

## Unit VI

Page No.

Date

- ① What is file? Enlist the operation that can be performed on file? If different file opening modes in C++?



- File can be defined as collection of records in which each record consists of one or more than one field.

For example:-

Rollno	Name	marks	Address
10	AAA	90	Bangalore
20	BBB	88	Pune
30	CCC	76	Gurgaon

\* Various operation that can be performed on file:-

- ① Insertion of record into file
- ② Deletion of record
- ③ Updation of record
- ④ Retrieval of record.

\* Different file ~~open~~ opening modes:-

- ① `ios::in` :- Open file for input operation
- ② `ios::out` :- Open file for output operation
- ③ `ios::binary` :- Open for binary operation

(4) `fpos::ate:-`

If flag is set initial position is set at ~~the~~ end of file otherwise initial position is set at beginning of file.

(5) `fpos::app:-` The output operation are appended to file.

Content are inserted at end of file.

(6) `fpos::trunc:-` The content of existing file get destroyed and replace by new file.

- we use this opening modes with help of open function:

`ofstream obj;`

For e.g. `obj.open("a.txt", fpos::out);`

(2) What is sequential file organization? State its advantage and disadvantage? Also explain operation:- ① Add ② Delete ③ Search

- File function `getw()`, `putw()`, `fsync()`, `FPrint()`, `get()`, `put()` are useful in reading and writing data sequentially.

- In sequential file organization both reading and write is done sequentially.

- Records are added in order they are arrived.

### \*Characteristics of Sequential Files:-

- In Sequential files, records are added in order they arrive.
- The length of record is not fixed. Such files is used to store text. In text, two lines need not to be have same size.
- Searching record is time consuming to access nth record all record storing from record no 1 to n-1 must be read to reach nth record.
- Insertion or deletion is also time consuming.

Record1 | R2 | R3 | R4 | R5 | ... | EOF

↳ End of file

### \*Advantage of Sequential Organization

- ① The Sequential organization is simple to implement.
- ② It is suitable for small database application.
- ③ Sequentially files organization is inherently accurate as it maintains simple order.

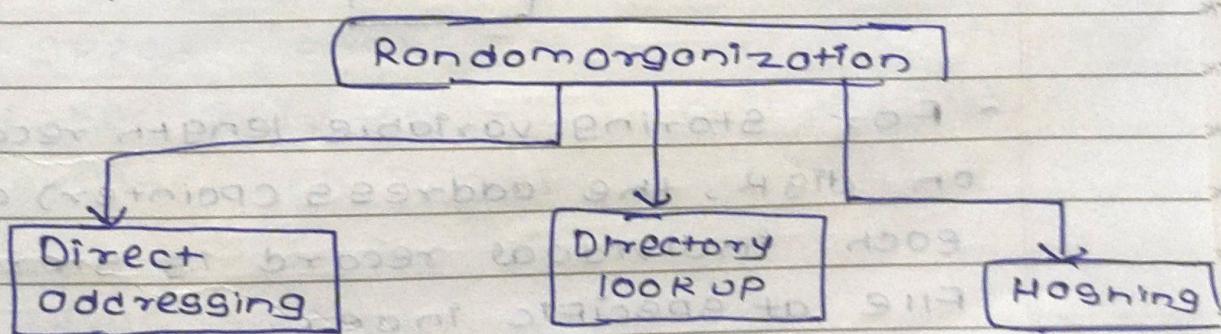
### \* Disadvantages

- ① If database is very huge then it is not efficient technique for retrieval of record.
- ② Searching, insertion, accessing is very time consuming.
- ③ When the record is deleted the space which is occupied by record remains unused.
- ④ Explain direct access file organization State its advantage and disadvantages?

• 881 950 6m

→ Direct access file / Random organization is kind of file organization in which records are stored at random location on disks.

There are three techniques used in Random organization



## ① Direct addressing :-

- In direct addressing two type of record are handled :-

① Fixed length record

② Variable length record

- For storing Fixed length record the disk space is divided into nodes.

- Every fixed length record is stored in node number # which has some primary key value.

eg - ~~record~~ primary key value

is 185 then record must be present in node 185.

For eg:-

100	
500	
1000	1000
-	
-	

- For storing variable length record on disk, the address (pointer) of each individual record is stored in file at specific index.

- we can locate variable length file using index of pointer.

H \* draw back  
① more disk access

## ② Directory lookup :-

- In this index for pointer to record is maintained.
- For retrieving the desired record first all index for record address is searched then using this record address actual record is accessed.

## ③ Hashing :-

- Hashing is a technique in which hash key is obtained using some hash function and record is placed in hash table using hash key.
- In random organization, record can quickly be searched with help of hash function being used.

### \* Advantage of direct access:-

- It provides fast and efficient access to specific data in file.
- It is also efficient for updation and editing data in file.

### \* Disadvantage:-

- Direct access is more complex and difficult to implement than sequential.
- It requires more memory to store index or address information which make file larger than sequential access.

- (4) Explain index sequential file organization and compare it with direct access?

\* Index sequential file organization :-

- The main drawback of sequential file is that searching operation is not efficient. Because in sequential organization primary key of every record is composed with searching key.

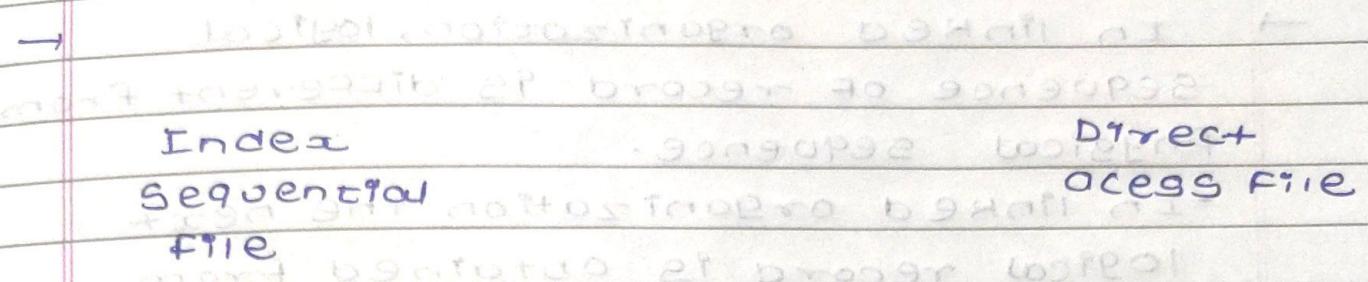
- To optimize searching index sequential file is introduced.

- In index sequential file a separate file for storing indexes of every record is maintained along with master file.

EmpID POS		emp_id Name sal			
10	5	0	20	BBB	2000
20	1	2	40	DDD	4000
40	3	3	5	10	AAA
		4			1000

- In this case, we need not have to scan the entire memory block of record instead of that using primary key and position we can access record.

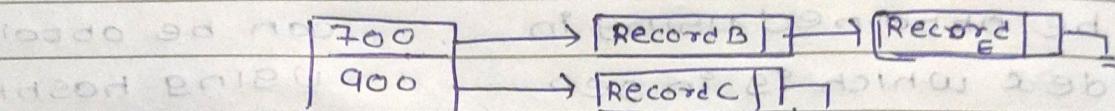
## \* Comparing index Sequential and direct access:



- |  |  |
|--|--|
| <p>① Desired record can be obtained using index which is in sequential file.</p> <p>② Insertion and deletion operation can be performed by index manipulation.</p> <p>③ Variable length records are allowed.</p> <p>④ Records are arranged sequentially in master file.</p> <p>⑤ It is less efficient.</p> | <p>① Desired record can be obtained using key.</p> <p>② On insertion or deletion of record collision may occur.</p> <p>④ Fixed length record should be there.</p> <p>⑤ Records are arranged randomly.</p> <p>⑥ It is more efficient.</p> |
|--|--|

Q) What is linked organization? Explain inverted file and coral rings?

- In linked organization, logical sequence of record is different from physical sequence.
- In linked organization the next logical record is obtained from linked value from present record.



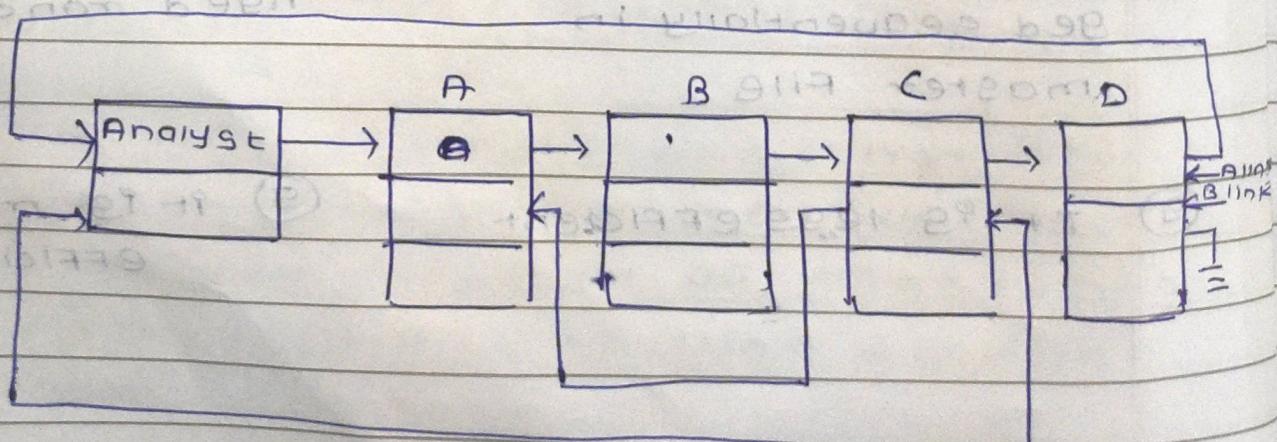
\*~~Inverted file~~:

### \* Coral Rings =

- The coral ring structure is similar to doubly linked multilist.

- Each list is structured as circular list with head node.

- The coral linked structure utilize back pointer to facilitate deletion without having to transverse over the entire structure.



- When deleting node back pointer allow direct access to previous node, enabling deleting without traversing from front.

- Here in Fig

Forward Circular list contains:  
A, B, C, D

Reverse Circular list contains:  
C, A

Each node has two link fields:-

① A link

② B link

① A link :- A link is used to link one record with some value of key K.

② B link :- B link is used to connect alternate record in reverse order.

\* Explain multilist linked file organization.  
 ⇒

- In Unlinked organization logical sequence of records is different from physical sequence. In a linked organization the next logical record is obtained by following a link value from the present record. It is a linked list.
- Linking records together in order of increasing primary key value facilitates easy insertion and deletion once the place at which the insertion or to be made is known.
- In multilist file organization the index contains 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 linked list of data records for each value of secondary key. multilist chains usually are bidirectional and occasionally are circular to improve update operation.

E#	Name	Sex	Salary
A 800	sid	M	10,000
B 510	atharv	M	15,000
C 950	gandharv	M	12,000
D 750	pranav	M	12,000
E 620	Ashwin	M	16,000

fig. sample data

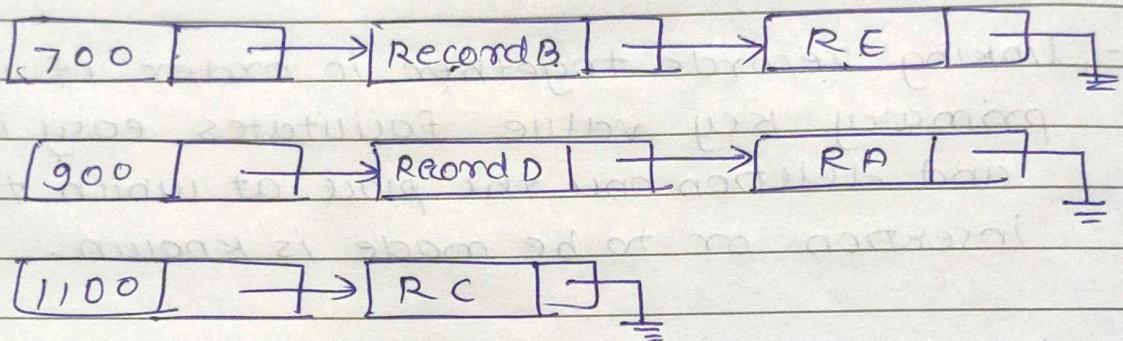


fig. Record Linking.

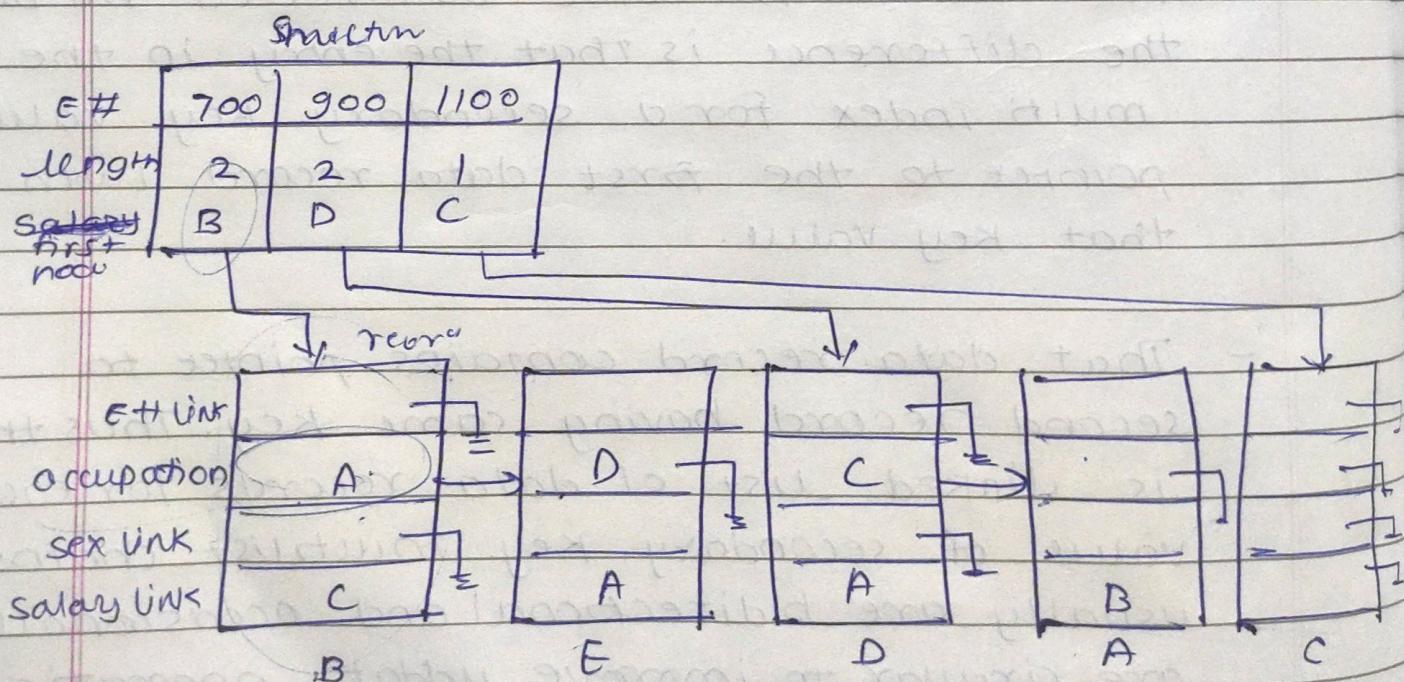


fig. Multilist representation

Q Explain cellular partition.

- 
- For reducing the searching time during file operations, the storage media may be divided into cells.
  - The cells can be of 2 types.
    - i) Entire disk pack can be cell.
    - ii) A cylinder can be cell.
  - If all the records lie on the same cylinder then without moving read/write head, the records can be accessed.
  - The primary goal of partitioning is to break down a large dataset or problem into smaller subsets that are easier to handle & process.

#### Advantages:

- 1) Various read operations can be performed parallelly in order to reduce the search time.
- 2) Faster execution of any query.

#### Disadvantages:

- If multiple records lies in same cell then reading a single cell become time consuming.

## 6.11 Inverted File Organisation

### University Questions

Q. What do you mean by inverted file organisation ?

Q. What is linked organization? Describe inverted Files and cellular partitions w. r. t. Linked organization.

SPPU - May 16, 6 Marks

Q. Explain linked organization with respect to inverted files.

SPPU - Dec. 19, 7 Marks

Conceptually, inverted files are similar to multilists. The difference is that while in multilists records, With the same key values are linked together. Information about link is kept in individual records. In case of inverted files this link information is kept in the index itself. Fig. 6.11.1 shows the indices for the file of Fig. 6.10.1.

E# index		Salary Index	
700	B,E	900	E
900	A,D	1200	A,C,D
1100	C	1500	B
Occupation index		Sex index	
Analyst	B,C	Female	B,C,D
Programmer	A,D,E	Male	A,E

Fig. 6.11.1 : Indices for fully inverted file

- In inverted files, only the index structure is important. Record can be stored in any way.
- Inverted files may also result in space saving when record retrieval does not require retrieval of key fields.
- In case of inverted files, the key fields may be deleted from the records.
- Insertion and deletion of records requires only the ability to insert and delete within indices.
- Index maintenance is more complex than for multilist.
- Number of disk accesses to process a query is equal to the number of records being retrieved plus the number to process the indices.