OSY - Notes


V2V EDTECH LLP
Online Coaching at an Affordable Price.
OUR SERVICES:
• Diploma in All Branches, All Subjects
• Degree in All Branches, All Subjects
• BSCIT / CS
• Professional Courses
+91 93260 50669
V2V EdTech LLP
v2vedtech.com
v2vedtech

Unit - V File Management

➢ 5.1 File Concepts

In an operating system, a file is a collection of data or information that is stored on a storage device (such as a hard disk, SSD, or USB drive). Files are used to store data in a structural format. We can perform different types of operations on file like creating file, Reading file, writing file, Deleting file, Renaming file, Moving file, Copying file. A file is a named collection of related information that is recorded on secondary storage such as magnetic disks, magnetic tapes and optical disks. Commonly files represent programs (source and object forms) and data. Data files may be numeric, alphabetic, alphanumeric or binary. In general a file is a sequence of bits, bytes, lines or records whose meaning is defined by the file's creator and user. The information in a file is defined by its creator.

- **File Attributes**

Describe any four file attributes (W – 19, S - 23)

1. **Name:**
   The symbolic file name is the only information kept in human readable form.

2. **Identifiers:**
   File System gives a unique tag or number that identifies file within file system and which is used to refer files internally.

3. **Type:**
   This information is needed for those system that support different types.

4. **Location:**
   This information is a pointer to a device and to the location of the file on the device.

5. **Size:**
   The current size of the file (in bytes, words or blocks) and possibly the maximum allowed size are included in this attribute.

6. **Protection:**
   Access control information determines that who can do reading, writing, executing and so on.

7. **Time, date and user Identification:**
   This information may be kept for creation, last modification and last use. These data can be useful for protection, security and usage monitoring.

- **Operations**

List any four operations performed on a file (W – 19, W-22, W -23, S - 24)
Write any four systems call related to file management (W - 22)

1. Creating a file
2. Writing file
3. Reading a file
4. Repositioning within a file
5. Deleting a file
6. Truncating a file
7. Appending new information to the end of the file
8. Renaming an existing file
9. Creating copy of a file, copy file to another I/O device such as printer or display.

1. **Creating a File:** This involves two main steps. First, space is allocated within the file system. Second, a new entry for the file is created in the directory, which includes the file's name and its location.

2. **Writing a File:** To write to a file, a system call is made, specifying both the file name and the information to be written. The system then searches the directory for the file's location. The write pointer is updated after each write operation.

3. **Reading a File:** To read from a file, a system call specifies the file's name and the memory location where the data should be placed. The system maintains a read pointer to indicate the next block of the file to be read. Once data is read, the read pointer is updated.

4. **Repositioning within a File:** This operation, also known as a file seek, involves searching the directory for the appropriate file entry and then setting the current file position to a specified value. It does not require any actual data movement.

5. **Deleting a File:** To delete a file, the directory is searched for the named file. Once found, all associated file space is released, and the directory entry is erased.

6. **Truncating a File:** Instead of deleting and recreating a file, this function allows a user to erase the contents of a file while retaining its attributes.

- **File types and File system structure**

In an operating system, a file is a collection of data or information that is stored on a storage device (such as a hard disk, SSD, or USB drive). Files are used to store data in a structural format. We can perform different types of operations on file like creating file, Reading file, writing file, Deleting file, Renaming file, Moving file, Copying file. A file is a named collection of related information that is recorded on secondary storage such as magnetic disks, magnetic tapes and optical disks. Commonly files represent programs (source and object forms) and data. Data files may be numeric, alphabetic, alphanumeric or binary. In general a file is a sequence of bits, bytes, lines or records whose meaning is defined by the file's creator and user. The information in a file is defined by its creator.

File Types :

The file name is split into two parts: a name of file and extension. For example "student.xls" the file name is student and it is created in MS - Excel application. The system uses the extension to indicate the type of the file and the type of operations that can be done on that file.

The following table indicates the common file types.

File Type	Usual Extension	Function
Executable	exe, com, bin, none	Ready to run machine language program.
Object	obj, o	Compiled, machine language, not linked.
Source code	c, cc, pas, asm, f77	Source code in various languages.
Text	text, doc	Textual data documents.
Batch	bat, sh	Commands to the command interpreter.
Word Processor	wp, rtf, tex, doc, etc.	Various Word Processor formats.
Library	lib, a, dll	Libraries of routines for programmers.

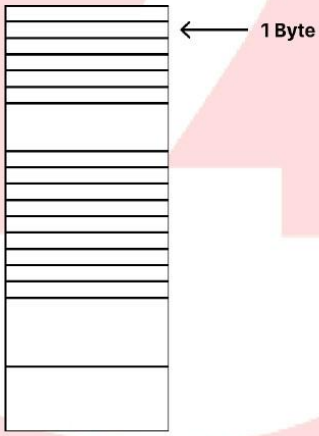
Print or view	ps, gif, dvi, pdf	ASCII or binary file in a format for printing or viewing.
Archive	arc, zip, tar	Related files grouped into one file; sometimes compressed, for archiving or storage.
Multimedia	avi, mp3, mov, mkv, mpeg, rm	Binary file containing audio or audio/video information.

File Structure :

Files can be structured in any of several ways.

1) Stream of Bytes :

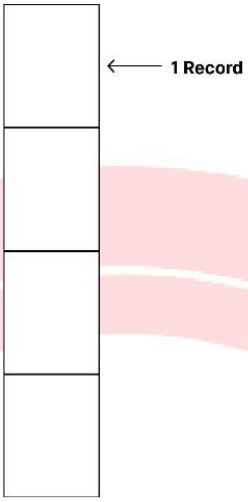
The file is treated as a continuous sequence of bytes. Data is stored in a linear fashion with no inherent structure. Commonly used for binary files where data is read and written in chunks.



2) Records :

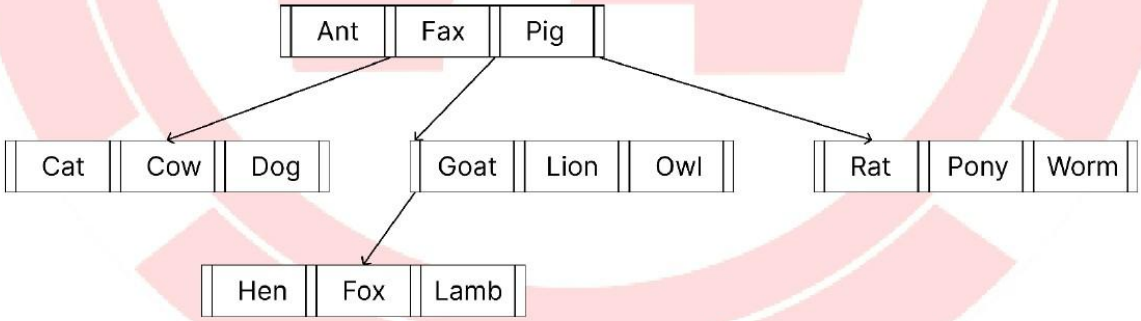
The file is divided into fixed or variable -sized records. Each record is a logical unit, often representing a single entry like a row in a database. Makes it easier to retrieve, update, or delete specific records.

Example :- employee record



3) Tree of Records :

Data is organized hierarchically, often in a tree structure. Each record can have "child" records, forming Parent-child relationships. Efficient for data retrieval, especially for hierarchical data structures like directories or XML data.

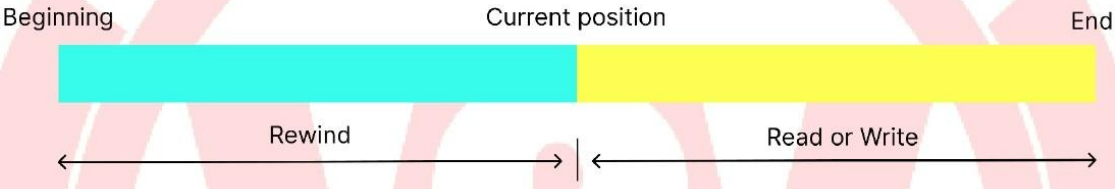


➢ 5.2 Accessing Methods

Describe sequential and direct access method (W - 19)

Sequential File Access :

Information from the file is processed in order i.e. one record after another. It is commonly used access mode. For example, editors and compilers access file in sequence. A read operation read information from the file in a sequence i.e reads the next portion of the file and automatically advances a file pointer. A write operation writes information into the file in a sequence i.e. appends to the end of the file and advances the end of the newly written material. In some operating system, a program may be able to skip forward or backward in records for some integer n.



Sequential file access

Advantages of sequential files:

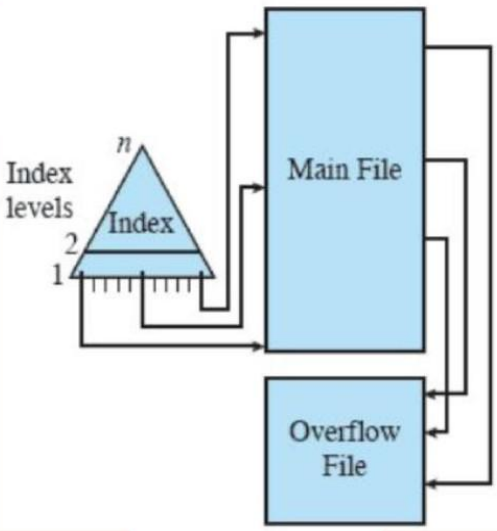
1. Easy to access the next record.
2. Data organization is very simple.
3. Automatic backup copy is created.

Disadvantages of sequential file:

1. Wastage of memory space because of master file and transaction file.
2. It is more time consuming since, reading, writing and searching always start from beginning of file.

Indexed Sequential File Access

An indexed sequential file is a sequential file in which the records are indexed. An indexed sequential file is an improvement over a sequential file. Two features are added in this file namely, an index to the file and an overflow file. Indexing of records provides the facility of searching the records randomly. An indexed file is a simple sequential file that contains an index as its records. Entries in indexed files are made up of two fields, the key field, which is the same as the key field in the main file and a pointer pointing to some record in the main file. To find a specific field in the main file, the index is searched for the highest key value, which is equivalent to the desired value. The pointer related to key field start searching the record at location it indicates. The search continues in the main file at the location indicated by the pointer.



Advantages of Index Sequential File :

1. Variable length records are allowed.
2. Indexed sequential file may be updated in sequential or random mode.
3. very fast operation.

Disadvantages of Index sequential File :

1. as the file grows, a performance deteriorates rapidly.
2. When a new record is added to the main file, all of the index files must be updated.
3. consumes large memory space for maintaining index files.

Direct access :

It is also called as relative access. A file is made up of fixed length logical records that allow programs to read and write records rapidly in no particular order. Direct access method is based on disk model of file which allows random access to any file block. For direct access, a file is viewed as a numbered sequence of block or record so we can directly read block 14, then block 53, and so on. Read n operation is used to read the nth block from the file whereas write n is used to write in that block. The block number provided by the user to the operating system is a relative block number. A relative block number is an index relative to the beginning of the file.

Direct file access

Advantages of Direct File Access:

1. Using this method we can access any records randomly.
2. It gives faster retrieval of records.

Disadvantages of Direct File Access:

1. Wastage of storage space, if hashing algorithm is not chosen properly.
2. This method is complex and expensive.

➢ 5.3 File Allocation Methods

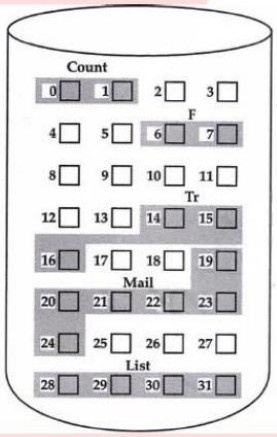
From the user's point of view, a file is an abstract data type. It can be created, opened, written, read, closed and deleted without any real concern for its implementation. The implementation of a file is a problem for the operating system. The main problem is how to allocate space to these files so that disk space is effectively utilized and file can be quickly accessed.

Three major methods of allocating disk space are in wide use:

i) Contiguous file allocation.
ii) Linked file allocation.
iii) Indexed file allocation.

• Contiguous Allocation

Enlist different file allocation methods? Explain contiguous allocation method in detail (W – 19, S – 23, S - 24)



Directory

File	Start	Length
Count	0	2
TR	14	3
Mail	19	6
List	28	4
F	6	2

The Contiguous allocation method requires each file to occupy a set of contiguous address on the disk. Contiguous allocation of a file is defined by the disk address of the first block and its length. If the file is 'n' blocks long and starts at location 'b', then it occupies blocks b, b+1, b+2, ... b+n-1. Contiguous allocation supports both sequential and direct access. The

difficulty with Contiguous allocation is finding space for a new file. If file to be created are 'n' blocks long, We must search free space list for 'n' free Contiguous blocks.

Advantages of Contiguous File Allocation Method

① Supports both sequential and direct access methods.

② Contiguous allocation is the best form of allocation for sequential files.

③ It is also easy to retrieve a single block from a file...

④ Reading all blocks belonging to each file is very fast.

⑤ provides good performance.

Disadvantages of Contiguous File Allocation Method

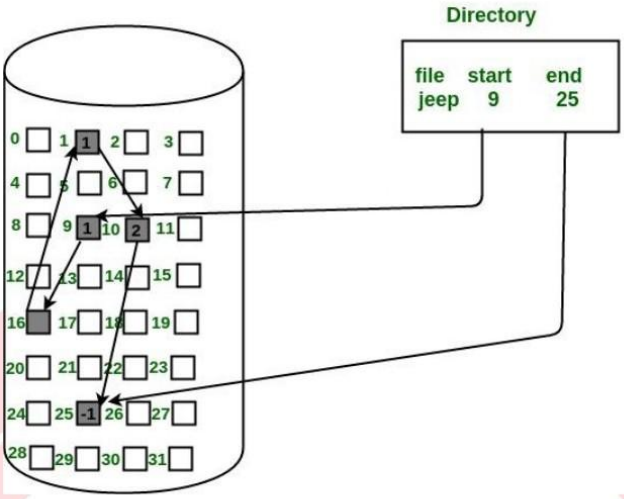
① Suffers from external Fragmentation.

② Very difficult to find contiguous blocks of space for new files.

③ Compaction may be required and it can be very expensive.

- **Linked file allocation :**

  Describe linked file allocation method with suitable example. Also list its one advantage (S -22, W – 22, W - 23)



Linked allocation solves all problems of contiguous allocation. This allocation is on the basis of an individual block. Each block contains a pointer to the next block in the chain. The disk block can be scattered anywhere on the disk. The directory contains a pointer to the first and last block of the file. To create a new file, simply create a new entry in the directory.

Advantages of Linked File Allocation Method :

① Any free blocks can be added to a chain.

② These is no external fragmentation.

③ Best suited for sequential files that are to be processed sequentially.

④ No need to know the size of the file in advance.

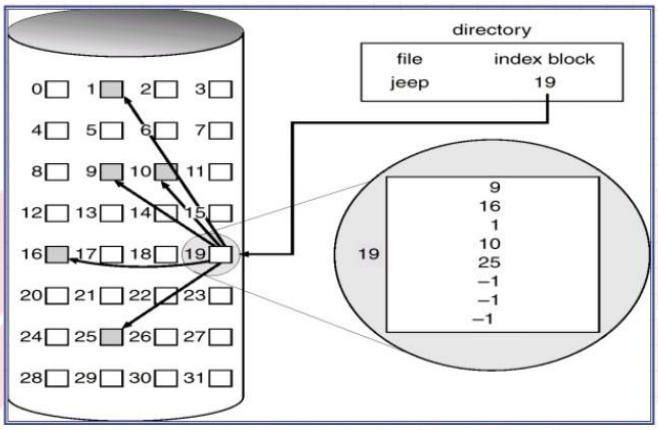
Disadvantages of Linked File Allocation Method :

① This method requires more space to store pointers.

② This method cannot support direct access.

③ It is not an efficient scheme because the list traversal needs to read each block which is quite time consuming.

- **Indexed File Allocation**



In this method, each file has its own index block. This index block is an array of disk block address. When a file is created, an index block and other disk blocks according to the file size are allocated to that file. Pointer to each allocated block is stored in the index block of that file. It contains file name and address of index block. When any block is allocated to the file, its address is updated in the index block. Each i^{th} entry in the index block points to the i^{th} block of the file. To find and read the i^{th} block, we use Pointer to that block from index block.

Advantages of Indexed File Allocation Method :

1. Does not suffer from external fragmentation.
2. Support both sequential and direct access to the file.
3. No Need for user to know size of the file in advance.
4. Entire block is available for data as no space is occupied by pointers.

Disadvantages of Indexed File Allocation Method :

1. It required lot of space for keeping pointers so wasted space of memory.
2. Indexed allocation is more complex and time consuming.
3. keeping index in memory requires space.

➢ 5.4 Directory Structure

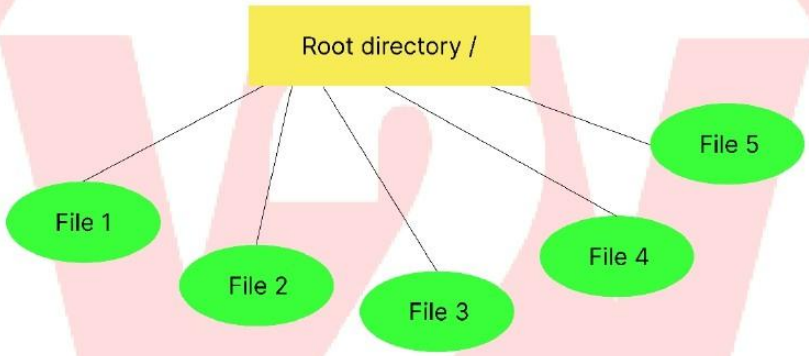
Describe following directory structures in short with neat sketches:

i) Single level

ii) Two level

iii) Tree structured (S – 22, W – 22, S – 23, W – 23, S - 24)

- **Single Level Directory Structure**



It is the simplest form of directory structure is having one directory containing all the files. Sometimes it is called the root directory. In this directory structure unique name must be assigned to each file of the directory. The single level directory structure appears as the list of files or a sequential file having the file names serving as the key. Single level directory structure was implemented in the older versions of single user system.

Advantages of single level directory structure

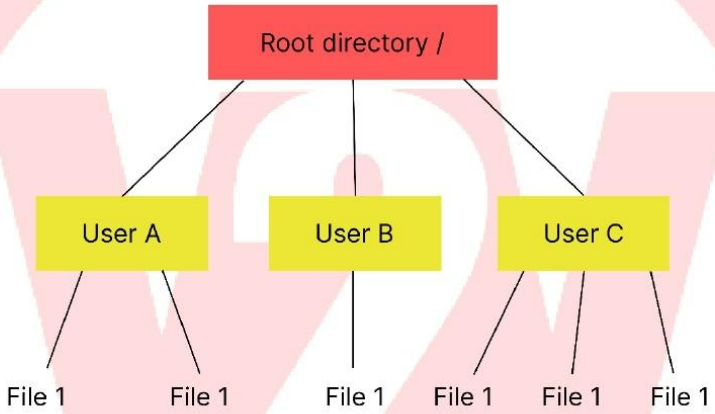
1. Single level directory structure is easy to implement and maintain.
2. It is simple directory structure.
3. Single level directory structure, the operations like creation, searching, deletion, updating are very easy and faster.

Disadvantages of Single Level directory structure

1. It having only one directory in a system so there may chance of name collision.

2. In single level directory structure, it is difficult to keep track of the files.
3. This directory is not used on multi- user systems.
4. The file such as graphics, text etc. are inconvenient for this data structure.

- **Two-level directory**



In two level directory structure, a separate directory is provided to each user and all these directories are contained and indexed in the master directory. The user directory represents a list of files of a specific user. In this directory structure, each user has its private directory known as user file directory (UFD). When a user refers to a particular file, only his own UFD is searched. Thus different users may have files with the same name. Two level directory structure used on a multi-user computer or on a simple network of personal computers that shared a common file server over a local area network.

A two-level directory as a tree of height 2:

i) The root of the tree is the MFD.

ii) Its direct descendants are the UFDs.

Advantages of Two Level Directory Structure

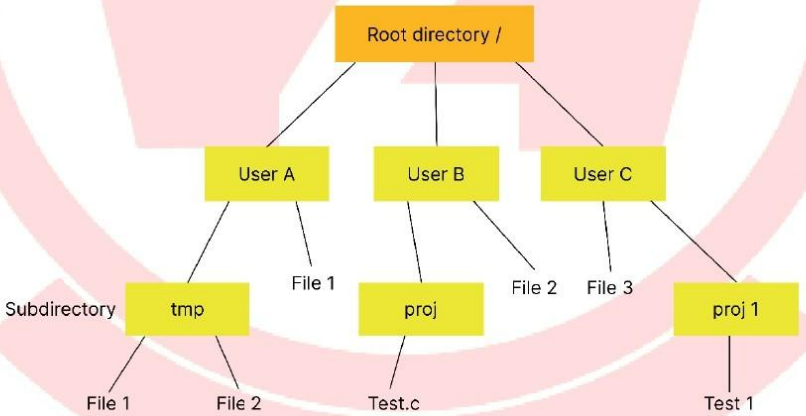
1. It solves the file name collision problem by creating own user directory.

2. This method isolates one user from another and protects user's files.
3. Different users may have files with same name.

Disadvantages of Two Level Directory:

1. Still it not very scalable, two files of the same type cannot be grouped together in the same user.
2. Sharing of files by different users is difficult.

- **Tree Structure**



In this directory structure users can create their own sub-directories and organize their files. The tree has a root directory and every file has a unique path name. A directory contains a set of files or subdirectories. All directories have the same internal format. One bit in each directory defines the entry as a file (0) or as a subdirectory (1). Each process has a current

directory. Current directory contains files that are currently required by the process. When reference is made to a file, the current directory is searched. If a file needed that is not in the current directory, then the user usually must either specify path name or change the current directory.

Advantages of Tree Structured Directory

1. User can create directory as well as subdirectory.
2. It provides a better structure to the system.
3. Managing millions of files is easy with tree structured directory.

Disadvantages of Tree Structured Directory

1. The tree structure can create duplicate copies of the files.
2. The users could not share files or directories.
3. It is inefficient, because accessing a file may go under multiple directories.4.
4. Search time may become unnecessarily long.