Chapter – 8

File Systems

File concept

- The operating system provides a uniform logical view of the information stored in terms of a logical storage unit is known as file.
- Structure of file depends on
  - Data
    - numeric
    - character
    - binary
  - Programs

Types of files

- Text file

  lines or pages, which are sequence of characters.

- Source file

  It consists of subroutines and functions which consists of declarations followed by executable statement.

- Object file

  It contains sequence of bytes forming

File Attributes

- **Name** – only information kept in human-readable form
- **Identifier** – unique tag (number) identifies file within file system
- **Type** – needed for systems that support different types
- **Location** – pointer to file location on device
- **Size** – current file size

File operations

- **Create**

  Free space on the disk is allocated.

  An entry for the file is made in the directory.

- **Write**

  Specify the name of the file and information to be written into the file. Write pointer is updated whenever write occurs.

- Repositioning in a file

The directory is searched for the required entry and the current file position pointer is set to a given value. Repositioning is also referred to as file seek.

- Deleting a file

The named file is searched in the directory ,on finding it, all of its free space is released and its entry in the directory is released.

- Truncating a file

  Erase only the contents of a file and not the attributes.

Other operations are

Append , copy and rename

File Types – Name, Extension

file type	usual extension	function
executable	exe, com, bin or none	ready-to-run machine-language program
object	obj, o	compiled, machine language, not linked
source code	c, cc, java, pas, asm, a	source code in various languages
batch	bat, sh	commands to the command interpreter
text	txt, doc	textual data, documents
word processor	wp, tex, rtf, doc	various word-processor formats
library	lib, a, so, dll	libraries of routines for programmers
print or view	ps, pdf, jpg	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	mpeg, mov, rm, mp3, avi	binary file containing audio or A/V information

File structure

- File types can also indicate the file structure of the file.
- Text files
- Executable file
- Binary file

Functions of file management system

- Mapping the logical file address to Physical address
- Management of disk space and its allocation and de-allocation
- Keeping track of all files
- To provide I/O support for various types of storage devices
- To provide I/O support for various users
- To provide support for protection and

File Access Methods

- A file stores the information which is accessed by user programs.
- Sequential Access
- Direct Access
- Indexed sequential Access

Sequential Access

- It is the simplest method of accessing a file.
- In this scheme , information is processed in sequential order, one record after the other.
- Read operation

  Reads the next record of the file and advances the file pointer.

- Write operation

Direct access methods

- It is based on the disk model of file, disks allow random access to any file block.
- Another name for this access is relative access
- A direct access is viewed as block of records.
- Any block can be read or written
- Read operation

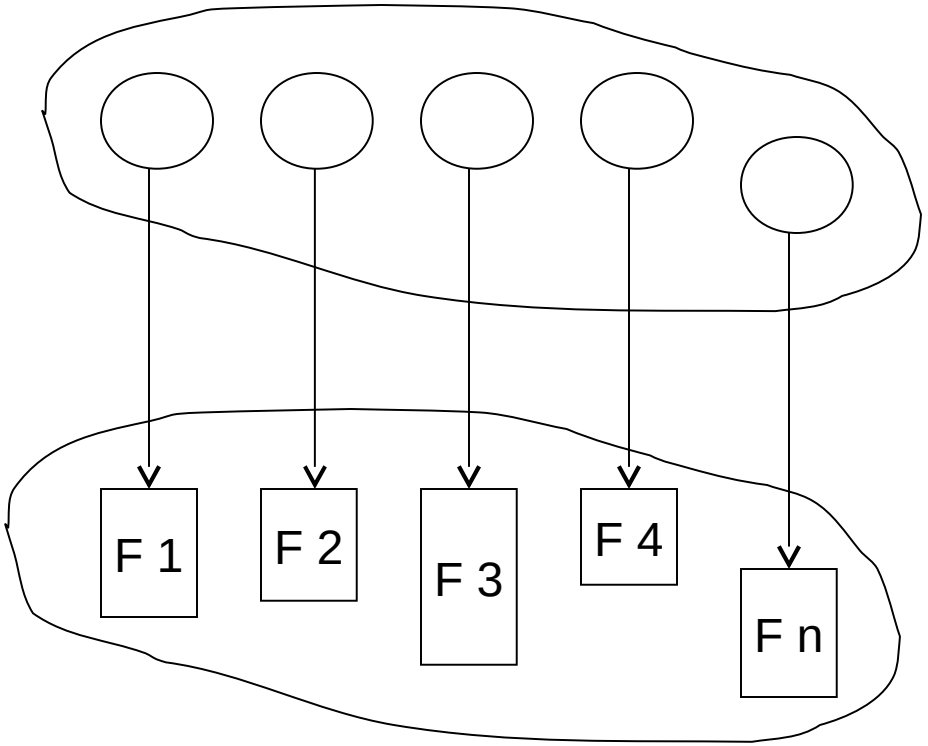
Read n means read the information in the

Indexed sequential file

- This method involves construction of an index for the file.
- Index is organized in sequence based on the key field and contains the pointer to the various blocks.
- For large files , index files itself large.
- To overcome this problem, two index files are created. The primary index file contains pointers to the secondary index files which in turn point to the actual data

DIRECTORY STRUCTURE

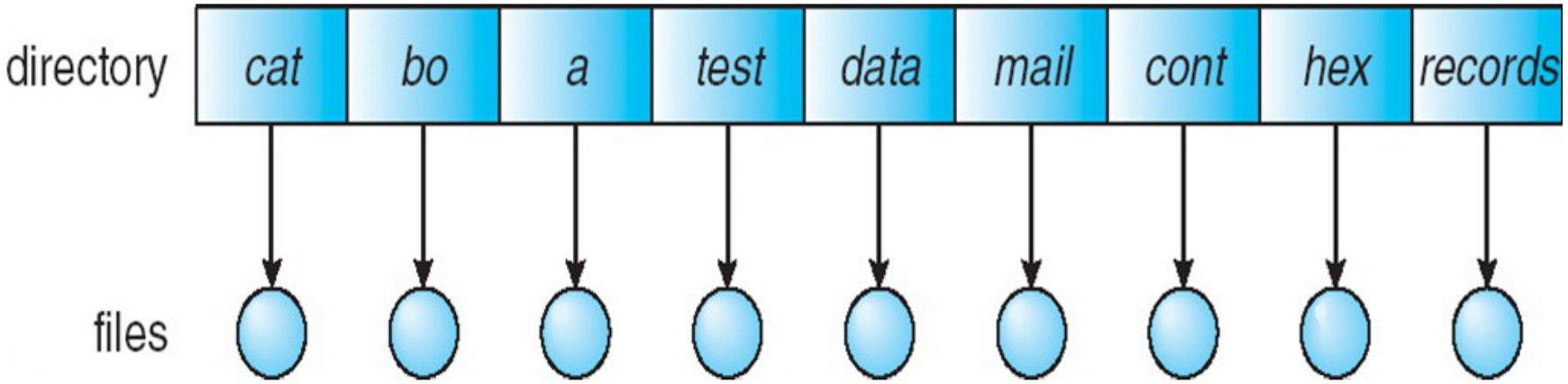
A collection of nodes containing information about all files



Operations in Directory

- Search for a file
- Create a file
- Delete a file
- List a directory
- Rename a file

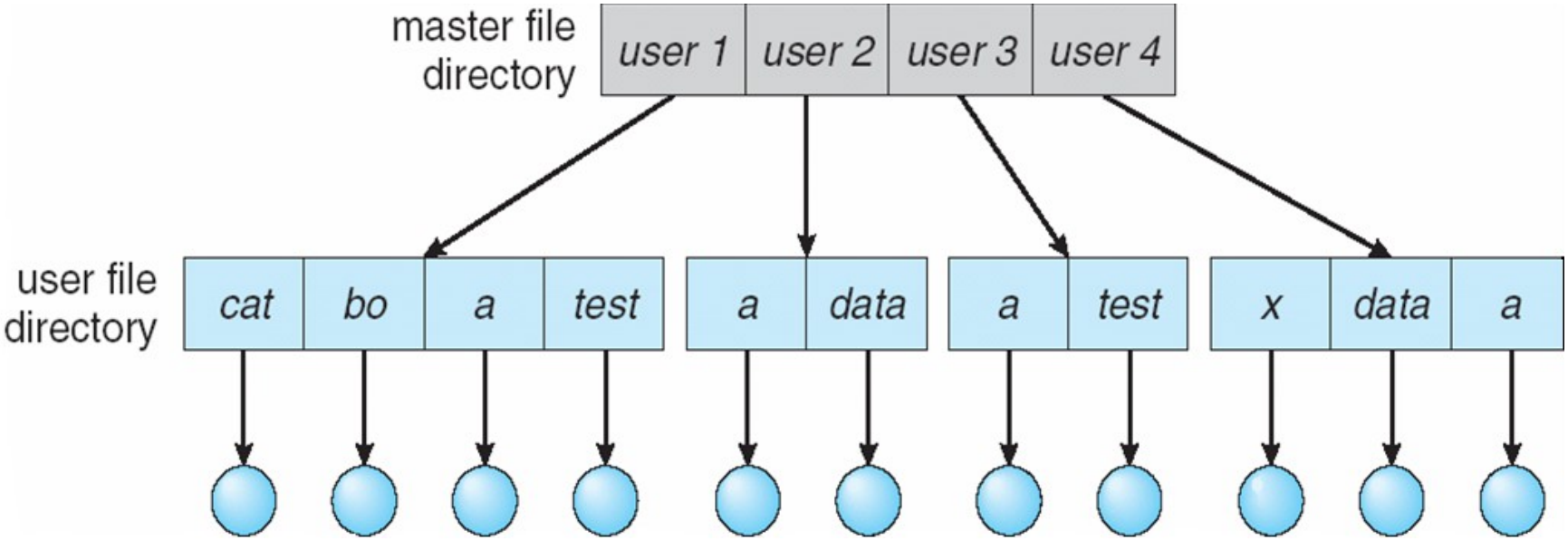
Single-Level Directory



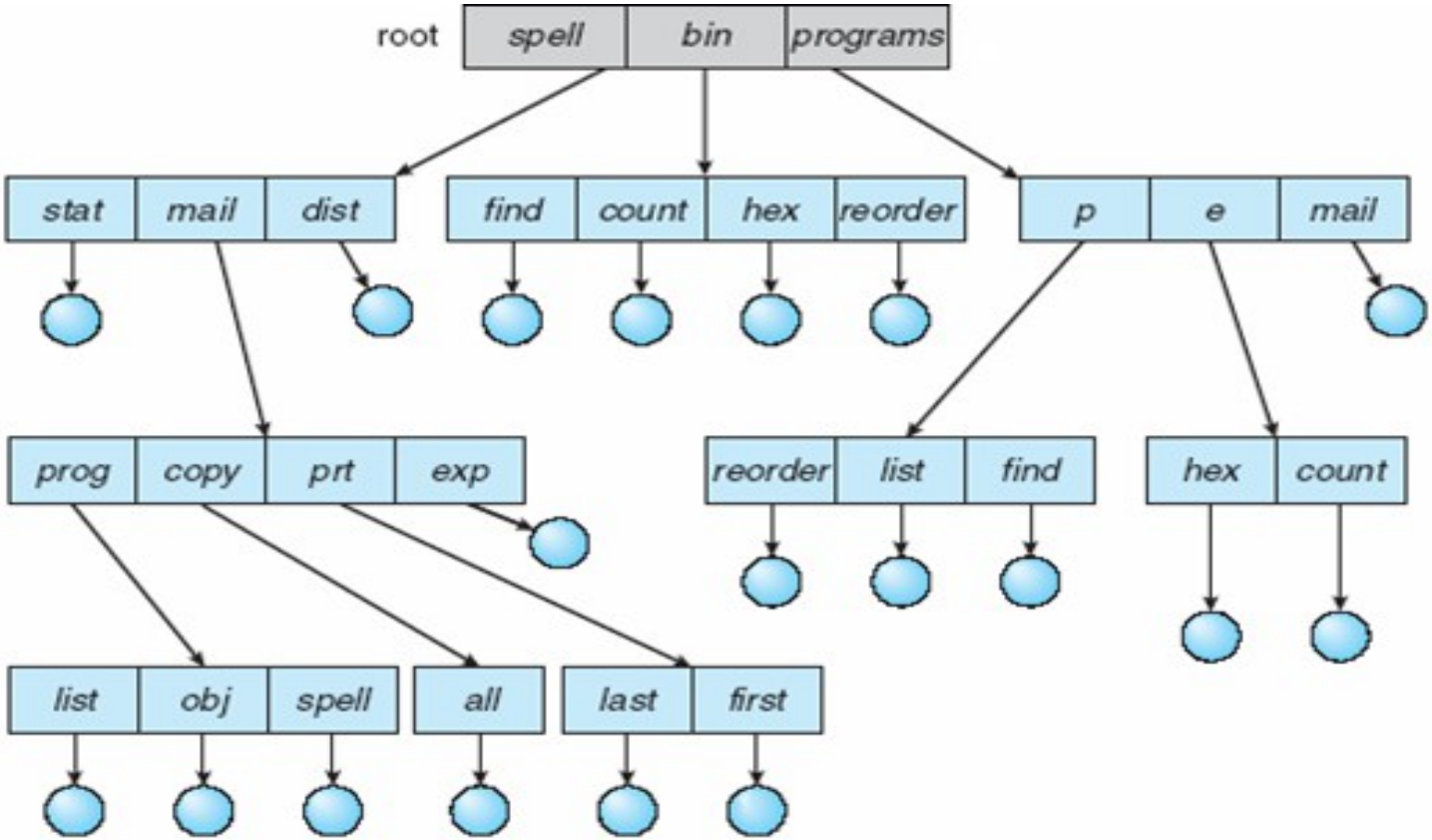
Two level directory

- Separate directory for each user – UFD (User File Directory)
- MFD (Master File Directory) – Root Directory
- Can have the same file name for different user
- Creation of file – File is created in UFD
- Deletion of file – OS searches for that file in UFD. It cannot delete another user’s file

Two level directory



Tree structured directories



Tree structured directory

- This structure allows the user to create their own sub directory and organize the files within them.
- A directory can consists of set of files or sub directories or both.
- A path name is the path from root through the sub directories to the required file.
- Absolute path
- Relative path

- Absolute path

  It starts at the root and follows through its sub directories to the specified file

E-x BCA/BCA2/OS/B2

- Relative path

  It describes the path starting from the current directory.

E-x OS/B2

Deletion of a directory

- In MS-DOS , a directory cannot be deleted unless it is empty. The user must delete all the files and sub directories in the directory
- In UNIX, rm command is used to delete all the files and sub directory in that directory.
- Users can access the files of other users too in this tree structured system.

File Protection

- A file in computer system must be protected from
- Physical damage
- Improper access

Physical damage can be caused by

- Error in reading or writing
- Power failures
- Head crashes

File protection

- Reliabilty

  Keeping back-up of all files

- Protection

A single user system is protected by physically removing the floppy disk or cd.

A multi-user systems can be protected by providing the access permissions to files.

Types of Access

- **Read**
- **Write**
- **Execute**
- **Append**
- **Delete**
- **List**
- **Rename,copy,edit**

Access control list

ACL is associated with each file and directory ,specifying the user name and type of access allowed for each user.

Access Groups

- **Owner** who has created the file
- **Group** is a set of users who share the file
- **Universal** consists of all other users in the system

In UNIX ,the operating system defines 3 files of 3 bits each rwx

r – controls read access

w- controls write access

Password

- The access to each file can be controlled by a password.

Disadvantages

- The number of passwords to be remembered may be large.
- If one password is used for all files, then once

  discovered ,all files will be accessible

File sharing

- Multiple users
- Remote file systems
- Consistency semantics

Multiple users

- Sharing and protection is often implemented by using the concept of owner and group
- The owner has total control over the file
- The group defines a set of users who can share access to the file.

Remote file systems

- FTP

  It is used to transfer the files between machines.

- Distributed File System

  Remote directories can be accessed from a local machine

- World Wide Web

  Browser is used to access to remote files

Consistency semantics

- It is one of the most important criteria of evaluating a file system that supports file sharing.
- They specify how multiple users of a system must access a shared file simultaneously.
- They also specify the changes made to the data by the user will be seen by other users.

- UNIX semantics
- Session semantics
- Immutable-shared files semantics

Unix Semantics

- Writing to an open file by a user is visible immediately to other users
- Since a file has a single image , user process will be delayed.
- Multiple users cannot perform both read and write operations concurrently.

Session semantics

- Writing to an open file by a user is not visible immediately to other users having the same file open.
- Only when the file is closed , the changes made to it are visible later.
- The file can be associated temporarily with several images at the same time. Multiple users can perform both read and write concurrently.

Immutable shared files semantics

- In this approach, once a file is declared as shared by its creator, it cannot be modified.
- Contents cannot be altered.
- The contents of immutable file is fixed. The sharing of file is read-only.

FILE SYSTEM IMPLEMENTATION

CHAPTER - 9

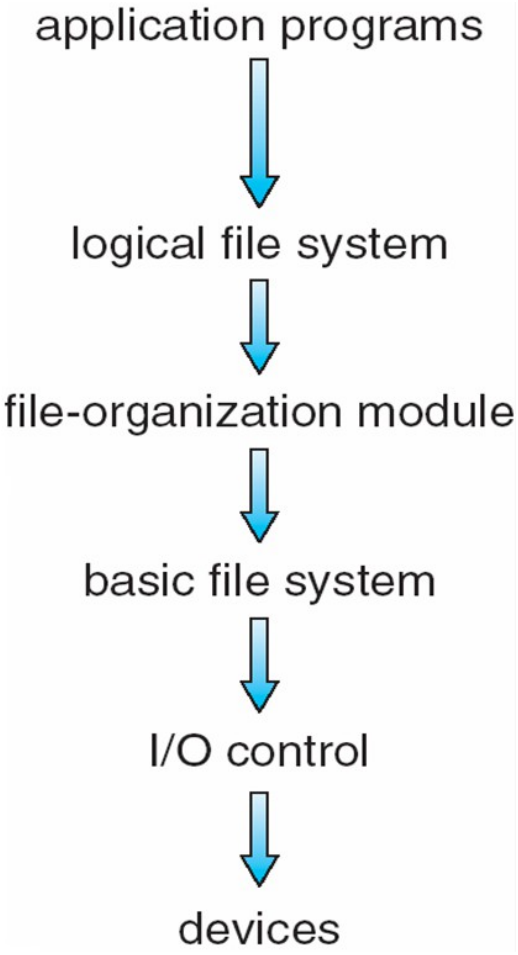
Functions of file system

- An important function of file system is to manage the disk space in a secondary storage.
- Most file systems are stored in the secondary storage, which can hold the amount of data permanently.

The characteristics of disk

- It is possible to read the block from the disk, modify it , rewrite it into the same place.
- It is easy to access any file sequentially and randomly.

Layers of a file system

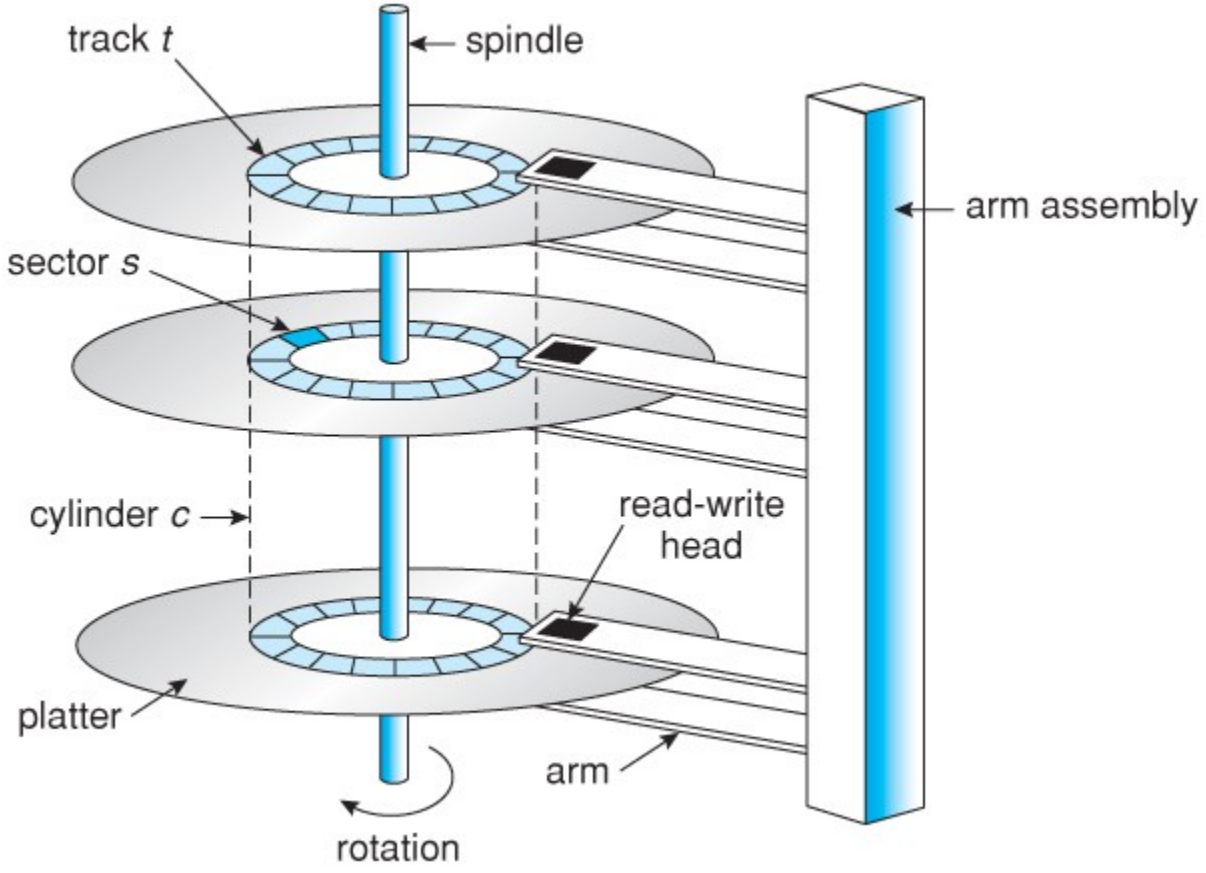


I/O control

- It consists of device drivers and interrupt handlers for transforming information between main memory and disk.
- A device driver is similar to translator. Its input is high level command and output is a low level hardware specific instruction., used to interface the I/O device with the rest of the system.

- Given commands like “read drive1, cylinder 72

Hard disk structure



Basic file system

- The basic file system is required to only issue general commands to the device driver so that it read and writes physical block of the system.
- For example

The address of physical block is drive1, cylinder 73,track 1,sector8.

File organization module

- It contains the information about files – type of the file, location of the file .
- It translates the logical block address to physical block address.
- Logical blocks of every file are numbered from 0 to N .Since the physical block consisting of data do not match the numbers of the logical block. Translation is performed to locate each block.
- This module includes free space manager

Logical file system

- It is used to manage the meta data information
- Meta data information - Information about the file structure and directory structure
- File structure information is stored in FCB (File Control Block) .The FCB is also consists of ownership, permission, File dates, File size,ACL and location of file contents.
- It takes care of file protection and security

Directory implementation

Two algorithms are used for directory implementation

- Linear list
- Hash table

Linear list

- To create a file ,the directory must be searched to make sure that there is no existing file with the same name. Only then a new entry is added at the end of the directory.
- To delete a file, the directory is searched for the given file name and after deleting it, releases the space allocation to it.
- To reuse the directory, one of the following methods can be used

- Searching a file

  linear search

  binary search(sorted list)

Advantages

  simple method

  Easy to implement

Disadvantages

  It is time consuming to execute

Hash table

- It uses the hash data structure to store the directory entries.
- Hash function is used for locating the data in the hash table.
- Insertion and deletion are easy.
- To overcome the problem of collisions, a chained overflow hash table can be used. Each entry in the hash table will be a linked list instead of single value.

Allocation methods

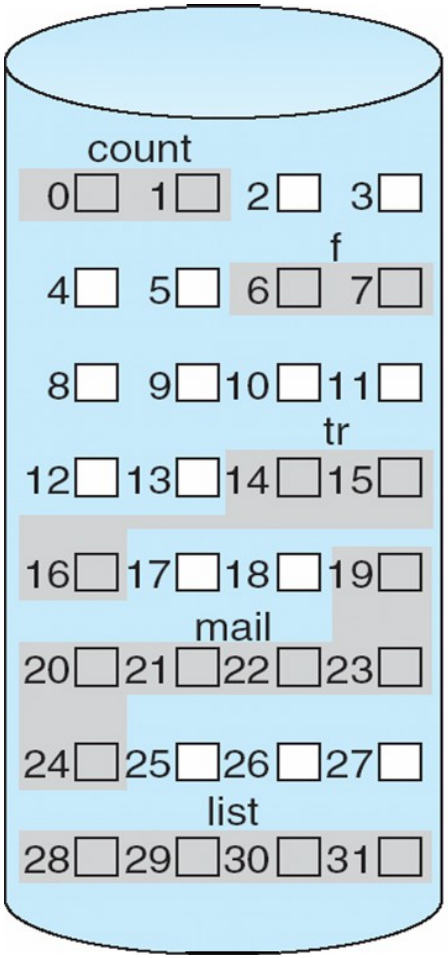
- A file system must keep track of disk blocks allocated to files and free blocks available for allocation.
- The three methods for disk space allocation are

  contiguous allocation

  Linked allocation

  indexed allocation

Contiguous allocation



directory

file	start	length
count	0	2
tr	14	3
mail	19	6
list	28	4
f	6	2

Contiguous allocation

- The file occupies the set of contiguous blocks on the disk.
- The directory entry for a file with contiguous allocation contains

  Address of starting block

  Length of the allocated portion.
- If the size of file n blocks long and starts at location L, then it occupies blocks L,L+1,L+2….L+n-1

Two main problems

- Finding contiguous free blocks of space for a new file
- External fragmentation (free blocks between two files)

Two common strategies for allocation of free blocks

- First fit

Searching is stopped as soon as the first free block big enough to hold the file is encountered.

- Best fit

It searches the whole list and allocates the smallest block to hold the file.

Compaction

It is a technique of combining all the free blocks together into one contiguous block.

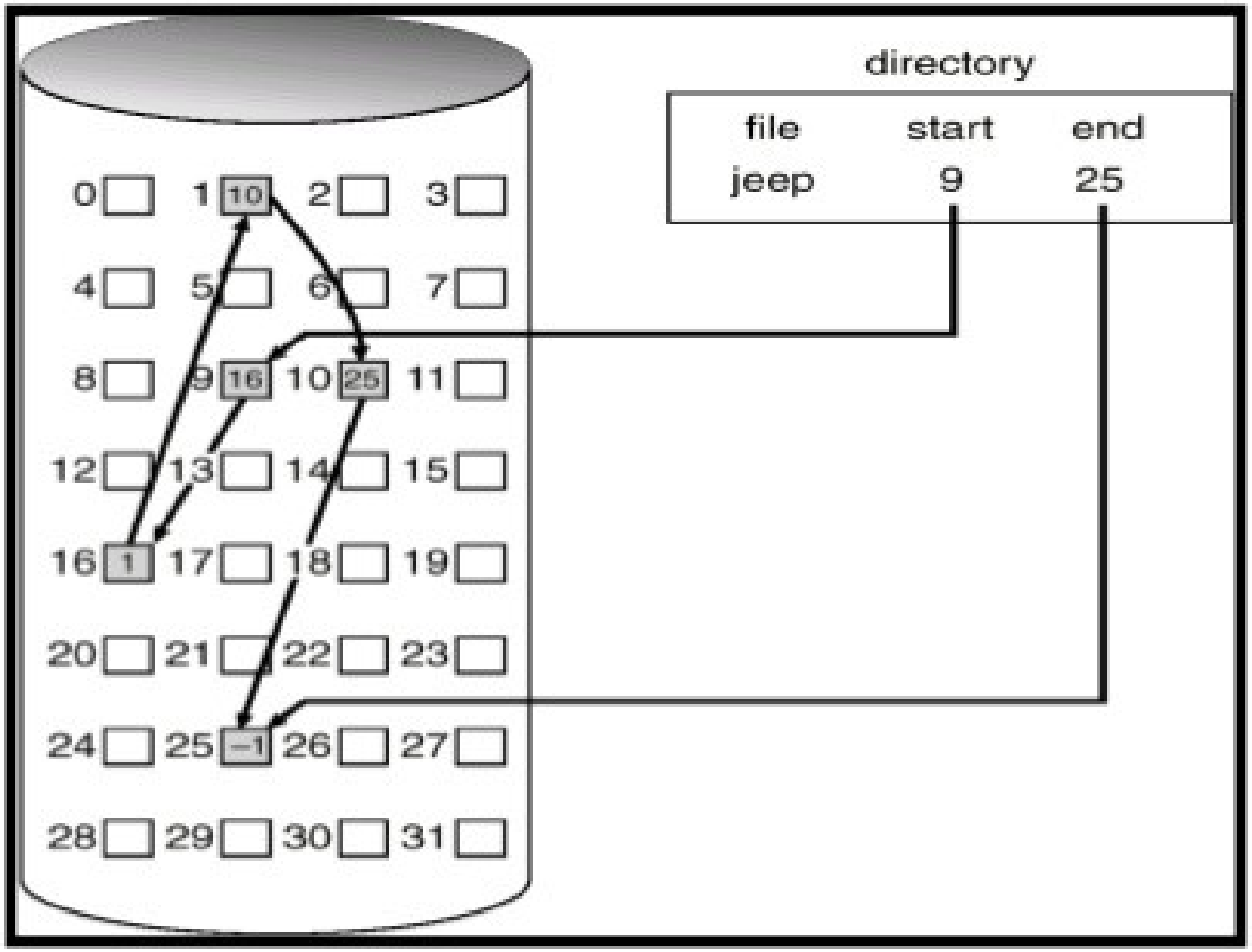
Advantages

- Accessing of files that are allocated continuously.
- Both sequential and direct access are supported.

Disadvantages

- It is difficult to find the contiguous free blocks
- External fragmentation
- When less space is allocated to a file it may be possible that the file cannot be extended
- When more space is allocated ,memory is wasted.

Linked Allocation



Linked allocation

- In linked allocation, each file is linked list of disk blocks. The disk block may be scattered thru out the disk.
- The file directory consists of a pointer to the first and last blocks of the file.
- Creation of a file
- Reading a file

File allocation table

- The table is stored at the beginning of each partition (either c: or D:)
- The table contains one entry for each block
- The table entry also contains a pointer , pointing to the next block.
- The directory entry contains block number of starting block.
- Free blocks are indicated by a 0 value in

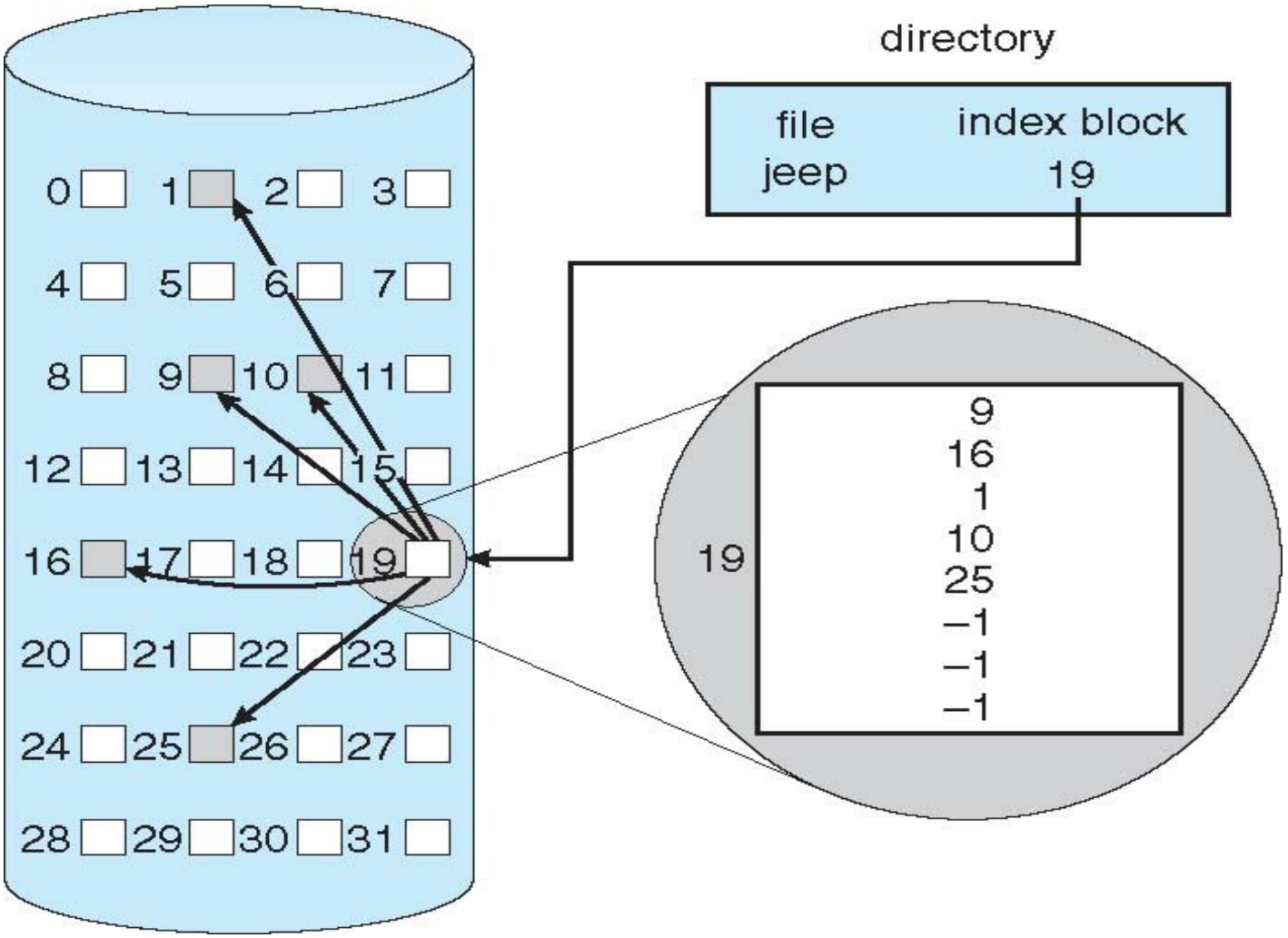
Advantages of linked allocation

- It is simple
- Disk compaction is not necessary
- No external fragmentation
- A file can continue to grow , since there are free blocks are available.

Disadvantages

- Direct accessing of a disk block is slow.
- Space is required for pointers and File allocation table.
- Non reliability – Disk blocks are linked together by a pointer , a single damaged pointer can make thousands of disk blocks inaccessible.

Indexed allocation



Indexed allocation

- Each file has its own index block which is an array of disk block pointers.
- The directory entry contains address of the index block.
- To access the kth block of a file , kth pointer is used.
- Creation of a file.

Schemes in index allocation

- Linked scheme
- Multi-level index
- Combined scheme

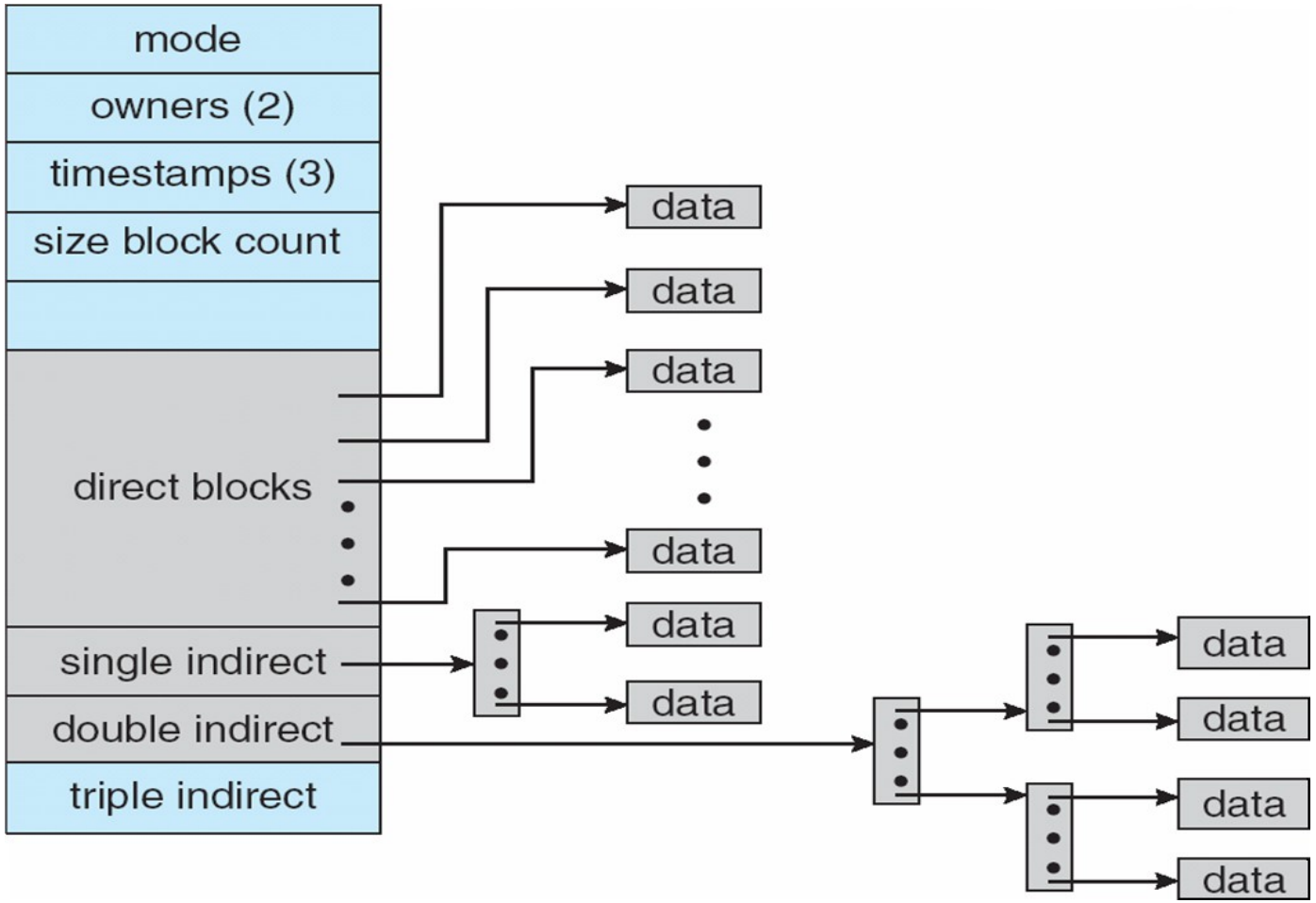
Linked scheme

- This scheme links two or more index blocks together for holding the pointers. Every index block would then contain a pointer or the address to the next index block.

Multi-level index

- In this policy, a first level index block is used to point to the second level index blocks which inturn points to the disk blocks occupied by the file. This can be extended to 3 or more levels depending on the maximum file size.

Combined scheme



Combined scheme

- In this scheme, a special block called the Inode (information Node) contains all the information about the file such as the name, size, authority, etc.
- The first few of these pointers in Inode point to the direct blocks i.e the pointers contain the addresses of the disk blocks that contain data of the file.
- The next few pointers point to indirect blocks. Indirect blocks may be single

Advantages of indexing

- It supports direct accessing
- Does not suffer from external fragmentation

Disadvantages

- The number of disk accesses to get the address of the required block.
- Indexed allocation requires plenty of space to keep pointers.

Free-space management

When a file is deleted, its disk space is added to the free-space list. The methods to manage the free disk blocks are

- Bit vector
- Linked list
- Grouping
- Counting

Bit Vector

- List of free disk space is implemented as a bit map or bit vector. When a block is free, the bit is 0 and when the block is allocated ,the bit is 1. For example

1 1 1 0 1 0 1 0

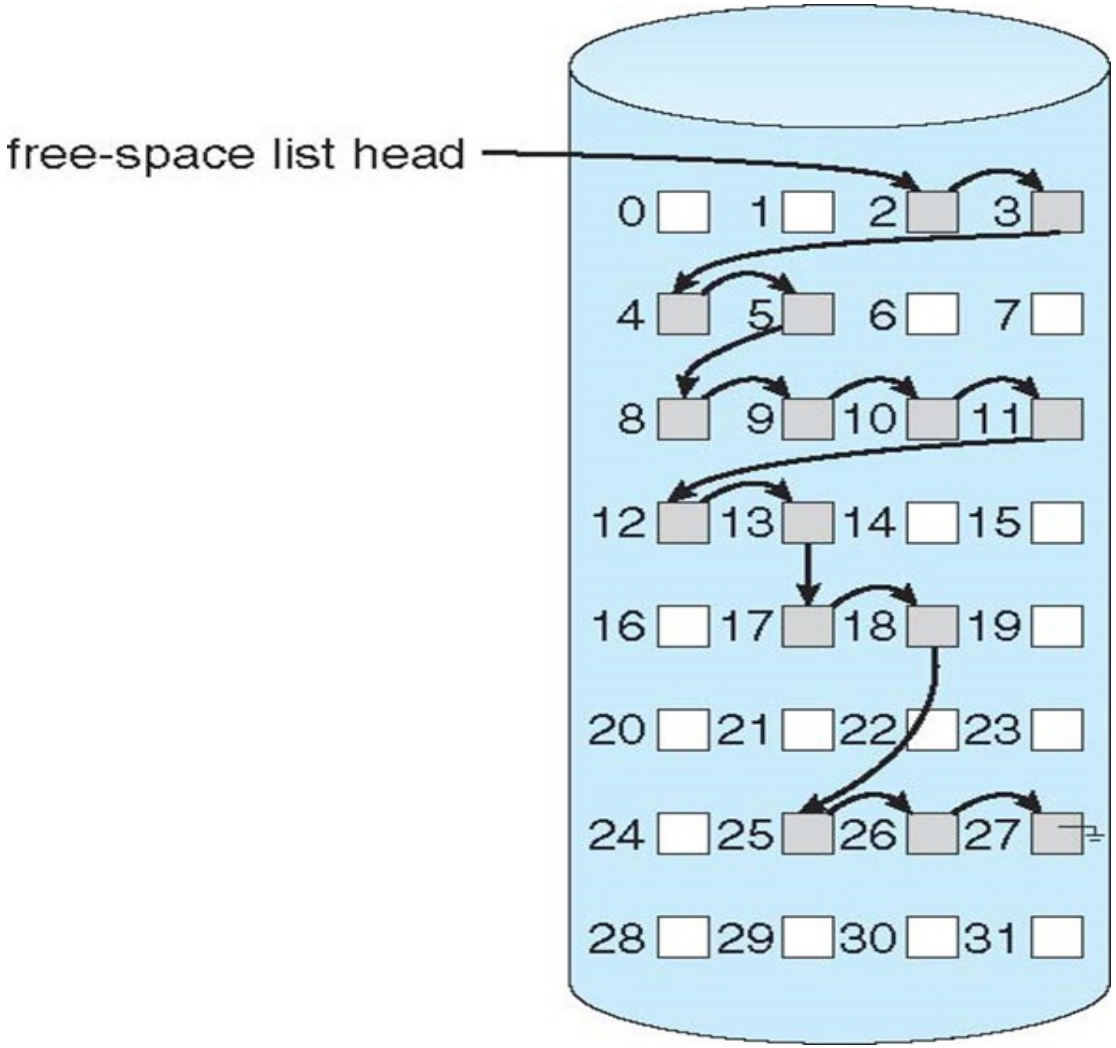
3 5 7

Advantages

- It is simple and efficient.

Disadvantage

Linked list


free-space list head
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

Linked list

- In this approach , all the free blocks are linked together. A pointer pointing to the first free block is kept in the special location on the disk. The first block again contains a pointer, pointing to the next free block and so on.

Grouping

- It stores the address of n free blocks in the first free block. The last block consists of the address of another n free blocks.

counting

- It is used to allocate several contiguous blocks. It consists of address of the first free block and count which gives n free contiguous blocks.

Recovery

- Consistency checking
- Backup and restore

Consistency checking

- Many file system read blocks ,change them and write them later. In case the system crashes before it writes all the modified blocks, the file system can then left in an inconsistent state.
- Frequently, a special program is run at reboot time to check for and correct disk inconsistencies.

- The consistency checker—a systems program such as chkdsk in MS-DOS and fsck in UNIX—compares

- the data in the directory structure with the data blocks on disk and tries to fix any inconsistencies it finds.

Backup and Restore

- system programs can be used to back up data from disk to another storage device, such as a floppy disk, magnetic tape, Compact Disk,or other hard disk.

- **A typical backup schedule may then be as follows:**
-
- Day 1: Copy to a backup medium all files from the disk. This is called a **full backup.**
- Day 2: Copy to another medium all files changed since day 1. This is

DISK MANAGEMENT

CHAPTER - 10

Floppy Disk

- A floppy disk is a magnetic storage medium consisting of a thin flexible circular piece of plastic material coated with magnetic oxide.
- Floppy disk consists of a motor that rotates that disk on a spindle and read/write head that can move to any spot on the disk.
- The disk spins at the speed of about 300 revolutions per Minute

- To store data on a disk, it must be formatted.
- Formatting of a disk creates a set of concrete circles called tracks. The tracks are further divided into sectors.
- A 3 ½ inch floppy disk contains 80 tracks. Each sector stores 512 bytes of data.

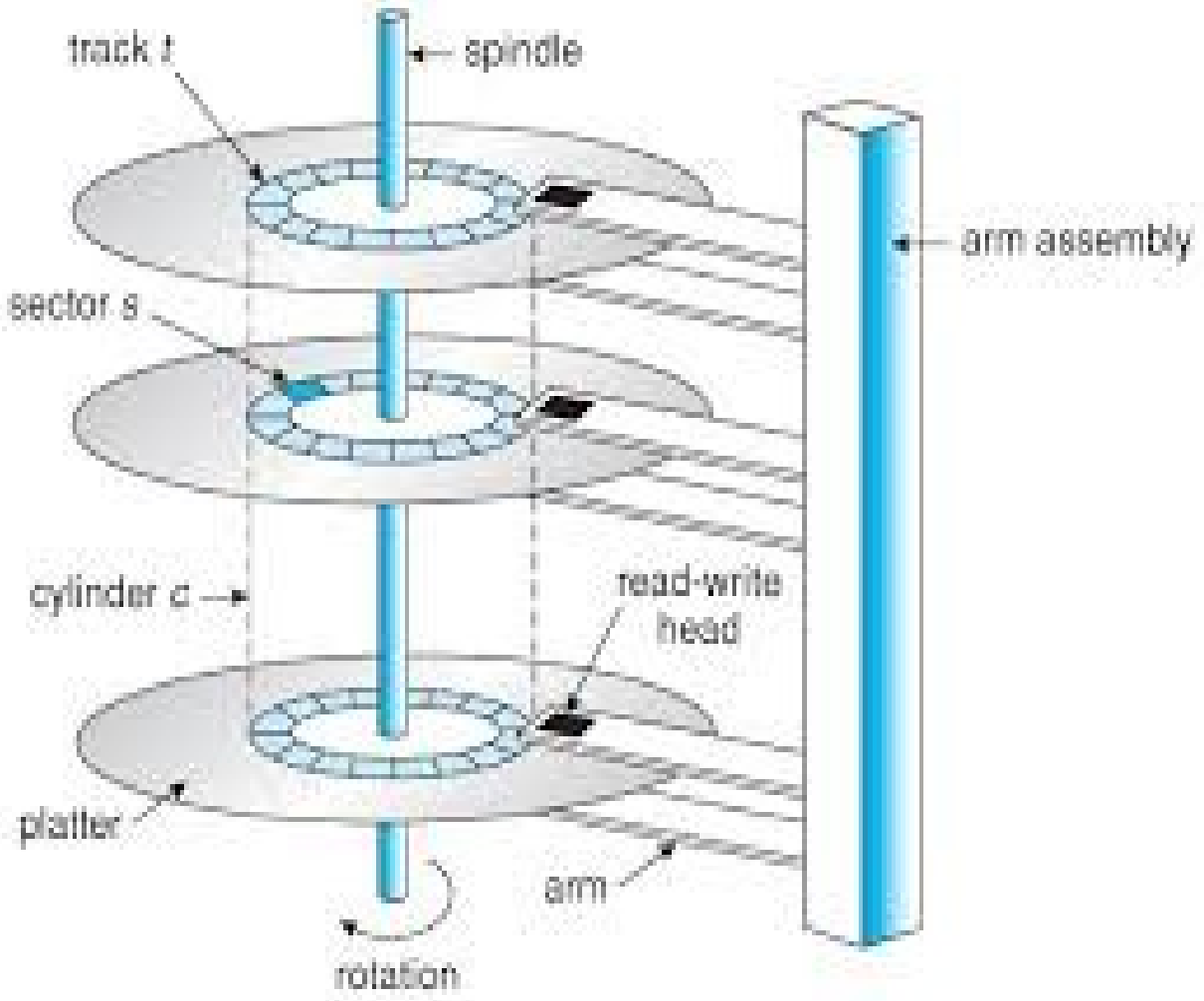
Advantages of floppy disk

- Inexpensive
- Portable
- Easy to use
- It is used to take backup of data

Disadvantages of floppy disk

- Gets corrupted easily
- Access rate is slow
- Small amount of data can be stored.

Hard disk


track t
spindle
arm assembly
sector s
cylinder c
read-write head
platter
arm
rotation

Hard disk

- It is a group of round flat metal platters coated with a magnetic medium.
- Platters are made up of aluminum or mixture of glass and ceramic. To allow the data storage, both sides of each platter are coated with magnetic medium.
- There is a 0.5 inch gap between the platters.
- The read/write head are used on both side of platters to read or write the data to the

- The common spindle that turns the whole assembly at several thousand revolutions per minute.(RPM).
- Small particle or dust on the disk surface will damage the read/write head. So the entire unit is sealed.
- The disk surface is divided into track and the tracks are divided into sectors. A hard disk can be viewed as a series of cylinders ,each of which contains a set of tracks.

- Hard disk comes in two forms 5.25 inch and 3.5 inch.
- The storage capacity of hard disk available today ranges 40 gb to 200 gb

Advantages and disadvantages of hard disk

Advantages

- High storage capacity
- High access time
- Reliable
- Non volatile

Disadvantages

They are not portable

Costly compared to a floppy disk

Magnetic Tapes

- Magnetic tapes are thin tapes coated with magnetic material
- Data is stored on tapes in the form of 1’s and 0’s.
- The data must be accessed sequentially.
- Similar to audio-cassette tape.

Advantages and disadvantages of magnetic tapes

Advantages

Bulk storage medium

Less expensive

Highly reliable and portable

Storage capacity 200 gb to 800 gb

Disadvantages

Data access is very slow

Insertion and deletion of data is not possible

Disk management

- Disk formatting
- Boot block
- Bad-block recovery.

Disk formatting

- To store data on a disk it must be divided into sectors. This process is known as physical formatting or low level formatting.