantages of database management system are as follows:

9.4.1. High Cost of DBMS

A complete database management system is very large and complex software. It is expensive to purchase database management software.

9.4.2. Higher Hardware Cost

Database management system is complicated and heavy software. It requires a large amount of memory and higher processing power to run. It requires powerful hardware to work properly and efficiently.

9.4.3. Appointing Technical Staff

The technical staff such as database administrator and application programmer is required to manage DBMS. The organization needs to pay good salaries to the technical staff.

9.4.4. Cost of Staff Training

DBMS is a complex system and it requires trained users to use it properly. The user training is required in all fields such as programming, application development and database administration. An organization may need to spend a lot of amount for staff training.

9.4.5. Problems in Wrong Database Environment

The problems may occur if wrong type of database environment is selected. A database system may also need changes due to change in requirements. The change can be costly due to conversion and testing of the existing programs. It may require a lot of cost to implement the changes.

9.4.6. Need of Data Dictionary

Another disadvantage of DBMS is the need of data dictionary. Data dictionary stores data definitions or description of the structure of data used in database. It ensures that data is according to the data definition rules. The rules include field names, field sizes and data types etc. Data dictionary is also used for data access authorization for database users.

Data dictionary is very useful tool but it is also expensive. It requires installation costs as well as hardware requirements.