# CONCEPT OF COMPUTER FILES

*A computer file is a collection of related records.*

*DEFINITION OF TERMS*

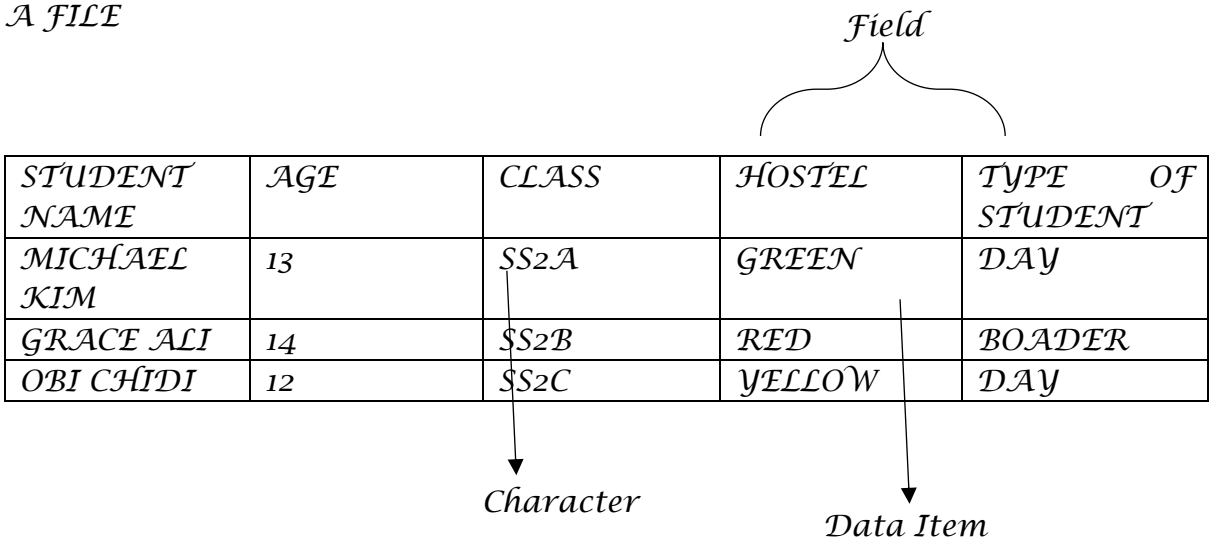
1. *Character: is the smallest element in a file and can be a letter, a number or special character.*
2. *Data item: is the smallest unit of information stored in a computer file.*

   *TYPES OF DATA ITEMS*

   I. ***Numeric** : this type of data consists of numbers 0-9.*
   II. ***Alphabetic:** this type of data consists of letters A-Z.*
   III. ***Alphanumeric or Alphameric:** this type of data is a combination of alphabetic and numeric character, and is used to describe the collection of Latin letters and Arabic digits or a text constructed from this collection. Eg No 40 block 2 kings way, Ikeja , Lagos state.*
3. ***Field :** is a collection of characters (bytes) that make up unit of information.*
4. ***Record:** is a collection of related data items or fields eg student's name, age, class, hostel and type of student.*
5. ***Computer files**: are the most basic unit of data that users can store on a disk, it can be video, song, image, program or document, it is a collection of related records.*
6. ***Database**: is a collection of related files.*

***DIAGRAM***

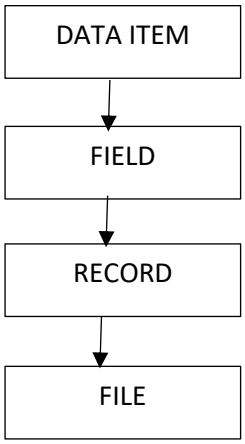
*A FILE*



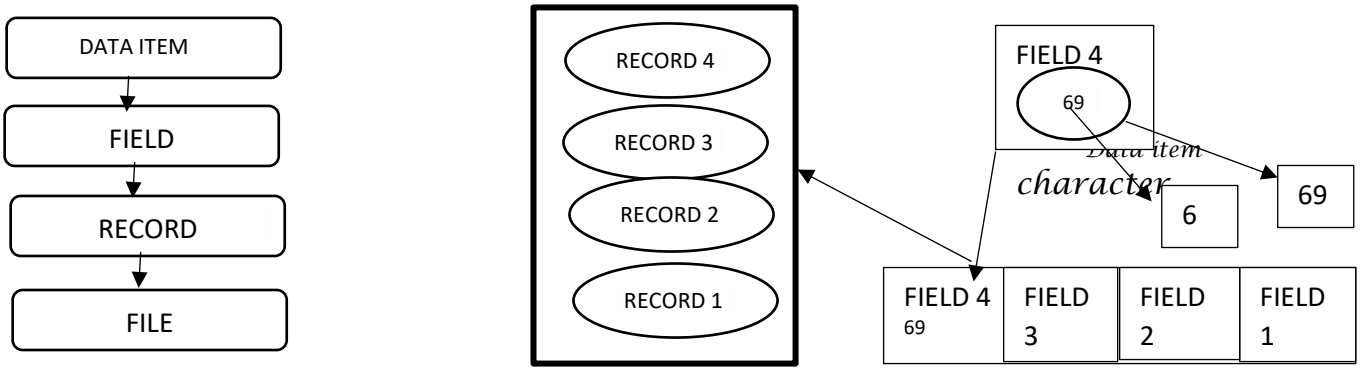
## FILE STRUCTURE FROM DATA ITEM TO FILE

*File structure is the way file is arranged and how the parts are connected to each other. It is organised from simple data item to the file.*

## FILE STRUCTURE




*DETAILED SKETCH OF A FILE STRUCTURE*




*TYPES OF FILE ORGANISATION*

***File organization*** *refers to the physical arrangement of data on storage devices such as the magnetic disk, magnetic tape, optical discs and so on.*

*There are four basic methods of organising files on storage media:*

- *Serial organisation*
- *Sequential organisation*
- *Indexed organisation*
- *Random (or Direct) organisation*

1. ***Serial organisation:*** *the type of file organization that records are arranged one after the other in no particular order. The common storage device used in serial file organization is magnetic tape, punched paper tape, punched cards or magnetic disks.*
   *Serial file organisation is also* ***the simplest file organisation method****. In serial files, records are entered in the order of their creation. As such, the file is unordered, and is at best in chronological order.*
2. ***Sequential organisation:*** *this is a type of file organisation where records are organised in the sequence by which they were entered.*
   *Records in sequential files can be read or written only sequentially. After you place a record into a sequential file, you cannot shorten, lengthen, or delete the record.* ***Magnetic tape*** *is majorly used.*
   *The key difference between a sequential file and a serial file is that it is ordered in a logical sequence based on a key field called* ***Primary key.*** *eg class list sorted on surname or alphabetically, students result sorted on registration number and admission list sorted by departments.*
3. ***Indexed or indexed sequential:*** *in this type of file organisation, records are stored in the file with the help of the primary key. For each primary key, an index value is created and mapped to the record. This index contains the address of the record in the file. This file has multiple keys. Eg Library that uses index.*
4. ***Random organisation****: Records are stored randomly but could be in a scattered manner using a record key as an address on the disk for its location or accessed. Magnetic and optical disks allow data to be stored and accessed randomly. The storage media used in random file organization are magnetic disk and optical disc.eg Online hostel booking and online hotel reservation.*

## *ADVANTAGES AND DISADVANTAGES OF DIFFERENT FILE ORGANISATION*

| *SERIAL ADVANTAGES* | *SERIAL DISADVANTAGES* |
|---|---|
| 1. *It is cheap.* | 1. *No provision for insertion.* |
| 2. *It is simple to manage.* | 2. *Updating is only by copying.* |
| 3. *It is easier to do file backup.* | 3. *It is cumbersome to access.* |
| 4. *It is easy to access the next.* | 4. *Wastage of space.* |

| | |
|---|---|
| | 5. It cannot support modern high-speed requirement for quick record access.<br>6. It takes long searching time. |
| SEQUENTIAL<br>ADVANTAGES<br>1. It allows read/ write sharing.<br>2. It is easy to access records.<br>3. Search time is less.<br>4. It is simple to understand.<br>5. it is easy to organize, and maintain.<br>6. It is efficient and economical. | SEQUENTIAL<br>DISADVANTAGES<br>1. Data redundancy is high.<br>2. Records cannot be retrieved randomly and so getting a record may require searching the entire file.<br>3. Retrieval of files takes a long time.<br>4. It involves sorting of records which will involve extra cost. |
| INDEXED / INDEXED SEQUENTIAL<br>ADVANTAGES<br>1. Access to records is fast.<br>2. Records can be accessed randomly or sequentially.<br>3. It permits the efficient and economical use of sequential processing techniques when the activity ratio is high. | INDEXED / INDEXED SEQUENTIAL<br>DISADVANTAGES<br>1. It requires more storage space to hold the index.<br>2. It is expensive since direct access storage device is involved relatively expensive hardware and software resources are required.<br>3. Access to records may be slower compared to random files. |
| RANDOM<br>ADVANTAGES<br>1. To access and retrieve records are quick and directly.<br>2. Update of files is faster and efficient.<br>3. It does not require any kind of sorting.<br>4. It is useful in accessing heavy databases.<br>Processing efficiency is high. | RANDOM<br>DISADVANTAGES<br>1. Expensive hardware and software resources are required.<br>2. File updating is more difficult compared to sequential files.<br>3. Data may be accidentally erased or overwritten.<br>4. May be less efficient in the use of storage space than sequentially organised files.<br>5. Special security measures are necessary for on line direct file. |

METHODS OF ACCESSING FILES

*To access files mean to retrieve or locate information stored on a file in a storage device. The method of accessing a file depends on how file is organised.*

*Methods of accessing files are:*

1. *Serial access*
2. *Sequential access*
3. *Random access*

1. ***Serial access:*** *in this access method retrieving a record will require searching a file one by one until the file / record to be retrieved is gotten. This means that file is accessed serially which means that the records are read from the storage medium into the main memory one after the other in the order they are stored on magnetic tape.*
2. ***Sequential access:*** *in this method records are accessed sequentially in the order in which they are stored. If the record required is 30$^{th}$, the file to get it into storage to process, computer will first have to read in all 29$^{th}$ preceding records.*
3. ***Random access:*** *this type of access method is where the records are accessed or located by searching directly using index and record key.*

***Computer File Classification***

1. ***Master file****: a computer file whose contents are relatively permanent in nature. It contains data and historical information which may be used as a source of reference .eg customer ledger, payroll, inventory, students master file and customer master file e.t.c.*
   *A feature to know is the constant/regular updating of files to show current position.*
   ***New master file content = old master file content + transaction file content.***
2. ***Transaction file:*** *these are files that are temporal which are created as events occur. They contain information about transactions or activities that are currently being processed. These files are used to update the master files from time to time.*
   ***Note as transaction file is used to update it is no longer required.***
   *Reference files: these are files that are permanent to some extent and are used to ensure validity of master files by constant comparison of both files eg price lists, names and addresses.*

   ***Other files***
   ***Program file:*** *it contains computer instructions or software for processing of data.*
   ***Data file:*** *for storing data*

***Output file:*** *contains data are the output of a device or program.*
***Input file:*** *contains data that serve as input to a device or program.*
***Read only file:*** *file that you can only read but not change or modify.*
***Video file:*** *used to store video.*
***Audio file:*** *used to store sound.*

*Criteria for classifying files*

1. ***Nature of content:*** *the type of data in a file can be used to classify the file. Eg program file contains programs or software and text file that contains text.*
2. ***Method of organisation:*** *the way a file is arranged on a disk can be used as a criteria to classify the file eg. Random file, sequential and serial file.*
3. ***Storage medium:*** *the way files are stored on a storage medium can be used to classify computer files eg Disk file is stored on magnetic disk and Tape file is stored on magnetic tape.*