

## 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. In general, a file is a sequence of bits, bytes, lines or records whose meaning is defined by the files creator and user.

## File Structure

A File Structure should be according to a required format that the operating system can understand.

A file has a certain defined structure according to its type.

A text file is a sequence of characters organized into lines.

A source file is a sequence of procedures and functions.

An object file is a sequence of bytes organized into blocks that are understandable by the machine.

When operating system defines different file structures, it also contains the code to support these file structure. Unix, MS-DOS support minimum number of file structure.

## File Type

File type refers to the ability of the operating system to distinguish different types of file such as text files source files and binary files etc. Many operating systems support many types of files. Operating system like MS-DOS and UNIX have the following types of files -

### Ordinary files

These are the files that contain user information.

These may have text, databases or executable program.

The user can apply various operations on such files like add, modify, delete or even remove the entire file.

### Directory files

These files contain list of file names and other information related to these files.

### Special files

These files are also known as device files.

These files represent physical device like disks, terminals, printers, networks, tape drive etc.

These files are of two types -

Character special files - data is handled character by character as in case of terminals or printers.

Block special files - data is handled in blocks as in the case of disks and tapes.

## File Access Mechanisms

File access mechanism refers to the manner in which the records of a file may be accessed. There are several ways to access files -

### Sequential access

Direct/Random access

Indexed sequential access

Sequential access

A sequential access is that in which the records are accessed in some sequence, i.e., the information in the file is processed in order, one record after the other. This access method is the most primitive one.

Example: Compilers usually access files in this fashion.

Direct/Random access

Random access file organization provides, accessing the records directly.

Each record has its own address on the file with by the help of which it can be directly accessed for reading or writing.

The records need not be in any sequence within the file and they need not be in adjacent locations on the storage medium.

Indexed sequential access

This mechanism is built up on base of sequential access.

An index is created for each file which contains pointers to various blocks.

Index is searched sequentially and its pointer is used to access the file directly.

Space Allocation

Files are allocated disk spaces by operating system. Operating systems deploy following three main ways to allocate disk space to files.

Contiguous Allocation

Linked Allocation

Indexed Allocation

Contiguous Allocation

Each file occupies a contiguous address space on disk.

Assigned disk address is in linear order.

Easy to implement.

External fragmentation is a major issue with this type of allocation technique.

Linked Allocation

Each file carries a list of links to disk blocks.

Directory contains link / pointer to first block of a file.

No external fragmentation

Effectively used in sequential access file.

Inefficient in case of direct access file.

Indexed Allocation

Provides solutions to problems of contiguous and linked allocation.

A index block is created having all pointers to files.

Each file has its own index block which stores the addresses of disk space occupied by the file.

Directory contains the addresses of index blocks of files.

An Operating System provides services to both the users and to the programs.

It provides programs an environment to execute.

It provides users the services to execute the programs in a convenient manner.

Following are a few common services provided by an operating system -

- Program execution
- I/O operations
- File System manipulation
- Communication
- Error Detection
- Resource Allocation
- Protection
- Program execution

Operating systems handle many kinds of activities from user programs to system programs like printer spooler, name servers, file server, etc. Each of these activities is encapsulated as a process.

A process includes the complete execution context (code to execute, data to manipulate, registers, OS resources in use). Following are the major activities of an operating system with respect to program management -

- Loads a program into memory.
- Executes the program.
- Handles program's execution.
- Provides a mechanism for process synchronization.
- Provides a mechanism for process communication.
- Provides a mechanism for deadlock handling.

I/O Operation

An I/O subsystem comprises of I/O devices and their corresponding driver software. Drivers hide the peculiarities of specific hardware devices from the users.

An Operating System manages the communication between user and device drivers.

I/O operation means read or write operation with any file or any specific I/O device.

Operating system provides the access to the required I/O device when required.

File system manipulation

A file represents a collection of related information. Computers can store files on the disk (secondary storage), for long-term storage purpose. Examples of storage media include magnetic tape, magnetic disk and optical disk drives like CD, DVD. Each of these media has its own properties like speed, capacity, data transfer rate and data access methods.

A file system is normally organized into directories for easy navigation and usage. These directories may contain files and other directions. Following are the major activities of an operating system with respect to file management -

Program needs to read a file or write a file.

The operating system gives the permission to the program for operation on file.

Permission varies from read-only, read-write, denied and so on.  
Operating System provides an interface to the user to create/delete files.

Operating System provides an interface to the user to create/delete directories.

Operating System provides an interface to create the backup of file system.

#### Communication

In case of distributed systems which are a collection of processors that do not share memory, peripheral devices, or a clock, the operating system manages communications between all the processes. Multiple processes communicate with one another through communication lines in the network.

The OS handles routing and connection strategies, and the problems of contention and security. Following are the major activities of an operating system with respect to communication -

Two processes often require data to be transferred between them. Both the processes can be on one computer or on different computers, but are connected through a computer network.

Communication may be implemented by two methods, either by Shared Memory or by Message Passing.

#### Error handling

Errors can occur anytime and anywhere. An error may occur in CPU, in I/O devices or in the memory hardware. Following are the major activities of an operating system with respect to error handling -

The OS constantly checks for possible errors.

The OS takes an appropriate action to ensure correct and consistent computing.

#### Resource Management

In case of multi-user or multi-tasking environment, resources such as main memory, CPU cycles and files storage are to be allocated to each user or job. Following are the major activities of an operating system with respect to resource management -

The OS manages all kinds of resources using schedulers.

CPU scheduling algorithms are used for better utilization of CPU.

#### Protection

Considering a computer system having multiple users and concurrent execution of multiple processes, the various processes must be protected from each other's activities.

Protection refers to a mechanism or a way to control the access of programs, processes, or users to the resources defined by a computer system. Following are the major activities of an operating system with respect to protection -

The OS ensures that all access to system resources is controlled.

The OS ensures that external I/O devices are protected from invalid access attempts.

The OS provides authentication features for each user by means of passwords.