DATABASICS

Chapter 1

1.1 OVERVIEW

Data: *Data* is a collection of facts, figures and statistics – related to an object, that can be processed to produce a meaningful information. In an organization, data is an asset that enables the managers to perform an effective and successful operation of management. It gives view of past activities (rise and fall) and enables to make better decisions for future. It is also useful for generating some useful reports, graphs, statistics etc.

Information: The manipulated and processed data is called *information* e.g., the percentage of students results. It refers to all the facts, figures or statistics that are precisely meaningful to the people. So by definition, it is an output of a certain process.

Operations: Manipulation of data (after capturing from different sources) to achieve the required objectives and results is called operation. For this purpose, a software (program) is used to process raw data which is converted to meaningful information. Thus, effectively, a series of actions/operations are performed on raw data to achieve some output or result.

These are categorized into three basic activities :

- **Data Capturing:** Data must be recorded or captured in some form before it can be processed. Data may first be recorded on source documents or given directly through input devices.
- **Data Manipulation:** The following operations may then be performed on the gathered data.
  - ❖ Classifying: Organizing data into classes /groups. Items may be assigned predetermined codes, they can be numeric, alphabetic or alphanumeric.
  - ❖ Calculations: Arithmetic manipulation of the data.
  - ❖ Sorting: Data is arranged in logical sequence (numerically or alphabetically).
  - ❖ Summarizing: Masses of data are reduced to a more concise and usable form.

- **Managing the Output Results:** Once the data is captured and manipulated, it may be :
  - ❖ Storing and Retrieval : Data is retained for future reference. Accessing / fetching the stored data and/or information is the *Retrieve* activity.
  - ❖ Communication and Reproduction: Data may be transferred from one location or operation to another, for further processing. It is sometimes necessary to copy or to make duplicate of data. This activity is called *Reproduction.*

1.2 TRADITIONAL FILE SYSTEM

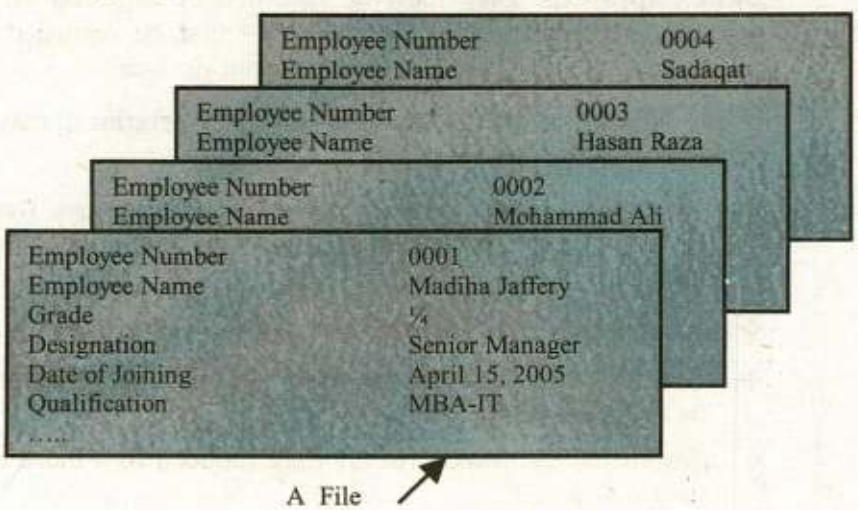
Record: A collection of related fields (facts about something) treated as a single unit is called a record. Let us assume an employee's biographic information in a bank.

Employee Number	0001
Employee Name	Madiha Jaffery
Grade	¼
Designation	Senior Manager
Date of Joining	April 15, 2005
Qualification	MBA-IT
.....	

A "Record"

As it belongs to one employee of the bank, so it is an individual employee's *record* (of biographic information).

File: A collection of related records treated as a single unit is called a file or a data set. If we collect records (as shown below) of all the employees, it becomes a *file* (bio-information) of all the employees of the bank .



Files are categorized according to different criteria as discussed below :

File Types

- Usage point of view

❖ **Master File :** These are the latest updated files which never become empty, ever since they are created. They maintain information that remains constant over a long period of time. Whenever the information changes in files/records, it is updated. Methods of updating are adding, deleting or editing records in a file.

❖ **Transaction File:** These are those files in which data prior to the stage of processing is recorded. It may be temporary file, retained till the master file is updated. It may also be used to maintain a permanent record of transaction data.

❖ **Backup File:** These are again permanent files and their purpose is the protection of vital files/data of an organization by creating them using some specific software utilities.

- Functional point of view

For this purpose, the files are given appropriate names, consisting of two parts i.e., *file name* and *file extension*, having a dot "." in between. Normally, the extension is given by the software being used at the time of initial *save*. These files are summarized as under:

❖ **Program files :** These files contain the software instructions i.e. source program files and executable files. The source program files may have the extension as *.com* and the executable files as *.exe*.

❖ **Data files :** These are the files that contain data and are created by the software being used. A few of them are given as under :

Software	File Types	
Word Processor	**.doc, .rtf**	(document)
Spread Sheet	**.xls** and **.wks**	(worksheet)
Data base	**.dat, .dbf** and **.mdb**	(data files)

- Some other types are as given below:

ASCII/Text files	.....	**.txt**
Image files	.....	**.tif , .jpg, .eps, .gif, .bmp**
Audio files	.....	**.wav, .mid**
Video files	.....	**.avi , .mpg**

File Organization (Storage point of view)

- **Sequential Files:** As the name refers, these files are stored or created on the storage media in the order the records are entered i.e., one after another in the sequence. They require comparatively more processing time.

- **Direct or Random Files:** These files reside on the storage media according to the address which is calculated against the value of the key field of the record. Some times, the same address is calculated, which leads to the concept of synonym.

- **Indexed Sequential:** The key field of the records (in a file)are stored separately along with the address of each record. These files can be processed sequentially as well as randomly. They require relatively more space on the storage media but the processing is as fast as random/direct files.

1.3 DATABASES

A *database* is a collection of logically related data sets or files. Normally, these files/datasets are of different nature, used for specific purposes. These may be organized in various ways to meet various processing and retrieval requirements of the organizations or users. For example;

A bank may have separate files for its clients i.e.,

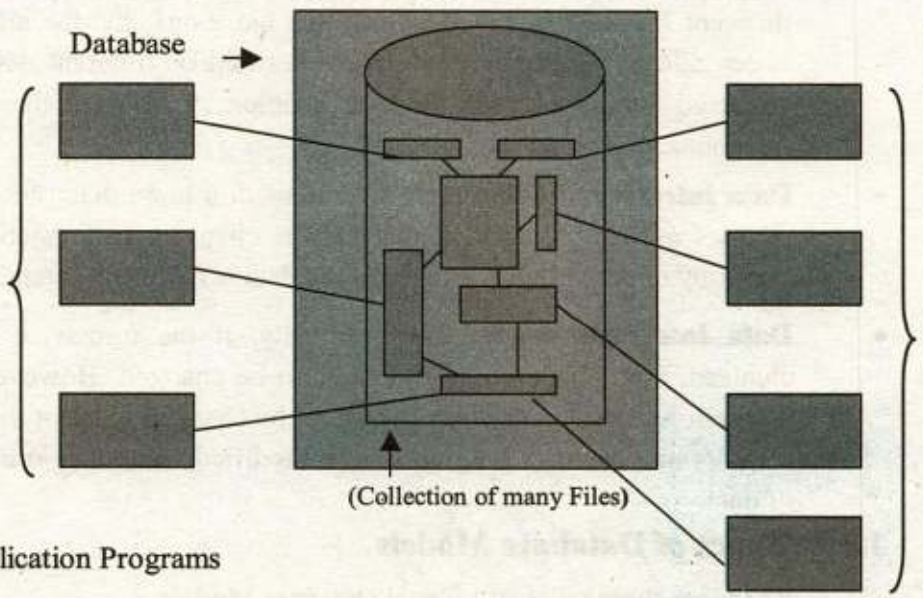
- Savings A/C
- Automobile loan
- Personal loan
- Clients biographic information etc.

The bank's clients/customer database would include records from each of these files. Using a series of programs, data for any client may be added, retrieved or updated depending upon the activity at a particular time.

It is a computerized system whose overall purpose is to maintain information and to make that information available at any time.

A data base system is just a computerized record keeping system. A data base itself can be regarded as a kind of electronic file cabinet, a warehouse or a repository for a collection of computerized data files. The user of the database normally has the following facilities to enjoy.

- Adding new, blank files to the database.
- Inserting new data into the existing files.
- Retrieving data from existing files.
- Updating data in existing files.
- Deleting data from existing files.
- Removing existing files, empty or otherwise from the database.



A Database System

The figure shown above is intended to illustrate the point that a database system involves four major components, namely.

- Data — *The Information*
- Hardware — *The physical* components i.e.,
  - Secondary Storage
  - I/O devices
  - Device controllers
  - I/O channels
  - Processors
  - Main memory

- Software — *User/system* software
  - Set of programs
  - Utilities
- Personnels — *The people*
  - Programmers/Analysts
  - End users
  - Database Administrators

1.3.1 Database Objectives

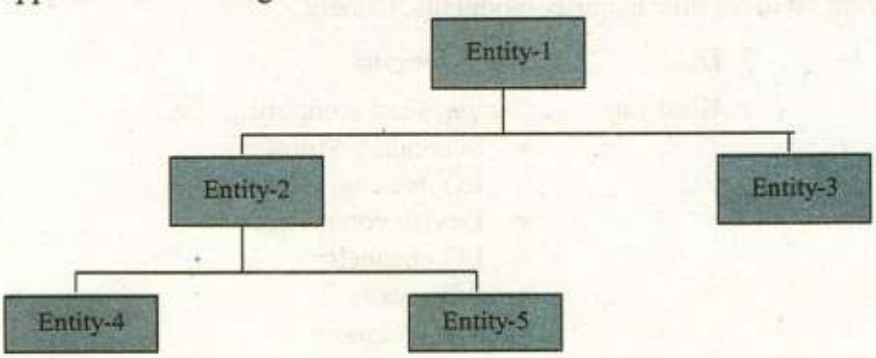
There are at least three main objectives for using the data base organization.

- **Data Integration:** In a database, information is coordinated from different files and operated on a single file. Logically, the information is centralized, physically data may be located on different devices i.e., scattered around over on different locations, connected through data communication links.
- **Data Integrity:** If a data item is contained in more than one file, then all files must be updated if that item is changed. In a database, only one copy of data is kept, therefore, the data is more consistent.
- **Data Interdependence:** Conventionally, if the format of a file is changed, then all the programs have to be changed. However, a data base allows the organization of data to be changed without the need to re-program. It allows programs to be modified without re-organization of data.

1.3.2 Types of Database Models.

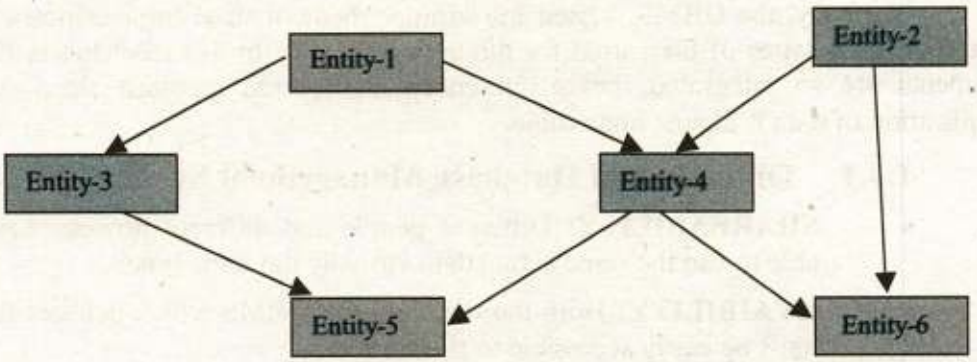
There are three types of Logical Database Models.

- **HIERARCHICAL Model:** This Model has the general shape or appearance of an Organizational Chart.



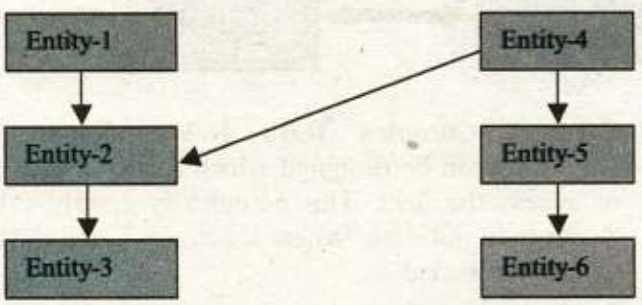
A node on the chart, representing a particular Entity is subordinate at the next highest level, just as on an organizational chart, an employee reports to only boss. This kind of structure is often referred to as an "Inverted Tree", with the top most referred to as the "Root".

- **NETWORK Model:** This is somewhat similar to that of the Hierarchical model but has one major difference.



Subordinate entities, depicted by arrows on the network diagram, may participate in as many subordinate relationships as desired. Therefore a much more complex diagrams may be used to represent the structure of the database. Networks provide more flexibility than a simple hierarchical system in the data relationships may be maintained.

- **RELATIONAL Model:** This system consists of a collection of simple files/Relations (Entities), each of which has no structural or physical connection such as those typically used in hierarchical or network systems.



The various entities possess the interrelationships as depicted by a network like diagram. But these relationships are based on the data content of the entities involved, not by pointer chains or other types of structural connection techniques.

1.4 DATABASE MANAGEMENT SYSTEM

The data management system (a collection of programs) which is used for storing and manipulating databases is called *database management system* (*DBMS*). It is an improvement over the traditional *file management systems*. It uses DBMS software (database manager) which controls the overall structure of a database and access to the data itself.

Normally, the DBMS is used for large or medium sized organizations, having heterogeneous types of files, used for different purposes. In this mechanism, the data elements are so integrated, cross referenced and shared amongst them that the duplication of data is almost impossible.

1.4.1 Objectives of Database Management System

- **SHAREABILITY:** Different people and different processes must be able to use the same actual data virtually the same time.
- **AVAIBILITY:** Both the data and the DBMS which delivers the data must be easily accessible to the users.
- **EVOLVABILITY:** The ability of the DBMS to change in response to growing user needs and advancing technology.
- **DATABASE INTEGRITY:** Since data is shared among multiple users, adequate integrity control measures must be maintained.

1.4.2 Advantages of Database Systems

- **Data Independence:** Application programs are not aware of the physical implementation of the data sets. The DBMS sits in between the application programs and the actual data sets that make up the database.

- **Support Complex Data Relationships:** Fairly complex data structures can be designed which allow various ways to logically view or access the data. This complexity greatly enhances the ability of a designer to put data "where it belongs", and provide a path to that data whenever needed.
- **Sophisticated Data Security Features:** Provide enhanced security mechanisms for access to data. Data base security mechanisms typically go much further in adding more extensive security features. If granted Read access to a file/table, the user may see each record in the file, and every data field it contains. Access intent of each application program (read, write, update, delete) can be specified

explicitly. An application program's view of data records may be controlled to the field level.

- **Data Base Backup / Recovery:** Provide sophisticated backup / recovery mechanism.

Backup / Recovery capabilities often distinguish between true DBMS and a software package that only claims this facility. A DBMS has a logging or recording mechanism that captures information on changes to data within a data base. In case of data base recovery, a utility within the DBMS rebuilds it by using a backup copy of the data and log of changes as input.

- **Advanced Capabilities:** DBMS normally have advance access capability for on-line and ad-hoc reporting capabilities. However, the ability to provide data independence to create complex data structures, to provide security to data access, and to provide backup / recovery capability are the primary requirements of a Database Management Systems.

1.4.3 Disadvantages of Database Systems

Although, the DBMS are very powerful tools to do the job, but they are not more in use. It is because of some disadvantages :

- **Require additional System Overhead:** Additional overhead is required to access data, in case of doing some simple jobs; like reading and processing a tape file, which might take a little time and resources to do the job. If we have to do it on DBMS, it is like "requiring too much to do too little".

- **Additional Training required for Training of Staff:** Application programmers require a sort of precise training to code efficient programs that will run under a DBMS. There is a possibility that inadequate training or experience of application development staff will lead the creation of grossly inefficient database calls. Quite often, the problem might not be found until the program reaches production *status.* The typical example is that of using proper and improper indexes for accessing the database.

- **Problems can multiply in selecting a wrong type of Dbase Environment:** A later change in structure, forced by changing requirements, can be costly in terms of conversion and testing of existing programs. Hierarchical data base systems are, in particular, more sensitive than network or relational systems towards this kind of problem, and implementing changes costs a great deal. On the other hand, doing these changes on relational data bases are fairly easy and less costly.

- **Data must be considered a corporate resource:** The data in a company's data base no longer belong to one organization alone. True, one organization normally has the primary responsibility for creating a data base. However, as data base systems mature, more companies or organizations can share the same data across applications.
- **A Need of a Dictionary:** In order to share data across application systems, or to simply given end users the ability to identify the location of information they need in order to do their jobs, the internal data contents of a company's data bases need to be documented in a consistent manner. For this purpose, they have to install a data dictionary system, which is another overhead on the DBMS.

1.4.4 Features of a DBMS

- **Data Dictionary:** Some databases have a data dictionary, a procedures document or disk file that stores the data definitions or a description of the structure of data used in the database. The data dictionary may monitor the data being entered to make sure it conforms to the data definition rules i.e., file names, field names, field sizes, data types etc. It may be used for data access authorization for the database users.
- **Utilities:** The DBMS utilities are the software programs that are used to maintain the database by manipulating the data, records and files. Some programs are also used for backup and recovery procedures of the databases.
- **Query Language:** Normally, *SQL* (**S**tructured **Q**uery **L**anguage) is used for creating table structures, entering data into them and retrieving/updating the selected records, based on the particular criteria and format indicated, within the databases. Typically, the query is in the form of a sentence or English-like command i.e., SELECT, DELETE, CREATE, MODIFY, UPDATE and INSERT commands.
- **Report Generator:** A report generator is a program that is used to produce an on-screen or printed document from the database. The report format can be specified in advance i.e., row headings, column headings, page headers etc. Even the non-experts can create very useful and attractive reports by using this facility.
- **Access Security:** By using this facility, the database administrators can assign specific access privileges for the users of the databases.
- **Backup and Recovery:** It is an important feature available in almost all the DBMS programs. By using this feature, we are able to have the backup of our data and can later, use it to reinstate it in case of data failure, corruption or loss.

Exercise 1c

1. Fill in the blanks:

(i) DBMS stands for _____

(ii) A _____ is a collection of related fields

(iii) A file is a collection of _____

(iv) Before processing the data is recorded in _____.

(v) A _____is a collection of logically related data

(vi) The data definitions is stored in _____.

(vii) SQL stands for _____

(viii) Hierarchical data Model has the general shape of a(n) _____.

(ix) Data is a collection of _____, _____ and _____

(x) Processed data is called _____

2. Select the correct option:

(i) Which of the following represents a collection of concepts that are used to describe the structure of a database?

a) data warehouse b) data model
c) data structure d) data type

(ii) Which of the following data model is more flexible?

a) Network data model b) Hierarchical data model
c) Relational data model d) object data model

(iii) Which of the following type of file require largest processing time?

a) Sequential file b) Random file
c) Indexed sequential file d) Direct access file

(iv) Which of the following may be a temporary file?

a) Master file b) Transaction file
c) Backup file d) None of these

(v) SQL is a(n):

a) Unstructured language b) Structured language
c) Object oriented language d) Software

3. Write T for true and F for false statement.

   (i) Data can only be processed through computers

   (ii) The traditional file system approach has many advantages over DBMS approach.

   (iii) Data dictionary is used to view the meanings of database terminology

   (iv) Master file is the latest updated file which never becomes empty, ever since it is created.

   (v) SQL is used to retrieve information from the database based on certain criteria.

   (vi) The Network Data Model is more popular and widely used than Relational Data Model.

   (vii) Indexed sequential files can be processed sequentially as well as randomly.

   (viii) Backup files store data prior to its processing.

4. (ix) Microsoft ACCESS is a relational database management system

   (x) A report generator is used to produce a printed document from the database.

5. a) Differentiate between *Data* and *Information* .

   b) What activities are involved in data processing? Discuss in details.

6. Define file, record and field in details?

7. Describe the file types from usage point of view and functional point of view?

8. How do we organize the files on storage media?

9. In general, what activities are to be performed on the databases? Discuss in details.

10. What are the four major components of the database systems? Write in details.

11. Discuss the objectives of the databases in your own words.

12. Describe the different database models?

13. Discuss the objectives and features of the DBMS?

14. What are the advantages and disadvantages of the DBMS?