Functions of Operating System

An Operating System acts as a communication bridge (interface) between the user and computer hardware. The purpose of an operating system is to provide a platform on which a user can execute programs conveniently and efficiently.
An operating system is a piece of software that manages the allocation of Computer Hardware. The coordination of the hardware must be appropriate to ensure the correct working of the computer system and to prevent user programs from interfering with the proper working of the system.
The main goal of the Operating System is to make the computer environment more convenient to use and the Secondary goal is to use the resources most efficiently.

What is an Operating System?

An operating system is a program that manages a computer's hardware. It also provides a basis for application programs and acts as an intermediary between the computer user and computer hardware. The main task an operating system carries out is the allocation of resources and services, such as the allocation of memory, devices, processors, and information. The operating system also includes programs to manage these resources, such as a traffic controller, a scheduler, a memory management module, I/O programs, and a file system. The operating system simply provides an environment within which other programs can do useful work.

Why are Operating Systems Used?

Operating System is used as a communication channel between the Computer hardware and the user. It works as an intermediate between System Hardware and End-User. Operating System handles the following responsibilities:

- It controls all the computer resources.
- It provides valuable services to user programs.
- It coordinates the execution of user programs.
- It provides resources for user programs.
- It provides an interface (virtual machine) to the user.
- It hides the complexity of software.
- It supports multiple execution modes.
- It monitors the execution of user programs to prevent errors.

Functions of an Operating System

Memory Management

The operating system manages the Primary Memory or Main Memory. Main memory is made up of a large array of bytes or words where each byte or word is assigned a certain address. Main memory is fast storage and it can be accessed directly by the CPU. For a program to be executed, it should be first loaded in the main memory. An operating system manages the allocation and deallocation of memory to various processes and ensures that the other process does not consume the memory allocated to one process. An Operating System performs the following activities for Memory Management:

- It keeps track of primary memory, i.e., which bytes of memory are used by which user program. The memory addresses that have already been allocated and the memory addresses of the memory that has not yet been used.
- In multiprogramming, the OS decides the order in which processes are granted memory access, and for how long.
- It Allocates the memory to a process when the process requests it and deallocates the memory when the process has terminated or is performing an I/O operation.

Processor Management

In a multi-programming environment, the OS decides the order in which processes have access to the processor, and how much processing time each process has. This function of OS is called Process Scheduling. An Operating System performs the following activities for Processor Management.

An operating system manages the processor's work by allocating various jobs to it and ensuring that each process receives enough time from the processor to function properly.

Keeps track of the status of processes. The program which performs this task is known as a traffic controller. Allocates the CPU that is a processor to a process. De-allocates processor when a process is no longer required.

Device Management

An OS manages device communication via its respective drivers. It performs the following activities for device management. Keeps track of all devices connected to the system. designates a program responsible for every device known as the Input/Output controller. Decide which process gets access to a certain device and for how long. Allocates devices effectively and efficiently. Deallocates devices when they are no longer required. There are various input and output devices. an OS controls the working of these input-output devices. It receives the requests from these devices, performs a specific task, and communicates back to the requesting process.

File Management

A file system is organized into directories for efficient or easy navigation and usage. These directories may contain other directories and other files. An Operating System carries out the following file management activities. It keeps track of where information is stored, user access settings, the status of every file, and more. These facilities are collectively known as the file system. An OS keeps track of information regarding the creation, deletion, transfer, copy, and storage of files in an organized way. It also maintains the integrity of the data stored in these files, including the file directory structure, by protecting against unauthorized access.

User Interface or Command Interpreter

The user interacts with the computer system through the operating system. Hence OS acts as an interface between the user and the computer hardware. This user interface is offered through a set of commands or a graphical user interface (GUI). Through this interface, the user makes interacts with the applications and the machine hardware.

Booting the Computer

The process of starting or restarting the computer is known as booting. If the computer is switched off completely and if turned on then it is called cold booting. Warm booting is a process of using the operating system to restart the computer.

Security

The operating system uses password protection to protect user data and similar other techniques. it also prevents unauthorized access to programs and user data. The operating system provides various techniques which assure the integrity and confidentiality of user data. The following security measures are used to protect user data:

- Protection against unauthorized access through login.
- Protection against intrusion by keeping the firewall active.
- Protecting the system memory against malicious access.
- Displaying messages related to system vulnerabilities.

Control Over System Performance

Operating systems play a pivotal role in controlling and optimizing system performance. They act as intermediaries between hardware and software, ensuring that computing resources are efficiently utilized. One fundamental aspect is resource allocation, where the OS allocates CPU time, memory, and I/O devices to different processes, striving to provide fair and optimal resource utilization. Process scheduling, a critical function, helps decide which processes or threads should run when preventing any single task from monopolizing the CPU and enabling effective multitasking.

Job Accounting

The operating system Keeps track of time and resources used by various tasks and users, this information can be used to track resource usage for a particular user or group of users. In a multitasking OS where multiple programs run simultaneously, the OS determines which applications should run in which order and how time should be allocated to each application.

Error-Detecting Aids

The operating system constantly monitors the system to detect errors and avoid malfunctioning computer systems. From time to time, the operating system checks the system for any external threat or malicious software activity. It also checks the hardware for any type of damage. This process displays several alerts to the user so that the appropriate action can be taken against any damage caused to the system.

Coordination Between Other Software and Users

Operating systems also coordinate and assign interpreters, compilers, assemblers, and other software to the various users of the computer systems. In simpler terms, think of the operating system as the traffic cop of your computer. It directs and manages how different software programs can share your computer's resources without causing chaos. It ensures that when you want to use a program, it runs smoothly without crashing or causing problems for others. So, it's like the friendly officer ensuring a smooth flow of traffic on a busy road, making sure everyone gets where they need to go without any accidents or jams.

Performs Basic Computer Tasks

The management of various peripheral devices such as the mouse, keyboard, and printer is carried out by the operating system. Today most operating systems are plug-and-play. These operating systems automatically recognize and configure the devices with no user interference.

Network Management

- **Network Communication:** Think of them as traffic cops for your internet traffic. Operating systems help computers talk to each other and the internet. They manage how data is packaged and sent over the network, making sure it arrives safely and in the right order.
- **Settings and Monitoring:** Think of them as the settings and security guard for your internet connection. They also let you set up your network connections, like Wi-Fi or Ethernet, and keep an eye on how your network is doing. They make sure your computer is using the network efficiently and securely, like adjusting the speed of your internet or protecting your computer from online threats.

Services Provided by an Operating System

The Operating System provides certain services to the users which can be listed in the following manner:

- **User Interface:** Almost all operating systems have a user interface (UI). This interface can take several forms. One is a command-line interface(CLI), which uses text commands and a method for entering them (say, a keyboard for typing in commands in a specific format with specific options). Another is a batch interface, in which commands and directives to control those commands are entered into files, and those files are executed. Most commonly, a graphical user interface (GUI) is used. the interface is a window system with a pointing device to direct I/O, choose from menus, and make selections and a keyboard to enter text.
- **Program Execution:** The Operating System is responsible for the execution of all types of programs whether it be user programs or system programs. The Operating System utilizes various resources available for the efficient running of all types of functionalities.
- **Handling Input/Output Operations:** The Operating System is responsible for handling all sorts of inputs, i.e., from the keyboard, mouse, desktop, etc. The Operating System does all interfacing most appropriately regarding all kinds of Inputs and Outputs.

For example, there is a difference between all types of peripheral devices such as mice or keyboards, the Operating System is responsible for handling data between them.

- **Manipulation of File System:** The Operating System is responsible for making decisions regarding the storage of all types of data or files, i.e., floppy disk/hard disk/pen drive, etc. The Operating System decides how the data should be manipulated and stored.
- **Resource Allocation:** The Operating System ensures the proper use of all the resources available by deciding which resource to be used by whom for how much time. All the decisions are taken by the Operating System.
- **Accounting:** The Operating System tracks an account of all the functionalities taking place in the computer system at a time. All the details such as the types of errors that occurred are recorded by the Operating System.
- **Information and Resource Protection:** The Operating System is responsible for using all the information and resources available on the machine in the most protected way. The Operating System must foil an attempt from any external resource to hamper any sort of data or information.
- **Communication:** The operating system implements communication between one process to another process to exchange information. Such communication may occur between processes that are executing on the same computer or between processes that are executing on different computer systems tied together by a computer network.
- **System Services:** The operating system provides various system services, such as printing, time and date management, and event logging.
- **Error Detection:** The operating system needs to be detecting and correcting errors constantly. Errors may occur in the CPU and memory hardware (for eg. a memory error or a power failure), in I/O devices (such as a parity error on disk, a connection failure on a network, or a lack of paper in the printer), and in the user program (an arithmetic overflow, an attempt to access an illegal memory location or a too-great use of CPU time). For each type of error, the operating system should take the appropriate action to ensure correct and consistent computing.

All these services are ensured by the Operating System for the convenience of the users to make the programming task easier. All different kinds of Operating Systems more or less provide the same services.

Characteristics of Operating System

- **Virtualization:** Operating systems can provide Virtualization capabilities, allowing multiple operating systems or instances of an operating system to run on a single physical machine. This can improve

resource utilization and provide isolation between different operating systems or applications.

- **Networking:** Operating systems provide networking capabilities, allowing the computer system to connect to other systems and devices over a network. This can include features such as network protocols, network interfaces, and network security.
- **Scheduling:** Operating systems provide scheduling algorithms that determine the order in which tasks are executed on the system. These algorithms prioritize tasks based on their resource requirements and other factors to optimize system performance.
- **Interprocess Communication:** Operating systems provide mechanisms for applications to communicate with each other, allowing them to share data and coordinate their activities.
- **Performance Monitoring:** Operating systems provide tools for monitoring system performance, including CPU usage, memory usage, disk usage, and network activity. This can help identify performance bottlenecks and optimize system performance.
- **Backup and Recovery:** Operating systems provide backup and recovery mechanisms to protect data in the event of system failure or data loss.
- **Debugging:** Operating systems provide debugging tools that allow developers to identify and fix software bugs and other issues in the system.

Mainframe Systems – Batch, Multi programmed, Multitasking, Time sharing

Batch Operating system

Batch Operating system is one of the important type of operating system.

The users who using a batch operating system do not interact with the computer directly. Each user prepares its job on an off-line device like punch cards and submits it to the computer operator. To speed up the processing, jobs with similar needs are batched together and run as a group. The programmers exit their programs with the operator and the operator then sorts the programs with similar requirements into batches.

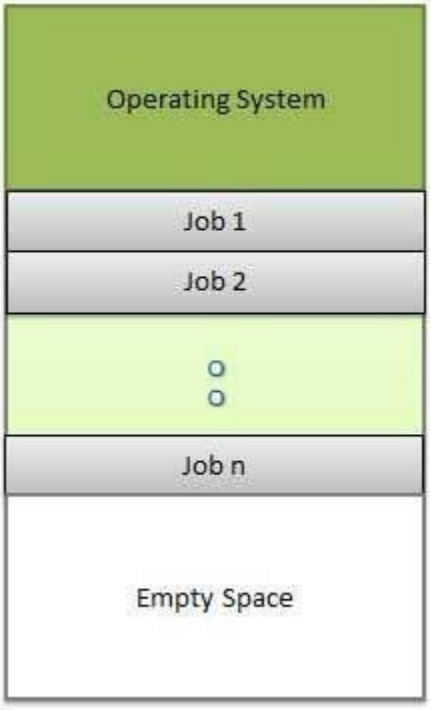
The problems that occurs with Batch Systems are as follows −

- There is a lack of interaction between the user and the job.
- CPU is being often idle, because the speed of the mechanical I/O devices is slower than the CPU.
- It is difficult to provide the desired priority.

Multiprogramming

Sharing the processor, when two or more programs reside in memory at the same time, is referred as **multiprogramming**. Multiprogramming assumes a single shared processor. Multiprogramming increases CPU utilization by organizing jobs so that the CPU always has one to execute.

The following figure shows the memory layout for a multiprogramming system.



An OS does the following activities related to multiprogramming.

- The operating system keeps several jobs in memory at a time.
- This set of jobs is a subset of the jobs kept in the job pool.
- The operating system picks and begins to execute one of the jobs in the memory.
- Multiprogramming operating systems monitor the state of all active programs and system resources using memory management programs to ensures that the CPU is never idle, unless there are no jobs to process.

Advantages

- High and efficient CPU utilization.
- User feels that many programs are allotted CPU almost simultaneously.

Disadvantages

- CPU scheduling is required.
- To accommodate many jobs in memory, memory management is required.

Time-sharing operating systems

Time-sharing is a technique which enables many people, located at various terminals, to use a particular computer system at the same time. Time-sharing or multitasking is a logical extension of multiprogramming. Processor's time which is shared among multiple users simultaneously is termed as time-sharing.

The main difference between Multiprogrammed Batch Systems and Time-Sharing Systems is that in case of Multiprogrammed batch systems, the objective is to maximize processor use, whereas in Time-Sharing Systems, the objective is to minimize response time.

Multiple jobs are executed by the CPU by switching between them, but the switches occur so frequently. Thus, the user can receive an immediate response. For example, in a transaction processing, the processor executes each user program in a short burst or quantum of computation. That is, if **n** users are present, then each user can get a time quantum. When the user submits the command, the response time is in few seconds at most.

The operating system uses CPU scheduling and multiprogramming to provide each user with a small portion of a time. Computer systems that were designed primarily as batch systems have been modified to time-sharing systems.

Advantages of Time sharing operating systems are as follows −

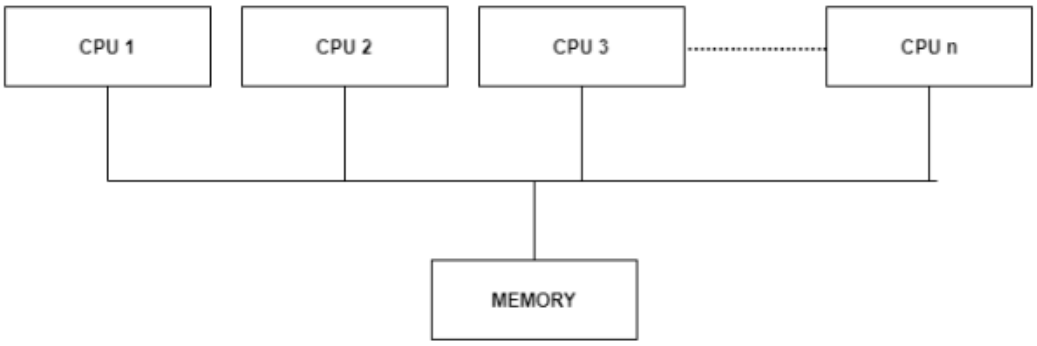
- Provides the advantage of quick response.
- Avoids duplication of software.
- Reduces CPU idle time.

Disadvantages of Time-sharing operating systems are as follows −

- Problem of reliability.
- Question of security and integrity of user programs and data.
- Problem of data communication.

1.3 Multiprocessor Systems

Most computer systems are single processor systems i.e they only have one processor. However, multiprocessor or parallel systems are increasing in importance nowadays. These systems have multiple processors working in parallel that share the computer clock, memory, bus, peripheral devices etc. An image demonstrating the multiprocessor architecture is –



Multiprocessing Architecture

Types of Multiprocessors

There are mainly two types of multiprocessors i.e. symmetric and asymmetric multiprocessors. Details about them are as follows −

Symmetric Multiprocessors

In these types of systems, each processor contains a similar copy of the operating system and they all communicate with each other. All the processors are in a peer to peer relationship i.e. no master - slave relationship exists between them.

An example of the symmetric multiprocessing system is the Encore version of Unix for the Multimax Computer.

Asymmetric Multiprocessors

In asymmetric systems, each processor is given a predefined task. There is a master processor that gives instruction to all the other processors. Asymmetric multiprocessor system contains a master slave relationship.

Asymmetric multiprocessor was the only type of multiprocessor available before symmetric multiprocessors were created. Now also, this is the cheaper option.

Advantages of Multiprocessor Systems

There are multiple advantages to multiprocessor systems. Some of these are −

More reliable Systems

In a multiprocessor system, even if one processor fails, the system will not halt. This ability to continue working despite hardware failure is known as graceful degradation. For example: If there are 5 processors in a multiprocessor system and one of them fails, then also 4 processors are still working. So the system only becomes slower and does not ground to a halt.

Enhanced Throughput

If multiple processors are working in tandem, then the throughput of the system increases i.e. number of processes getting executed per unit of time increase. If there are N processors then the throughput increases by an amount just under N.

More Economic Systems

Multiprocessor systems are cheaper than single processor systems in the long run because they share the data storage, peripheral devices, power supplies etc. If there are multiple processes that share data, it is better to schedule them on multiprocessor systems with shared data than have different computer systems with multiple copies of the data.

Disadvantages of Multiprocessor Systems

There are some disadvantages as well to multiprocessor systems. Some of these are:

Increased Expense

Even though multiprocessor systems are cheaper in the long run than using multiple computer systems, still they are quite expensive. It is much cheaper to buy a simple single processor system than a multiprocessor system.

Complicated Operating System Required

There are multiple processors in a multiprocessor system that share peripherals, memory etc. So, it is much more complicated to schedule processes and impart resources to processes.than in single processor systems. Hence, a more complex and complicated operating system is required in multiprocessor systems.

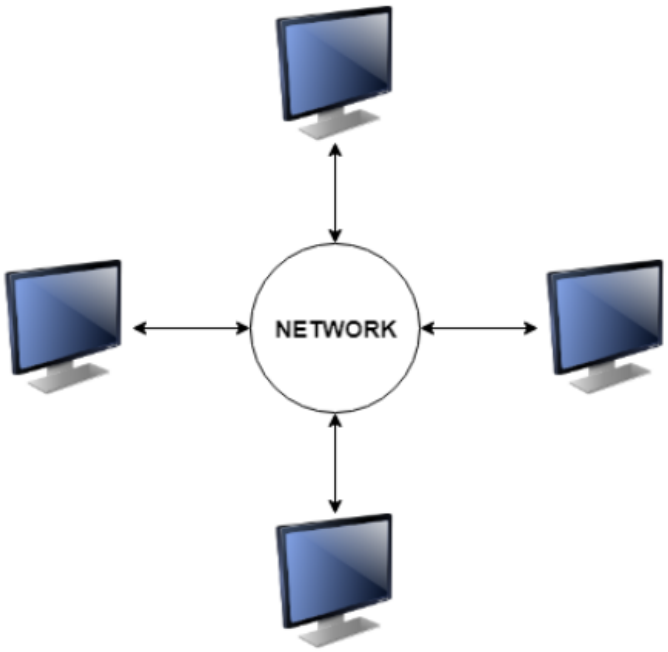
Large Main Memory Required

All the processors in the multiprocessor system share the memory. So a much larger pool of memory is required as compared to single processor systems.

1.4 Distributed Systems.

A distributed system contains multiple nodes that are physically separate but linked together using the network. All the nodes in this system communicate with each other and handle processes in tandem. Each of these nodes contains a small part of the distributed operating system software.

A diagram to better explain the distributed system is −



DISTRIBUTED OPERATING SYSTEM

Types of Distributed Systems

The nodes in the distributed systems can be arranged in the form of client/server systems or peer to peer systems. Details about these are as follows −

Client/Server Systems

In client server systems, the client requests a resource and the server provides that resource. A server may serve multiple clients at the same time while a client is in contact with only one server. Both the client and server usually communicate via a computer network and so they are a part of distributed systems.

Peer to Peer Systems

The peer to peer systems contains nodes that are equal participants in data sharing. All the tasks are equally divided between all the nodes. The nodes interact with each other as required as share resources. This is done with the help of a network.

Advantages of Distributed Systems

Some advantages of Distributed Systems are as follows −

- All the nodes in the distributed system are connected to each other. So nodes can easily share data with other nodes.
- More nodes can easily be added to the distributed system i.e. it can be scaled as required.
- Failure of one node does not lead to the failure of the entire distributed system. Other nodes can still communicate with each other.

- Resources like printers can be shared with multiple nodes rather than being restricted to just one.

Disadvantages of Distributed Systems

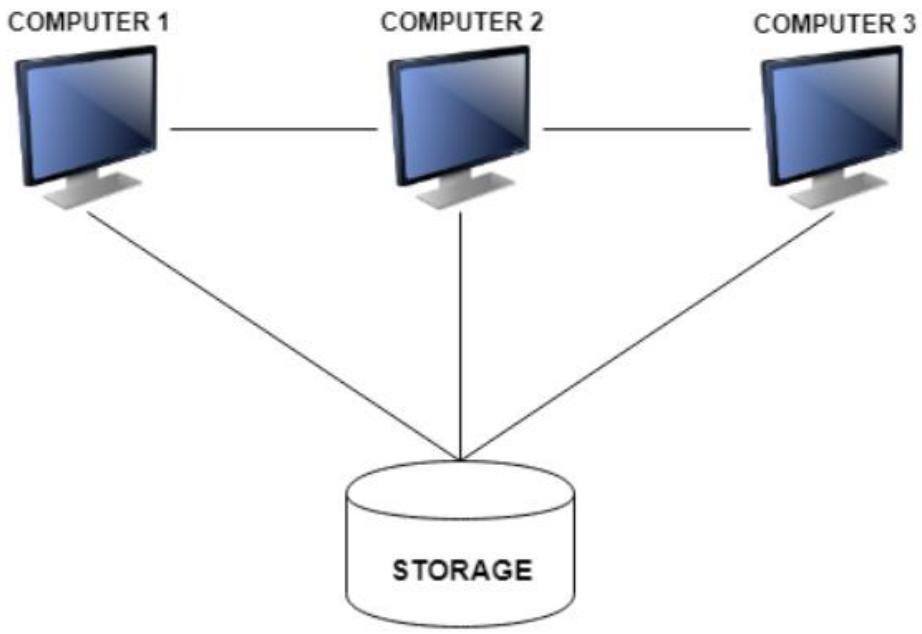
Some disadvantages of Distributed Systems are as follows −

- It is difficult to provide adequate security in distributed systems because the nodes as well as the connections need to be secured.
- Some messages and data can be lost in the network while moving from one node to another.
- The database connected to the distributed systems is quite complicated and difficult to handle as compared to a single user system.
- Overloading may occur in the network if all the nodes of the distributed system try to send data at once.

1.5 Clustered Systems.

Clustered systems are similar to parallel systems as they both have multiple CPUs. However a major difference is that clustered systems are created by two or more individual computer systems merged together. Basically, they have independent computer systems with a common storage and the systems work together.

A diagram to better illustrate this is −



The clustered systems are a combination of hardware clusters and software clusters. The hardware clusters help in sharing of high performance disks between the systems. The software clusters makes all the systems work together .

Each node in the clustered systems contains the cluster software. This software monitors the cluster system and makes sure it is working as required. If any one of the nodes in the clustered system fail, then the rest of the nodes take control of its storage and resources and try to restart.

Types of Clustered Systems

There are primarily two types of clustered systems i.e. asymmetric clustering system and symmetric clustering system. Details about these are given as follows −

Asymmetric Clustering System

In this system, one of the nodes in the clustered system is in hot standby mode and all the others run the required applications. The hot standby mode is a failsafe in which a hot standby node is part of the system . The hot standby node continuously monitors the server and if it fails, the hot standby node takes its place.

Symmetric Clustering System

In symmetric clustering system two or more nodes all run applications as well as monitor each other. This is more efficient than asymmetric system as it uses all the hardware and doesn't keep a node merely as a hot standby.

Attributes of Clustered Systems

There are many different purposes that a clustered system can be used for. Some of these can be scientific calculations, web support etc. The clustering systems that embody some major attributes are −

- **Load Balancing Clusters**

In this type of clusters, the nodes in the system share the workload to provide a better performance. For example: A web based cluster may assign different web queries to different nodes so that the system performance is optimized. Some clustered systems use a round robin mechanism to assign requests to different nodes in the system.

- **High Availability Clusters**

These clusters improve the availability of the clustered system. They have extra nodes which are only used if some of the system components fail. So, high availability clusters remove single points of failure i.e. nodes whose failure leads to the failure of the system. These types of clusters are also known as failover clusters or HA clusters.

Benefits of Clustered Systems

The difference benefits of clustered systems are as follows −

- **Performance**

Clustered systems result in high performance as they contain two or more individual computer systems merged together. These work as a parallel unit and result in much better performance for the system.

- **Fault Tolerance**

Clustered systems are quite fault tolerant and the loss of one node does not result in the loss of the system. They may even contain one or more nodes in hot standby mode which allows them to take the place of failed nodes.

- **Scalability**

Clustered systems are quite scalable as it is easy to add a new node to the system. There is no need to take the entire cluster down to add a new node.

1.6 Real Time Systems.

Real time system means that the system is subjected to real time, i.e., response should be guaranteed within a specified timing constraint or system should meet the specified deadline. For example: flight control system, real time monitors etc.

Types of real time systems based on timing constraints:

1. **Hard real time system –**
   This type of system can never miss its deadline. Missing the deadline may have disastrous consequences. The usefulness of result produced by a hard real time system decreases abruptly and may become negative if tardiness increases. Tardiness means how late a real time system completes its task with respect to its deadline. Example: Flight controller system.

2. **Soft real time system –**
   This type of system can miss its deadline occasionally with some acceptably low probability. Missing the deadline have no disastrous consequences. The usefulness of result produced by a soft real time system decreases gradually with increase in tardiness. Example: Telephone switches.

1.7 Special-Purpose Systems

Our coverage of operating-system issues thus far has focused mainly on general-purpose computing systems. There are, however, special purpose systems with requirements different from those of many of the systems we have described.

A real-time system is a computer system that requires not only those computed results be "correct" but also that the results be produced within a specified deadline period. Results produced after the deadline has passed—even if correct—may be of no real value. For such systems, many traditional operating-system scheduling algorithms must be modified to meet the stringent timing deadlines.

A multimedia system must be able to handle not only conventional data, such as text files, programs, and word-processing documents, but also multimedia data. Multimedia data consist of continuous-media data (audio and video) as well as conventional data. Continuous-media data—such as frames of video—must be delivered according to certain time restrictions (for example, 30 frames per second). The demands of handling continuous-media data require significant changes in operatingsystem structure, most notably in memory, disk, and network management.

1.8 Open-Source Operating System

Open source software is software with source code that anyone can inspect, modify, and enhance.

"Source code" is the part of software that most computer users don't ever see; it's the code computer programmers can manipulate to change how a piece of software—a "program" or "application"—works. Programmers who have access to a computer program's source code can improve that program by adding features to it or fixing parts that don't always work correctly.