

File Structures

- **Record:** (Group/ Segment) it is a collection of information items about a particular entity. E.g. A collection of information about a student in a college.
- **Field/Item:** An item of a record is a unit of meaningful information about an entity. E.g. Name, Id no, Address of a student.
- **File:** A collection of records involving a set of entities with certain aspects in common and organized for some particular purpose is called file. E.g. Collection of all students' records makes a file.
- **Key:** A record item or field which uniquely identifies a record in a file is called a key. E.g. IdNo Of a student for student record.
- **Database:** If the set of files is used by the application programs for some particular application area and if these files exhibit certain associations or relationships between the records of the files then such a collection of files is called as database. e.g student database which has files/tables like student information, result, attendance, registration, placement etc. In which all these different have relationship and association between them.
- **Transaction:** A particular operation involving a record or set of records is called as transaction. E.g. "Delete all students of batch = 00"
- Fields, records, files and databases are logical terms.
- **Block/Physical record:** Several logical records can be groped together to form a single physical entity to be stored on physical device is called as block or physical record.
- **Serial-access device:** The devices on which data is stored serially and can be retrieved only serially and access time of any data is proportional to the location of the data on device, are called as serial device. These devices are cheap compared to direct access devices. E.g. Magnetic tape.
- **Direct-access Device:** The devices for which access time for two records don't depend upon their location in memory are called as direct access devices. These devices are costly. E.g. Magnetic disk.
- **File Organization:** The technique used to represent and store the records on a file is called as file organization.

The 4 fundamental File Organizations Are:

1. Sequential
2. Direct/ Random
3. Index Sequential
4. Multi key

• **Sequential File Organization:**

In Sequentially organized file, the records are written consecutively when the file is created and must be accessed consecutively when the file is later used for input or output. Generally all records are of same format and length. But it is possible to have records of different format and variable length in one sequential file. In that case there has to be flags for start and end of record in each record. Generally Sequential file are stored physically by some order like sorted on primary key. One sequential file can not be stored sorted based on two keys. For that we need to have two sequential files having same records but sorted on different keys. Generally they are used in batch processing i.e. when all the records of file have to be performed with some operation. E.g. Calculation of result for all students.

○ Storing Sequential file: Sequential files can be stored on serial- access as well as direct- access devices.

○ Advantages:

- 1. Faster access to next record.
- 2. Access time is very good.
- 3. Simplicity

○ Disadvantages:

- 1. If pattern of access do not match with record ordering pattern then Access time goes very high.
- 2. Data can be accessed sequentially only.

Record 1	Record 2	Record i	Record N-1	Record N
----------	----------	-------	----------	-------	------------	----------

Structure of Sequential File

• **Random/Direct File organization:**

When records in a file are arranged in such a way so that the individual records can be directly accessed whenever they are needed is called as direct File organization. In this file organization there exists a direct relationship between the key used to identify the particular record and the location of the record in a file. The records do not necessarily appear physically in sorted order by their key values as in sequential files. When relative file is created the relationship that will be used to translate between key values and physical address of record in file is assigned. That relationship is called as mapping function R,

R (key value) Physical address

Generally direct file are used when only selected records are accessed from file not all at a time i.e. in interactive processing. E.g. online banking system in which user wants to access only his account.

- Storing direct file: Direct files can only be stored on direct- access devices.
- Advantages:
 1. Ability to access individual record directly. i.e. a record can be retrieved, inserted, modified or deleted without affecting other records.
 2. Access time for one record at a time is very good.
- Disadvantages:
 1. When all records are to be accessed at a time then it gives poor performance.

Key value Physical position

Beginning of file

In file

COW	1
ZEBRA	2
.	
.	
.	
.	
AXE	I
.	
.	
.	
BAT	N-1
CAT	N

End of file

- **Index Sequential File Organization:**
 An effective way to organize a collection of records when there is a need of both accesses. To access the records sequentially by some key value and also to access the records individually by some key is called as index sequential file organization. It supports combination of access types that are supported by a sequential file and a relative file. It is used when there is a need for both batch and interactive processing. E.g. Student management system in which we need to calculate result for every student as well we have requirement like Requirement of address to send a letter to only that

student who is detained. So the file can be stored sequentially based on student id. And index table is also provided with key as student id which directly maps student id to its record as shown in figure.

○ Advantages:

1. Ability to access individual record directly as well as all the records sequentially.

Index Table	Sequential Data File
00CP001	1
00CP007	2
00CP010	3
.	
.	
00CP015	N/2
.	
.	
.	
00CP100	N

Position
00CP050 N/2

00CP007 2 NajfhajkefhNaofjadfj N/

.....

.....

00CP010 3 NajfhajkefhNaofjadfj N/

00CP100 N NajfhajkefhNaofjadfj N/

00CP001 1 NajfhajkefhNaofjadfj N/

● **Multikey File Organization:**

When a file records are made accessed based on more than one key are called as Multikey file organization. This file organization is needed many times. E.g. In banking system we keep records of accounts in file. Now account holder needs account information which can be access through account no, while loan officer needs account records with a given value of overdue limit. So we need to provide to access path to the record based on different need. Generally these files are index sequential file in which file is stored sequentially based on primary key and more than one index table are provided based on different keys. Basically there are 2 approaches for implementing multi key organization.

1. Inverted File Organization:

In this file organization a key's inversion index contain all of the values that the key presently has in the records of the data file. Each key-value entry in the inversion index points to all of the data records that have the corresponding value. The data file is said to be inverted on that key. Inverted files are sorted on inversion index so that binary search can be applied to find out index of record. Whenever record is added in data file its corresponding entry has to be made in inverted file.

2. Multi List File Organization:

In multi list file organization the index contain all values that the secondary key has in data file same as inverted file but the difference is that the entry in the multi index for a secondary key value is pointer to the first data record with that key value. That data record contains pointer to second record having same key. Thus there is a linked list of data records for each value of secondary key. Multi list chains usually are bidirectional and occasionally are circular to improve update operation.

Sequential Data File Sorted on primary key A/C No

Physical Address	A/C No	Name	Amount	A/C Type
1	1111	ABC	500	01
2	2574	XYZ	2000	02
3	2389	STU	3000	03
4	3000	KBC	4000	01
5	2494	YQR	800	01
6	3678	SPZ	500	02

Inverted Index File for secondary key A/C type

A/C Type	Physical Address
01	1,4,5
02	2,6
03	3

Multi List File for secondary key A/C type

Physical Address	A/C No	Name	Amount	A/C Type
1	1111	ABC	500	01
2	2574	XYZ	2000	02
3	2389	STU	3000	03
4	3000	KBC	4000	01
5	2494	YQR	800	01
6	3678	SPZ	500	02

A/C Type	Pointer
01	
02	
03	

You can see the advantage of Multi key file organization that you can search all the records of particular A/C type directly from inverted file as well as all records based on A/C no also.