✔ File Organisation

File Structure : Basic Terminology :

(1) Byte : A byte is an arbitrary set of eight bits that represents a character. It is the smallest addressable unit in today's computers.

(2) Data item : One or more bytes are combined into a data item to describe an attribute of an object. Individual elements of data are called data items.

(3) Records : The data item related to an object are combined into a record.

(4) Record key : To distinguish one specific record from another, systems analysts select one data item in the record that is likely to be unique in all records of a file and uses it for identification purposes. This item, called the record key, key attribute or simply key, is already part of record, no additional data added to it just for the purpose of identification.

(5) Entity : An entity is any person, place, thing or event of interest to the organization about which data are captured, stored or processed. Parents and tests are entities of interests in hospitals, while banking entities includes customers and checks.

(6) File : A file is a collection of related records. Each record in a file is included because it pertains to the same entity. A file of checks, for example, consists only of checks. Inventory records and invoices do not belong to a check file, since they pertain to different entities.

(7) Database : A database is an integrated collection of data stored in different types of records, and in a way that makes them accessible for multiple applications. The inter-relation of the records derives from the relationships in the data, not from their physical storage location.

✔ Types of Files

(a) Master File : A master file is a collection of records about an important aspect of an organization's activities. It may contain data describing the current status of specific events or business indicators.

For example the master file in an accounts payable system shows the balance owned to every vendor or supplier from whom the organization purchases supplies or services. The balance owned to supplier reflects the current status of all accounts.

(b) Transaction File : A tranasction file is a temporary file with two purposes accumulating data about events as they occur and updating master files to reflect the results of current transactions. The term transaction refers to any business event that affects the organization and about which data are captured. Examples of common tranasctions in organizations are making purchases, paying for purchases, hiring personnel, paying employees and recording sales. Data important to an organization are collected about each event and accumulated into the transaction file. As explained previously, a file is a collection of records that are related to each other in the sense that they all pertain to the same entity. Transaction files are used in updating master files. Master files are permanent and exist long as the system exists. However, the contents of the files change as a result of processing and updating.

(c) Table File : A special type of master file is included in many systems to meet special processing requirements involving data that must be referenced repeatedly. Table files contain reference data used in processing transactions, updating master files or producing output. Table files conserve storage space and ease program maintenance by storing in a file data that otherwise it would be included in programs or master file records.

(d) Report File : Report file are temporary files used when printing time is not available for all the reports produced, a situation that frequently arises in overlapped processing. The computer writes the report or document to a file on a magnetic disk or tape, where it remains until it can be printed. This process is known as spooling *i.e.,* output that cannot be printed when it is produced is spooled into a report file.

(e) Other File : Other kind of files as well as special uses of the files types previously discussed, play a vital role in information systems. For *e.g.,* a backup file is a copy of master, transaction, or table file made to ensure that a duplicate is available if any thing happens to the original.

✓ File Access Methods

Various files acccess methods are as follows :

(1) Sequential : It is the simplest method to store and retrieve data from a file. Sequential organization simply means storing and sorting in physical form on tape or disk. In a sequential organization a record can be added only at the end of the file. That is, in a sequential file, records are stored one after the other without concern for the actual value of the data in the records. It is not possible to insert a record in the middle of the file without re-writing the file. In a sequential file update, transaction records are in the same sequence as in the master file. Records from both files are matched, one record at a time, resulting in an updated master file. It is a characteristic of sequential files that all records are stored by position, the first one is at the first position, the second one occupies the second position and third is at third and so on, there are no addressed or location assignments in sequential.

To read a sequential file, the system always starts at the beginning of the file. If the record sought is somewhere in the file, the system reads its ways upto it, one record at a time. For example, if a particular record happens to be the fifteenth one in a file, the system starts from the first one and reads ahead one record at a time until the fifteenth one is reached. It cannot jump directly to the fifteenth one in a sequential file without starting from the beginning.

(2) Random or Direct : For a proposed system, when the sequential files are assumed as a disadvantages, another file organisation called direct organization is used. As with a sequential file, each record in a direct file must contain a key field. However, the records need not appear on the file in key field sequence. In addition any record stored in a direct file can be accessed if its location or address is known. All the previous records needs not to be accessed. The problem, however, is to determine how to store the data records so that, given the key field of the desired record, its storage location on the file can be determined. In other words, if the program knows the record key, it can determine the location address of a record and retrieve it independently of any other records in the file. It would be ideal if the key field could also be the location of the record on the file. This method is known as direct addressing method. This is quite simple method but the requirements of this method often prevents its use. Because of many other factors, this method could not become popular. Hence, it is rarely used.

Therefore, before a direct organized file can be created, a formula or method must be devised to convert the key field value for a record to the address or location of the record on the file. This formula or method is generally known as an algorithm. Otherwise known as hashing addressing.

(3) Indexed : The third way of accessing records in the system is through an index. The basic form of an index includes a record key and the storage address for a record.

To find a record, when the storage address is unknown it is necessary to scan the records. However, if an index is used, the search will be faster since it takes less time to search an index than an entire file of data.

Index file offers the simplicity of sequential file while at the same time offering a capability for direct access. The records must be initially stored on the file in sequential order according to a key field. In addition, as the records are being recorded on the file, one or more indices are established by the system to associate the key field value with the storage location of the record on the file. These indices are then used by the system to allow a record to be directly accessed. To find a specific record when the file is

stored under an indexed organization, the index is searched first to find the key of the record wanted. When it is found, the corresponding storage address is noted and then the program can access the record directly. This method uses a sequential scan of index, followed by direct access to the appropriate record. The index helps to speed up the search compared with a sequential file, but it is slower as compared it to the direct addressing.

✔ Database Design

Database design is the process of producing a detailed data model of a database. The logical data model contains all the needed logical and physical design choices and physical storage parameters that are used to generate a design in a Data Definition Language, which can then be used to create a database. A fully attribute data model contains detailed attributes for each entity.

The term database design can be used to describe many different parts of the design of an overall database system. Principally, and most correctly, it can be thought of as the logical design of the base data structures used to store the data.

The process of database design generally consists of a number of steps which will be carried out by a database designer. Usually the designer must—

(a) Determine the relationships between the different data elements.

(b) Superimpose a logical structure upon data on the basis of their relationships.

In a majority of cases, a person who is doing the design of a database is a person with expertise in the area of database design, rather than expertise in the domain from which the data to be stored is darawn e.g., financial information, biological information etc. Therefore, the data to be stored in the database must be determined incorporation with a person who does have expertise in that domain, and who is aware of what data must be stored with in the system.

✔ Process of Database Design

Step 1	Determine the purchase of database
Step 2	Find and Organize the information required
Step 3	Divide the information into tables
Step 4	Turn information items into columns
Step 5	Specify primary keys
Step 6	Set up the table relationships
Step 7	Define the design
Step 8	Apply the normalization rules