# Operating System

Introduction - What Is an Operating System-Operating System Software -A Brief History of Machine Hardware -Types of Operating Systems - Brief History of Operating System Development-Object-Oriented Design- Memory Management-Virtual Memory-Process Management

An Operating System acts as a communication interface between the user and computer hardware. Its purpose is to provide a platform on which a user can execute programs conveniently and efficiently. An operating system is software that manages the allocation of Computer Hardware. The coordination of the hardware must be appropriate to ensure the computer system's correct operation and to prevent user programs from interfering with it. 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. In this article we will see functions of operating system in detail.

**Why 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:

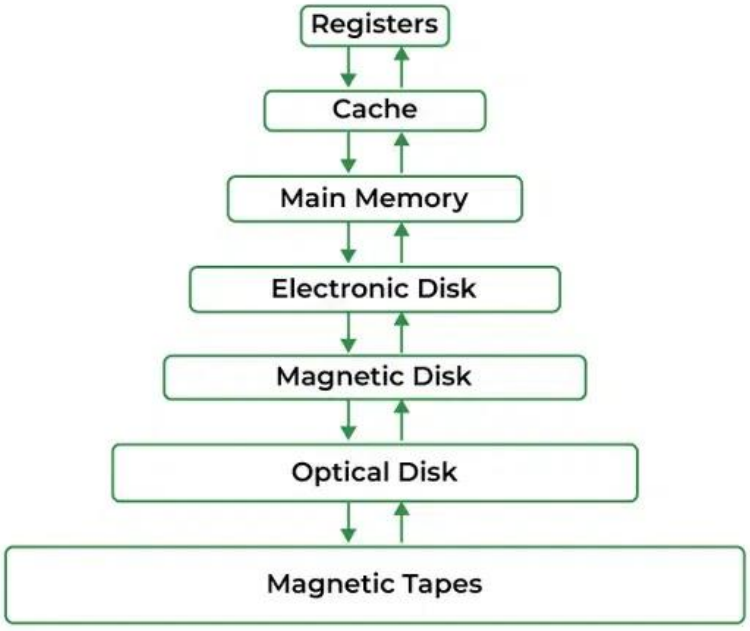
- It controls all the computer resources.
- It provides valuable services to user programs.
- It coordinates the execution of user programs.
- It provides resources for user programs.
- It provides an interface (virtual machine) to the user.
- It hides the complexity of software.
- It supports multiple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



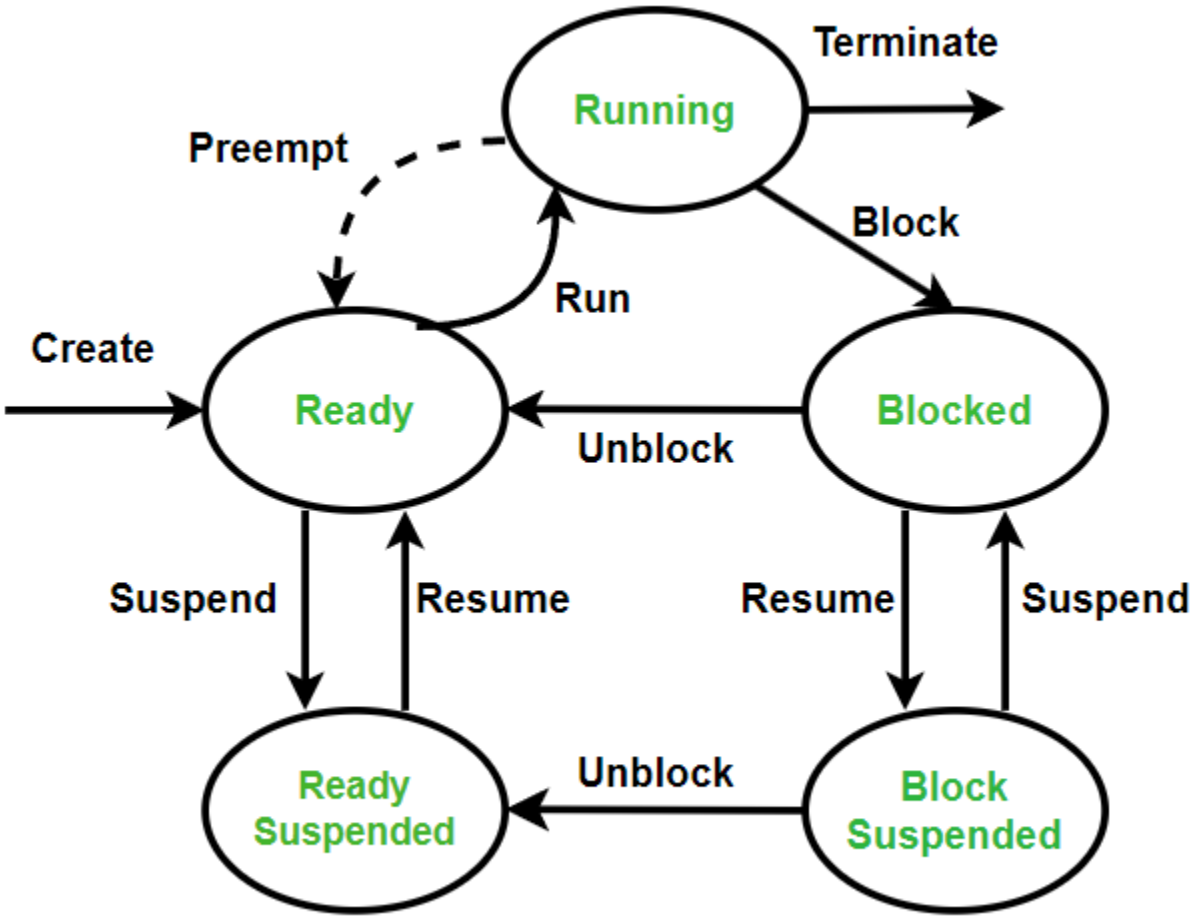

*Memory Management*

**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.




*Process management*

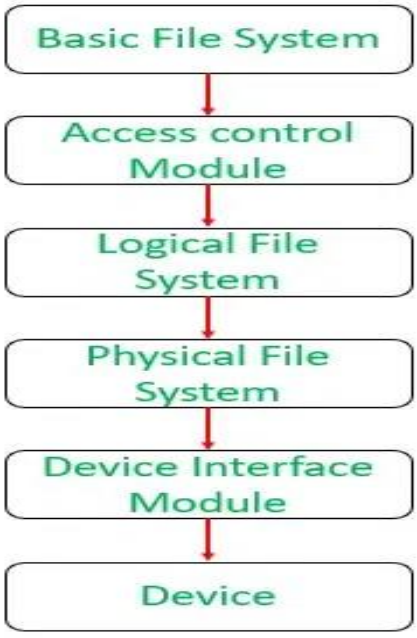
**Device Management**

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.




*File Management*

**I/O Management**

**I/O management** is the important function of operating system refers to how the OS handles **input** and **output** operations between the computer and external devices, such as keyboards, mice, printers, hard drives, and monitors.

**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.


*Command Interpreter*

**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.

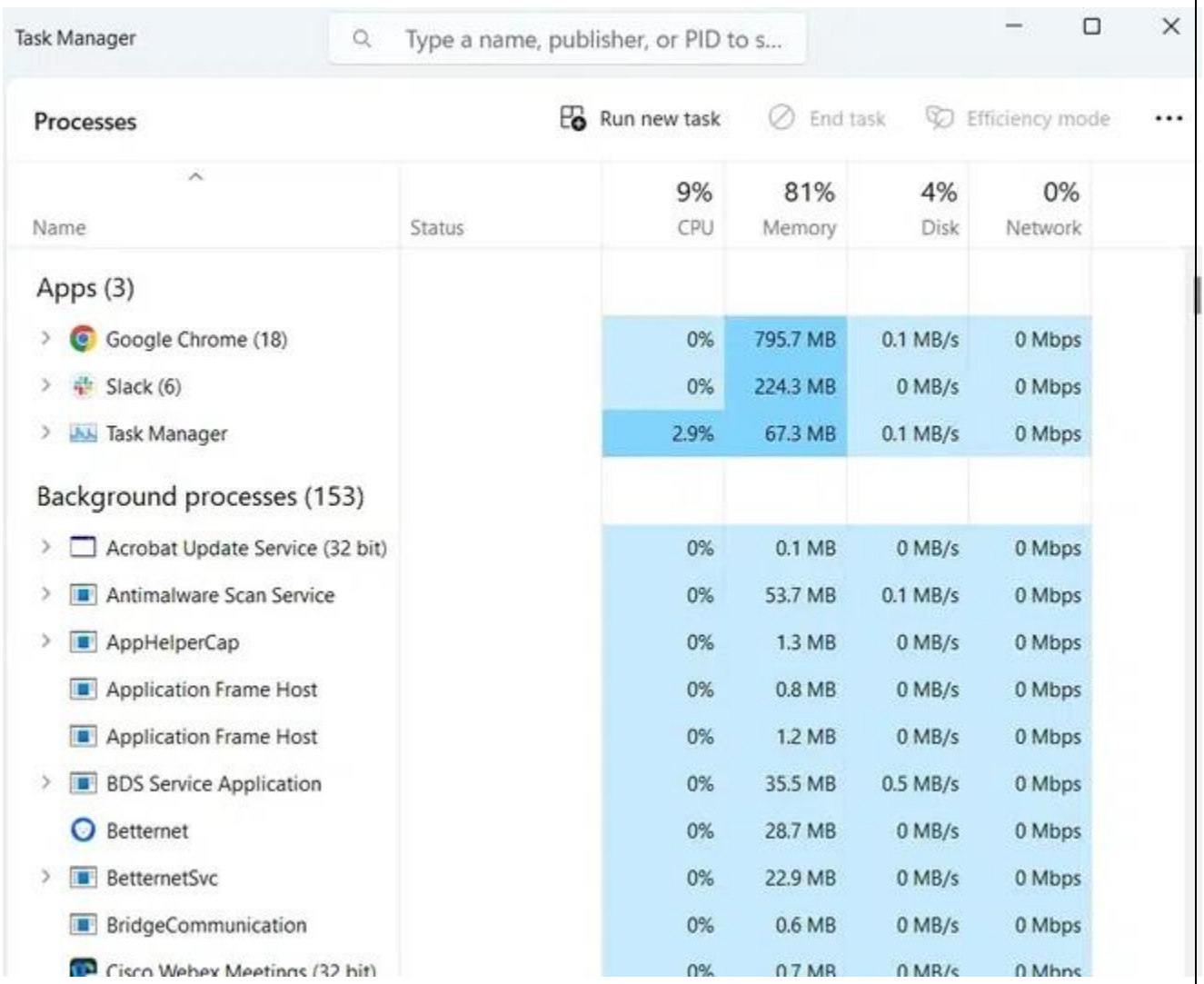
**Security**

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.




*Control Over System Performance*

**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.

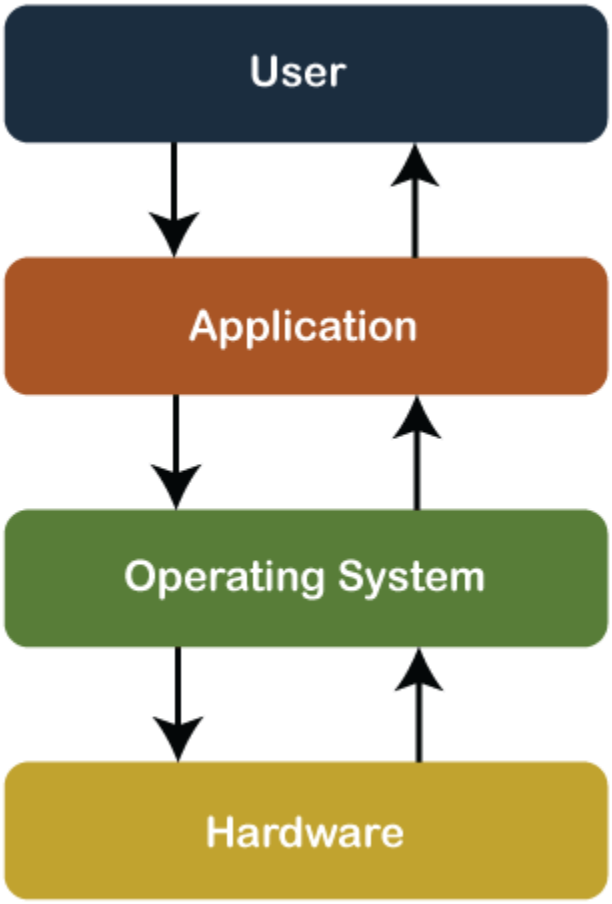
**Characteristics of Operating System**

- **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.

History of the Operating System

An operating is a software program that manages and controls the execution of application programs, software resources and computer hardware. It also helps manage the software/hardware resource, such as file management, memory management, input/ output and many peripheral devices like a disk drive, printers, etc. These are the popular operating system: Linux OS, Windows OS, Mac OS, VMS, OS/400 etc.




Functions of Operating System

- Processor management
- Act as a Resource Manager

- Memory Management
- File Management
- Security
- Device Management
- Input devices / Output devices
- Deadlock Prevention
- Time Management
- Coordinate with system software or hardware

**Types of Operating System**

1. Batch Operating System
2. Time-Sharing Operating System
3. Embedded Operating System
4. Multiprogramming Operating System
5. Network Operating System
6. Distributed Operating System
7. Multiprocessing Operating System
8. Real-Time Operating System

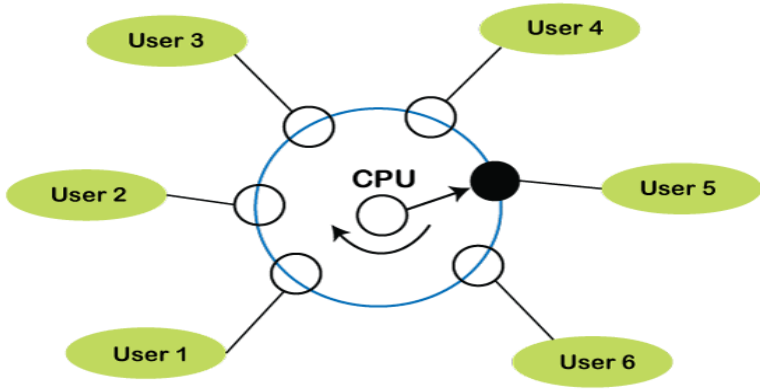
Batch Operating System

In Batch Operating System, there is no direct interaction between user and computer. Therefore, the user needs to prepare jobs and save offline mode to punch card or paper tape or magnetic tape. After creating the jobs, hand it over to the computer operator; then the operator sort or creates the similar types of batches like B2, B3, and B4. Now, the computer operator submits batches into the CPU to execute the jobs one by one. After that, CPUs start executing jobs, and when all jobs are finished, the computer operator provides the output to the user.

Time-Sharing Operating System

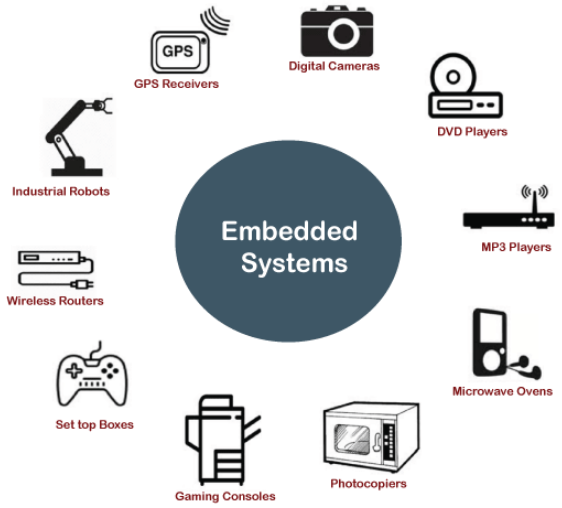
It is the type of operating system that allows us to connect many people located at different locations to share and use a specific system at a single time. The time-sharing operating system is the logical extension of the multiprogramming through which users can run multiple tasks concurrently. Furthermore, it provides each user his terminal for input or output that impacts the program or processor currently running on the system. It represents the CPU's time is shared between many user processes. Or, the processor's time that is shared between multiple users

simultaneously termed as time-sharing.



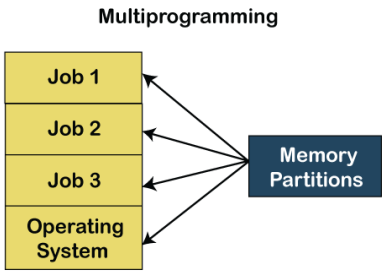

## Embedded Operating System

The Embedded operating system is the specific purpose operating system used in the computer system's embedded hardware configuration. These operating systems are designed to work on dedicated devices like automated teller machines (ATMs), airplane systems, digital home assistants, and the internet of things (IoT) devices.



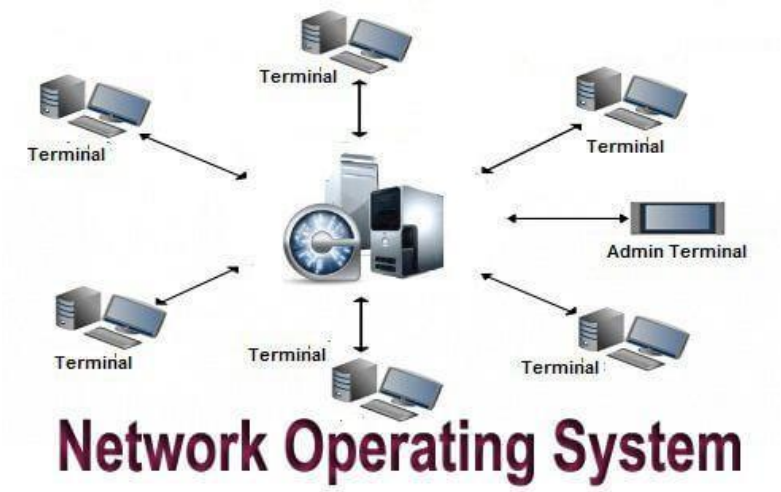

## Multiprogramming Operating System

Due to the CPU's underutilization and the waiting for I/O resource till that CPU remains idle. It shows the improper use of system resources. Hence, the operating system introduces a new concept that is known as multiprogramming. A **multiprogramming operating system** refers to the concepts wherein two or more processes or programs activate simultaneously to execute the processes one after another by the same computer system. When a program is in run mode and uses CPU, another program or file uses I/O resources at the same time or waiting for another system resources to become available. It improves the use of system resources, thereby increasing system throughput. Such a system is known as a multiprogramming operating system.



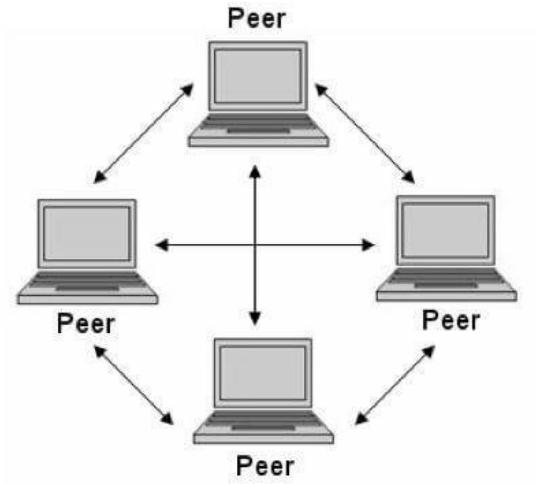

Network Operating System

A network operating system is an important category of the operating system that operates on a server using network devices like a switch, router, or firewall to handle data, applications and other network resources. It provides connectivity among the autonomous operating system, called as a network operating system. The network operating system is also useful to share data, files, hardware devices and printer resources among multiple computers to communicate with each other.

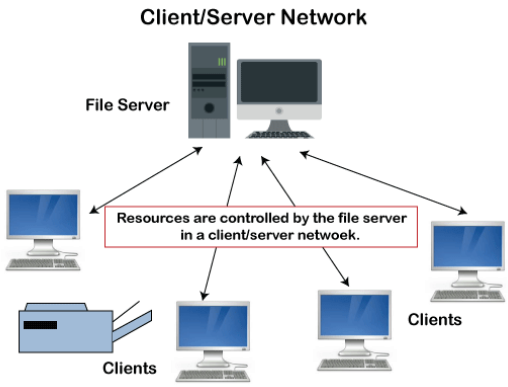



**Types of network operating system**

- **Peer-to-peer network operating system:** The type of network operating system allows users to share files, resources between two or more computer machines using a LAN.

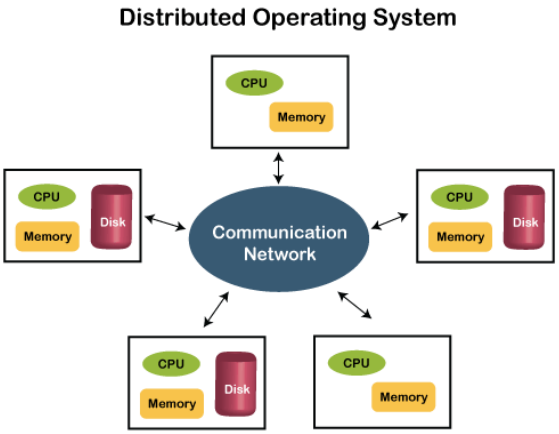



- **Client-Server network operating system:** It is the type of network operating system that allows the users to access resources, functions, and applications through a common server or center hub of the resources. The client workstation can access all resources that exist in the central hub of the network. Multiple clients can access and share different types of the resource over the network from different locations.




Distributed Operating system

A distributed operating system provides an environment in which multiple independent CPU or processor communicates with each other through physically separate computational nodes. Each node contains specific software that communicates with the global aggregate operating system. With the ease of a distributed system, the programmer or developer can easily access any operating system and resource to execute the computational tasks and achieve a common goal. It is the extension of a network operating system that facilitates a high degree of connectivity to communicate with other users over the network.




Multiprocessing Operating System

It is the type of operating system that refers to using two or more central processing units (CPU) in a single computer system. However, these multiprocessor systems or parallel operating systems are used to increase the computer system's efficiency. With the use of a multiprocessor system, they share computer bus, clock, memory and input or output device for concurrent execution of process or program and resource management in the CPU.

Real-Time Operating System

A real-time operating system is an important type of operating system used to provide services and data processing resources for applications in which the time interval required to process & respond to input/output should be so small without any delay real-time system. For example, real-life situations governing an automatic car, traffic signal, nuclear reactor or an aircraft require an immediate response to complete tasks within a specified time delay. Hence, a real-time operating system must be fast and responsive for an embedded system, weapon system, robots, scientific research & experiments and various real-time objects.

Types of the real-time operating system:

- **Hard Real-Time System**
  These types of OS are used with those required to complete critical tasks within the defined time limit. If the response time is high, it is not accepted by the system or may face serious issues like a system failure. In a hard real-time system, the secondary storage is either limited or missing, so these system stored data in the ROM.
- **Soft Real-Time System**
  A soft real-time system is a less restrictive system that can accept software and hardware resources delays by the operating system. In a soft real-time system, a critical task prioritizes less important tasks, and that priority retains active until completion of the task. Also, a time limit is set for a specific job, which enables short time delays for further tasks that are acceptable. For example, computer audio or video, virtual reality, reservation system, projects like undersea, etc.

HISTORY OF OS

Generations of Operating System

**The First Generation (1940 to early 1950s)**

When the first electronic computer was developed in 1940, it was created without any operating system. In early times, users have full access to the computer machine and write a program for each task in absolute machine language. The programmer can perform and solve only simple mathematical calculations during the computer generation, and this calculation does not require an operating system.

**The Second Generation (1955 - 1965)**

The first operating system (OS) was created in the early 1950s and was known as **GMOS. General Motors** has developed OS for the **IBM** computer. The second-generation operating system was based on a single stream batch processing system because it collects all similar jobs in groups or batches and then submits the jobs to the operating system using a punch card to complete all jobs in a machine. At each completion of jobs (either normally or abnormally), control transfer to the operating system that is cleaned after completing one job and then continues to read and initiates the next job in a punch card. After that, new machines were called mainframes, which were very big and used by professional operators.

**The Third Generation (1965 - 1980)**

During the late 1960s, operating system designers were very capable of developing a new operating system that could simultaneously perform multiple tasks in a single computer program called multiprogramming. The introduction of **multiprogramming** plays a very important role in developing operating systems that allow a CPU to be busy every time by performing different tasks on a computer at the same time. During the third generation, there was a new development of minicomputer's phenomenal growth starting in 1961 with the DEC PDP-1. These PDP's leads to the creation of personal computers in the fourth generation.

**The Fourth Generation (1980 - Present Day)**

The fourth generation of operating systems is related to the development of the personal computer. However, the personal computer is very similar to the minicomputers that were developed in the third generation. The cost of a personal computer was very high at that time; there were small fractions of minicomputers costs. A major factor related to creating personal computers was the birth of Microsoft and the Windows operating system. Microsoft created the first **window** operating system in 1975. After introducing the Microsoft Windows OS, Bill Gates and Paul Allen had the vision to take personal computers to the next level. Therefore, they introduced the **MS-DOS** in 1981; however, it was very difficult for the person to understand its cryptic commands. Today, Windows has become the most popular and most commonly used operating system technology. And then, Windows released various operating systems such as Windows 95, Windows 98, Windows XP and the latest operating system, Windows 7. Currently, most Windows users use the Windows 10 operating system. Besides the Windows operating system, Apple is another popular operating system built in the 1980s, and this operating system was developed by Steve Jobs, a co-founder of Apple. They named the operating system Macintosh OS or Mac OS.

Advantages of Operating System

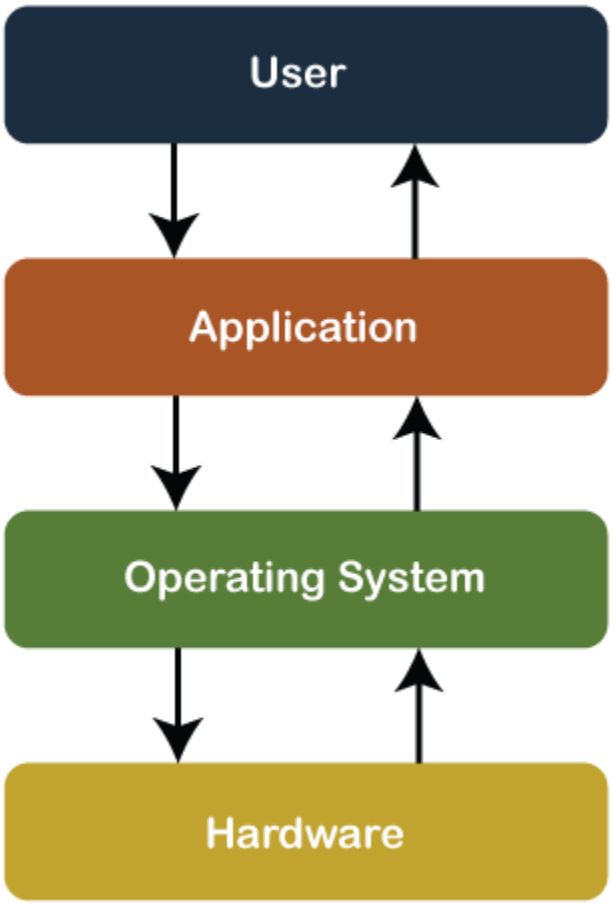
- It is helpful to monitor and regulate resources.
- It can easily operate since it has a basic graphical user interface to communicate with your device.
- It is used to create interaction between the users and the computer application or hardware.
- The performance of the computer system is based on the CPU.
- The response time and throughput time of any process or program are fast.
- It can share different resources like fax, printer, etc.
- It also offers a forum for various types of applications like system and web application.

Disadvantage of the Operating System

- It allows only a few tasks that can run at the same time.
- It any error occurred in the operating system; the stored data can be destroyed.
- It is a very difficult task or works for the OS to provide entire security from the viruses because any threat or virus can occur at any time in a system.
- An unknown user can easily use any system without the permission of the original user.
- The cost of operating system costs is very high.

History of the Operating System

An operating is a software program that manages and controls the execution of application programs, software resources and computer hardware. It also helps manage the software/hardware resource, such as file management, memory management, input/ output and many peripheral devices like a disk drive, printers, etc. These are the popular operating system: Linux OS, Windows OS, Mac OS, VMS, OS/400 etc.




Functions of Operating System

- Processor management
- Act as a Resource Manager
- Memory Management
- File Management
- Security
- Device Management
- Input devices / Output devices
- Deadlock Prevention

- Time Management
- Coordinate with system software or hardware

**Types of Operating System**

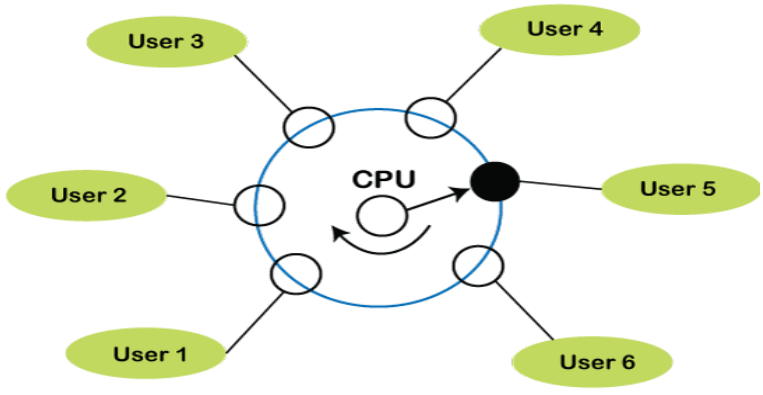
1. Batch Operating System
2. Time-Sharing Operating System
3. Embedded Operating System
4. Multiprogramming Operating System
5. Network Operating System
6. Distributed Operating System
7. Multiprocessing Operating System
8. Real-Time Operating System

Batch Operating System

In Batch Operating System, there is no direct interaction between user and computer. Therefore, the user needs to prepare jobs and save offline mode to punch card or paper tape or magnetic tape. After creating the jobs, hand it over to the computer operator; then the operator sort or creates the similar types of batches like B2, B3, and B4. Now, the computer operator submits batches into the CPU to execute the jobs one by one. After that, CPUs start executing jobs, and when all jobs are finished, the computer operator provides the output to the user.

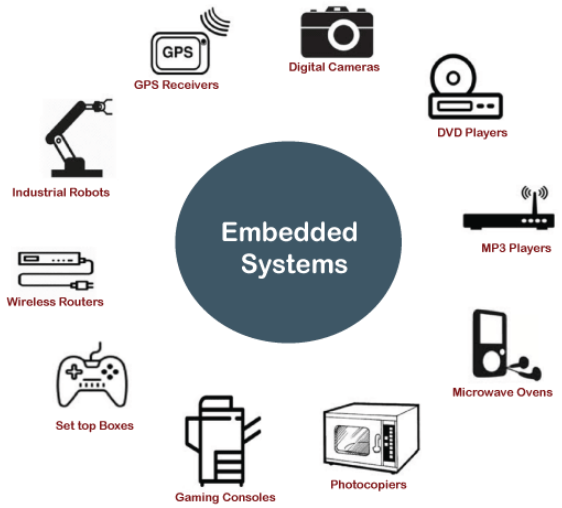
Time-Sharing Operating System

It is the type of operating system that allows us to connect many people located at different locations to share and use a specific system at a single time. The time-sharing operating system is the logical extension of the multiprogramming through which users can run multiple tasks concurrently. Furthermore, it provides each user his terminal for input or output that impacts the program or processor currently running on the system. It represents the CPU's time is shared between many user processes. Or, the processor's time that is shared between multiple users simultaneously termed as time-sharing.



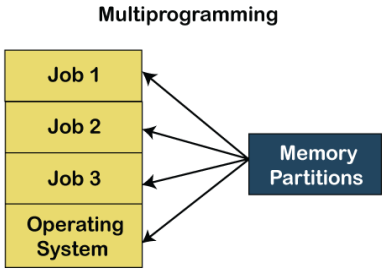

Embedded Operating System

The Embedded operating system is the specific purpose operating system used in the computer system's embedded hardware configuration. These operating systems are designed to work on dedicated devices like automated teller machines (ATMs), airplane systems, digital home assistants, and the internet of things (IoT) devices.




## Multiprogramming Operating System

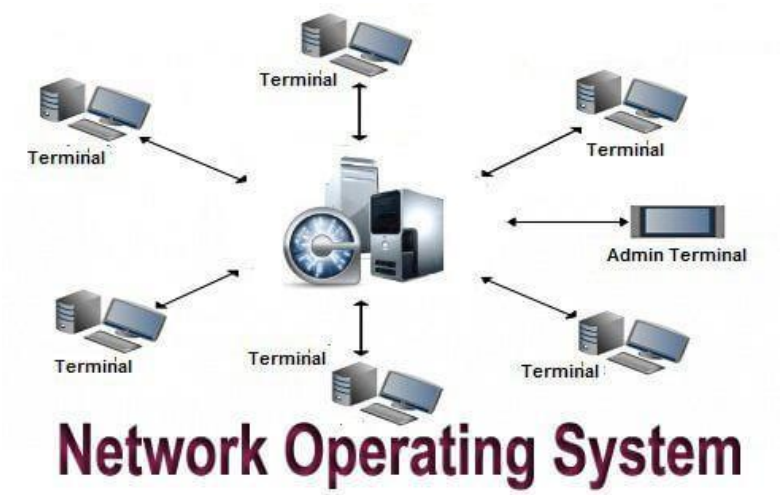
Due to the CPU's underutilization and the waiting for I/O resource till that CPU remains idle. It shows the improper use of system resources. Hence, the operating system introduces a new concept that is known as multiprogramming. A **multiprogramming operating system** refers to the concepts wherein two or more processes or programs activate simultaneously to execute the processes one after another by the same computer system. When a program is in run mode and uses CPU, another program or file uses I/O resources at the same time or waiting for another system resources to become available. It improves the use of system resources, thereby increasing system throughput. Such a system is known as a multiprogramming operating system.




## Network Operating System

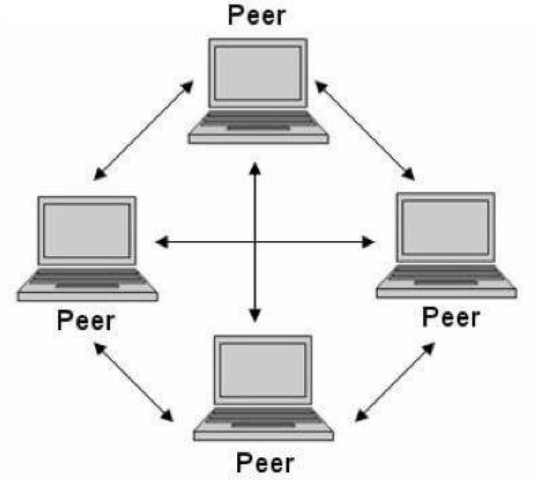
A network operating system is an important category of the operating system that operates on a server using network devices like a switch, router, or firewall to handle data, applications and other network resources. It provides connectivity among the autonomous operating system, called as a network operating system. The network operating system is also useful to share data, files,

hardware devices and printer resources among multiple computers to communicate with each other.



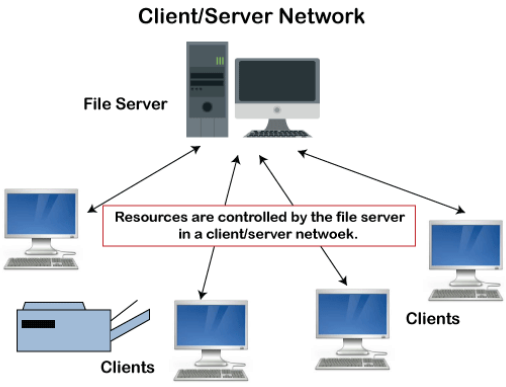

**Types of network operating system**

- **Peer-to-peer network operating system:** The type of network operating system allows users to share files, resources between two or more computer machines using a LAN.



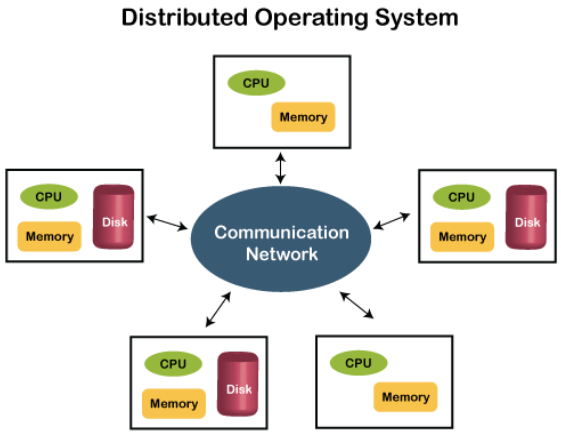

- **Client-Server network operating system:** It is the type of network operating system that allows the users to access resources, functions, and applications through a common server or center hub of the resources. The client workstation can access all resources that exist in the central hub of the network. Multiple clients can access and share different types of the

resource over the network from different locations.




Distributed Operating system

A distributed operating system provides an environment in which multiple independent CPU or processor communicates with each other through physically separate computational nodes. Each node contains specific software that communicates with the global aggregate operating system. With the ease of a distributed system, the programmer or developer can easily access any operating system and resource to execute the computational tasks and achieve a common goal. It is the extension of a network operating system that facilitates a high degree of connectivity to communicate with other users over the network.




Multiprocessing Operating System

It is the type of operating system that refers to using two or more central processing units (CPU) in a single computer system. However, these multiprocessor systems or parallel operating systems are used to increase the computer system's efficiency. With the use of a multiprocessor system, they share computer bus, clock, memory and input or output device for concurrent execution of process or program and resource management in the CPU.

Real-Time Operating System

A real-time operating system is an important type of operating system used to provide services and data processing resources for applications in which the time interval required to process & respond to input/output should be so small without any delay real-time system. For example, real-life situations governing an automatic car, traffic signal, nuclear reactor or an aircraft require an immediate response to complete tasks within a specified time delay. Hence, a real-time operating system must be fast and responsive for an embedded system, weapon system, robots, scientific research & experiments and various real-time objects.

Types of the real-time operating system:

- **Hard Real-Time System**
  These types of OS are used with those required to complete critical tasks within the defined time limit. If the response time is high, it is not accepted by the system or may face serious issues like a system failure. In a hard real-time system, the secondary storage is either limited or missing, so these system stored data in the ROM.
- **Soft Real-Time System**
  A soft real-time system is a less restrictive system that can accept software and hardware resources delays by the operating system. In a soft real-time system, a critical task prioritizes less important tasks, and that priority retains active until completion of the task. Also, a time limit is set for a specific job, which enables short time delays for further tasks that are acceptable. For example, computer audio or video, virtual reality, reservation system, projects like undersea, etc.

HISTORY OF OS

Generations of Operating System

**The First Generation (1940 to early 1950s)**

When the first electronic computer was developed in 1940, it was created without any operating system. In early times, users have full access to the computer machine and write a program for each task in absolute machine language. The programmer can perform and solve only simple mathematical calculations during the computer generation, and this calculation does not require an operating system.

**The Second Generation (1955 - 1965)**

The first operating system (OS) was created in the early 1950s and was known as **GMOS. General Motors** has developed OS for the **IBM** computer. The second-generation operating system was based on a single stream batch processing system because it collects all similar jobs in groups or batches and then submits the jobs to the operating system using a punch card to complete all jobs in a machine. At each completion of jobs (either normally or abnormally), control transfer to the operating system that is cleaned after completing one job and then continues to read and initiates the next job in a punch card. After that, new machines were called mainframes, which were very big and used by professional operators.

**The Third Generation (1965 - 1980)**

During the late 1960s, operating system designers were very capable of developing a new operating system that could simultaneously perform multiple tasks in a single computer program called multiprogramming. The introduction of **multiprogramming** plays a very important role in developing operating systems that allow a CPU to be busy every time by performing different tasks on a computer at the same time. During the third generation, there was a new development of minicomputer's phenomenal growth starting in 1961 with the DEC PDP-1. These PDP's leads to the creation of personal computers in the fourth generation.

**The Fourth Generation (1980 - Present Day)**

The fourth generation of operating systems is related to the development of the personal computer. However, the personal computer is very similar to the minicomputers that were developed in the third generation. The cost of a personal computer was very high at that time; there were small fractions of minicomputers costs. A major factor related to creating personal computers was the birth of Microsoft and the Windows operating system. Microsoft created the first **window** operating system in 1975. After introducing the Microsoft Windows OS, Bill Gates and Paul Allen had the vision to take personal computers to the next level. Therefore, they introduced the **MS-DOS** in 1981; however, it was very difficult for the person to understand its cryptic commands. Today, Windows has become the most popular and most commonly used operating system technology. And then, Windows released various operating systems such as Windows 95, Windows 98, Windows XP and the latest operating system, Windows 7. Currently, most Windows users use the Windows 10 operating system. Besides the Windows operating system, Apple is another popular operating system built in the 1980s, and this operating system

was developed by Steve Jobs, a co-founder of Apple. They named the operating system Macintosh OS or Mac OS.

Advantages of Operating System

- It is helpful to monitor and regulate resources.
- It can easily operate since it has a basic graphical user interface to communicate with your device.
- It is used to create interaction between the users and the computer application or hardware.
- The performance of the computer system is based on the CPU.
- The response time and throughput time of any process or program are fast.
- It can share different resources like fax, printer, etc.
- It also offers a forum for various types of applications like system and web application.

Disadvantage of the Operating System

- It allows only a few tasks that can run at the same time.
- It any error occurred in the operating system; the stored data can be destroyed.
- It is a very difficult task or works for the OS to provide entire security from the viruses because any threat or virus can occur at any time in a system.
- An unknown user can easily use any system without the permission of the original user.
- The cost of operating system costs is very high.

What is Memory Management?

In a multiprogramming computer, the Operating System resides in a part of memory, and the rest is used by multiple processes. The task of subdividing the memory among different processes is called Memory Management.

Main memory is associated with the processor, so moving instructions and information into and out of the processor is extremely fast. Main memory is also known as RAM (Random Access Memory). This memory is volatile. RAM loses its data when a power interruption occurs.

The main aim of memory management is to achieve efficient utilization of memory.

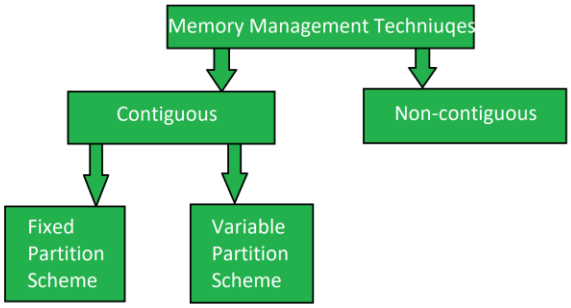
Memory Management Techniques are classified broadly into two categories:

- Contiguous
- Non-contiguous

**What is Contiguous Memory Management?**

contiguous memory allocation is a memory allocation strategy. As the name implies, we utilize this technique to assign contiguous blocks of memory to each task. Thus, whenever a process asks to access the main memory, we allocate a continuous segment from the empty region to the process based on its size. In this technique, memory is allotted in a continuous way to the processes. Contiguous Memory Management has two types:

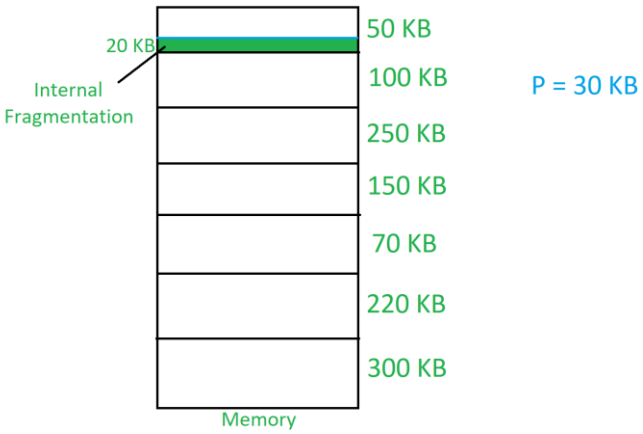
- Fixed(or Static) Partition
- Variable(or Dynamic) Partitioning




1. Fixed Partition Scheme

In the fixed partition scheme, memory is divided into fixed number of partitions. Fixed means number of partitions are fixed in the memory. In the fixed partition, in every partition only one process will be accommodated. Degree of multi-programming is restricted by number of partitions in the memory. Maximum size of the process is restricted by maximum size of the partition. Every partition is associated with the limit registers.

- **Limit Registers:** It has two limit:
- **Lower Limit:** Starting address of the partition.
- **Upper Limit:** Ending address of the partition.




Internal Fragmentation is found in fixed partition scheme. To overcome the problem of internal fragmentation, instead of fixed partition scheme, variable partition scheme is used.

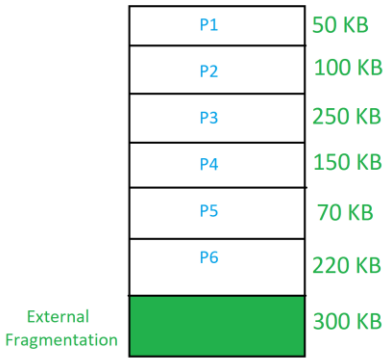
Disadvantages Fix partition scheme

- If a process of 19kb wants to allocate and we have free space which is not continuous we are not able to allocate the space.

2. Variable Partition Scheme

In the variable partition scheme, initially memory will be single continuous free block. Whenever the request by the process arrives, accordingly partition will be made in the memory. If the smaller processes keep on coming then the larger partitions will be made into smaller partitions.

- In variable partition schema initially, the memory will be full contiguous free block
- Memory divided into partitions according to the process size where process size will vary.
- One partition is allocated to each active partition.




External Fragmentation is found in variable partition scheme. To overcome the problem of external fragmentation, compaction technique is used or non-contiguous memory management techniques are used.

Solution of External Fragmentation

**1. Compaction**

Moving all the processes toward the top or towards the bottom to make free available memory in a single continuous place is called compaction. Compaction is undesirable to implement because it interrupts all the running processes in the memory.

**Disadvantage of Compaction**

- Page fault can occur.
- It consumes CPU time (overhead).

**2. Non-contiguous memory allocation**

1. **Physical address space:** Main memory (physical memory) is divided into blocks of the same size called frames. frame size is defined by the operating system by comparing it with the size of the process.
2. **Logical Address space:** Logical memory is divided into blocks of the same size called process pages. page size is defined by hardware system and these pages are stored in the main memory during the process in non-contiguous frames.

Advantages of Variable Partition Scheme

- Portion size = process size
- There is no internal fragmentation (which is the drawback of fixed partition schema).
- Degree of multiprogramming varies and is directly proportional to a number of processes.

Disadvantage Variable Partition Scheme

- External fragmentation is still there.

**Advantages of Contiguous Memory Management**

- It's simple to monitor how many memory blocks are still available for use, which determines how many more processes can be allocated RAM.
- Considering that the complete file can be read from the disc in a single session, contiguous memory allocation offers good read performance.
- Contiguous allocation is simple to set up and functions well.

**Disadvantages of Contiguous Memory Management**

- Fragmentation is not a problem. Since new files can be written to the disk after older ones.
- To select the appropriate hole size while creating a new file, it needs know its final size.
- The extra space in the holes would need to be reduced or used once the disk is full.

Fixed partitioning creates memory partitions of identical sizes. Variable partitioning offers more flexibility because it creates partitions of varying sizes but still results in memory fragmentation. Dynamic partitioning allocates partitions of the required sizes requested by the programs respectively.

## Partitioning Algorithms

There are various algorithms which are implemented by the Operating System in order to find out the holes in the linked list and allocate them to the processes.

The explanation about each of the algorithm is given below.

### 1. First Fit Algorithm

First Fit algorithm scans the linked list and whenever it finds the first big enough hole to store a process, it stops scanning and load the process into that hole. This procedure produces two partitions. Out of them, one partition will be a hole while the other partition will store the process.

First Fit algorithm maintains the linked list according to the increasing order of starting index. This is the simplest to implement among all the algorithms and produces bigger holes as compare to the other algorithms.

### 2. Next Fit Algorithm

Next Fit algorithm is similar to First Fit algorithm except the fact that, Next fit scans the linked list from the node where it previously allocated a hole.

Next fit doesn't scan the whole list, it starts scanning the list from the next node. The idea behind the next fit is the fact that the list has been scanned once therefore the probability of finding the hole is larger in the remaining part of the list.

Experiments over the algorithm have shown that the next fit is not better then the first fit. So it is not being used these days in most of the cases.

**3. Best Fit Algorithm**

The Best Fit algorithm tries to find out the smallest hole possible in the list that can accommodate the size requirement of the process.

Using Best Fit has some disadvantages.

1. 1. It is slower because it scans the entire list every time and tries to find out the smallest hole which can satisfy the requirement the process.
2. Due to the fact that the difference between the whole size and the process size is very small, the holes produced will be as small as it cannot be used to load any process and therefore it remains useless.
   Despite of the fact that the name of the algorithm is best fit, It is not the best algorithm among all.
3.

**4. Worst Fit Algorithm**

The worst fit algorithm scans the entire list every time and tries to find out the biggest hole in the list which can fulfill the requirement of the process.

Despite of the fact that this algorithm produces the larger holes to load the other processes, this is not the better approach due to the fact that it is slower because it searches the entire list every time again and again.

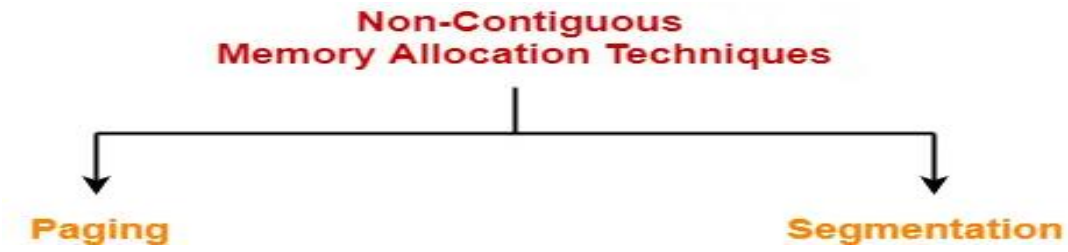
**5. Quick Fit Algorithm**

The quick fit algorithm suggestsmaintaining the different lists of frequently used sizes. Although, it is not practically suggestible because the procedure takes so much time to create the different lists and then expending the holes to load a process.

The first fit algorithm is **the best algorithm** among all because

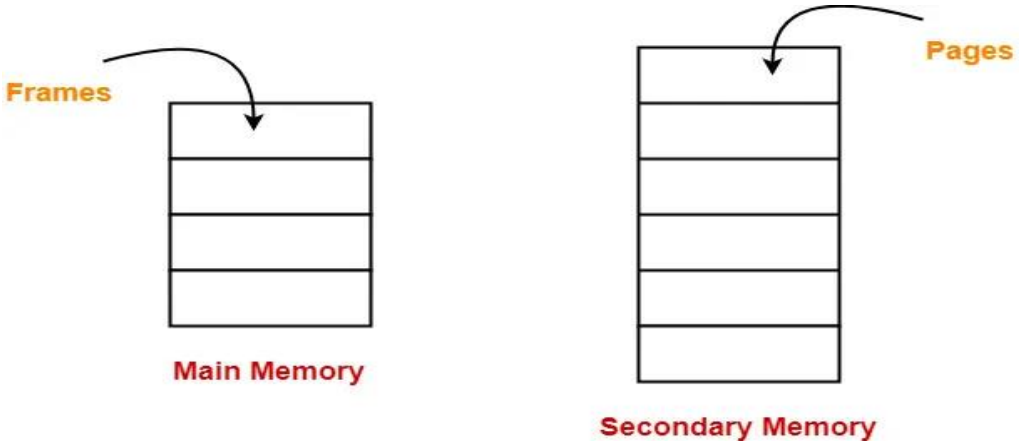
1. It takes lesser time compare to the other algorithms.
2. It produces bigger holes that can be used to load other processes later on.
3. It is easiest to implement.

https://www.geeksforgeeks.org/best-fit-allocation-in-operating-system/?ref=lbp




**Paging-**

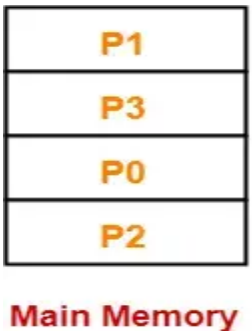
Paging is a fixed size partitioning scheme.
In paging, secondary memory and main memory are divided into equal fixed size partitions.
The partitions of secondary memory are called as pages.
The partitions of main memory are called as frames.




Each process is divided into parts where size of each part is same as page size.
The size of the last part may be less than the page size.
The pages of process are stored in the frames of main memory depending upon their availability.

**Example-**

Consider a process is divided into 4 pages P0, P1, P2 and P3.
Depending upon the availability, these pages may be stored in the main memory frames in a non-contiguous fashion as shown-



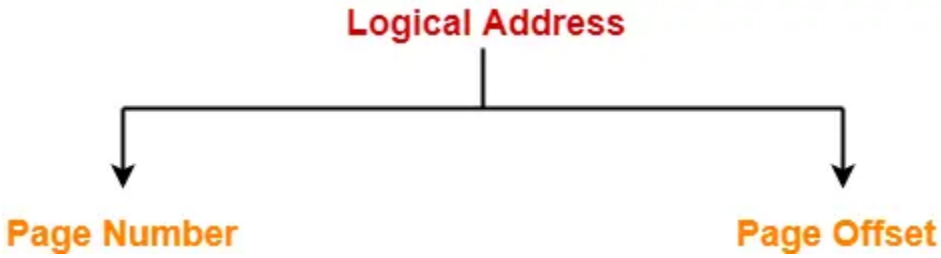

Translating Logical Address into Physical Address-

CPU always generates a logical address.
A physical address is needed to access the main memory.

Following steps are followed to translate logical address into physical address-

Step-01:

CPU generates a logical address consisting of two parts-
Page Number
Page Offset



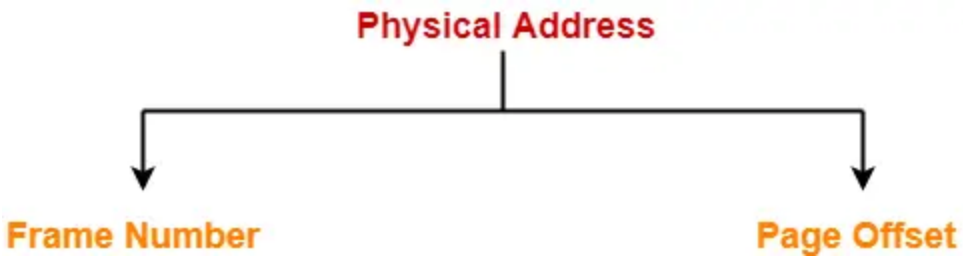

Page Number specifies the specific page of the process from which CPU wants to read the data.
Page Offset specifies the specific word on the page that CPU wants to read.

Step-02:

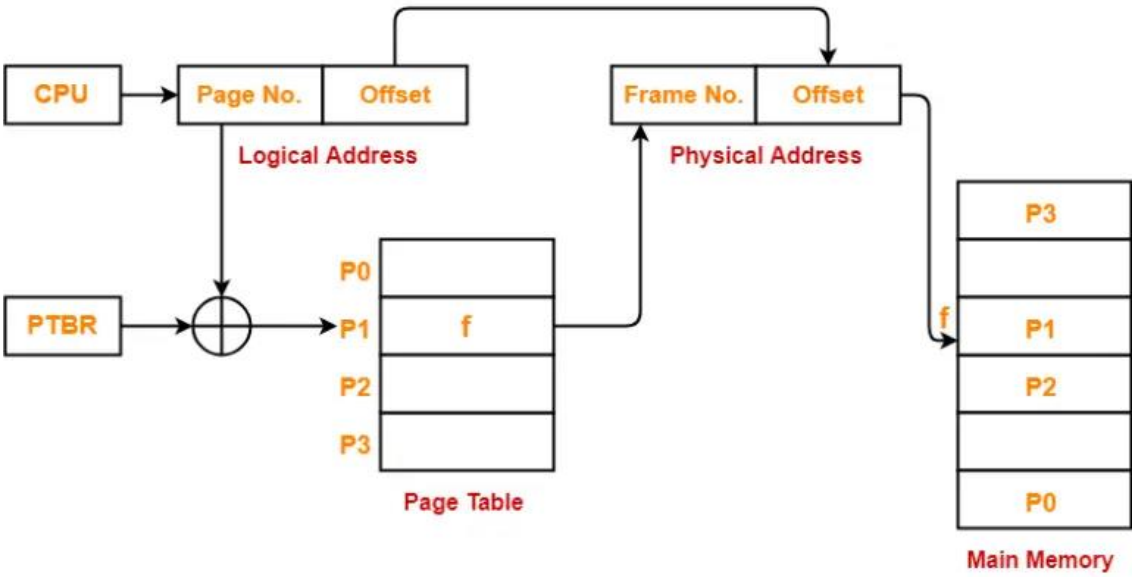
For the page number generated by the CPU,
**Page Table** provides the corresponding frame number (base address of the frame) where that page is stored in the main memory.

Step-03:

The frame number combined with the page offset forms the required physical address.




Frame number specifies the specific frame where the required page is stored.
Page Offset specifies the specific word that has to be read from that page.




Translating Logical Address into Physical Address

The **advantages o**f paging are-
It allows to store parts of a single process in a non-contiguous fashion.
It solves the problem of external fragmentation.

Disadvantages-

The disadvantages of paging are-
It suffers from internal fragmentation.
There is an overhead of maintaining a page table for each process.
The time taken to fetch the instruction increases since now two memory accesses are req

**Example:**

Consider a single level paging scheme. The virtual address space is 4 MB and page size is 4 KB. The maximum page table entry size possible such that the entire page table fits well in one page is:
For page table, to fit well in one page, we must have-

Page table size <= Page size

Number of pages the process is divided

= Process size / Page size

= 4 MB / 4 KB  MB = $2^{20}$, KB = $2^{10}$

$2^{20}$ / $2^{10}$

= $2^{10}$ pages

**Page table size**

= Number of entries in the page table x Page table entry size

= Number of pages the process is divided x Page table entry size

= $2^{10}$ x B bytes

Now,

According to the above condition, we must have-

$2^{10}$ x B bytes <= 4 KB

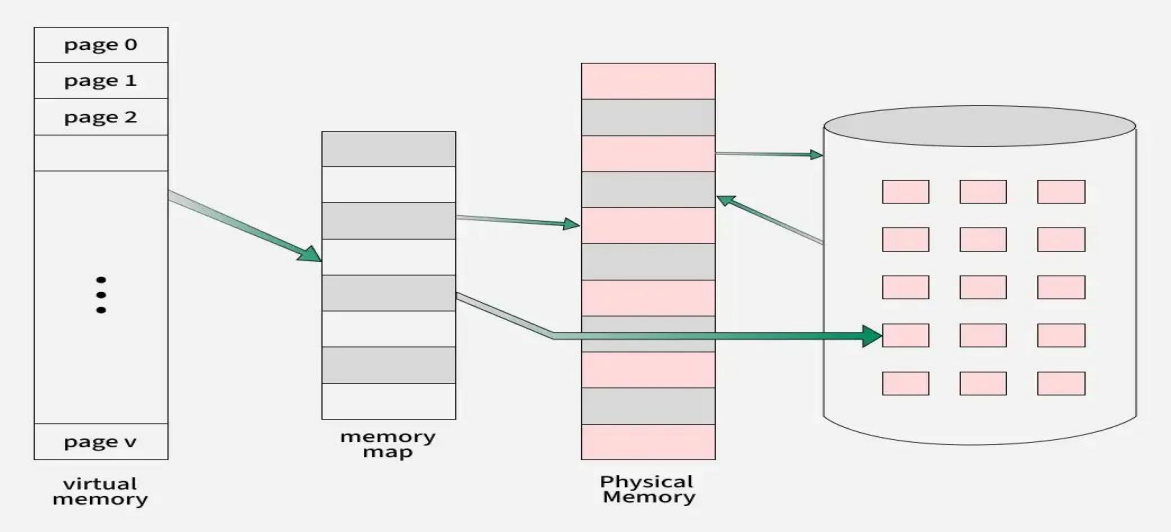
$2^{10}$ x B <= $2^{12}$

B <= 4

Thus, maximum page table entry size possible = 4 bytes.

Virtual memory is a memory management technique used by operating systems to give the appearance of a large, continuous block of memory to applications, even if the physical memory (RAM) is limited. It allows larger applications to run on systems with less RAM.

- The main objective of virtual memory is to support multiprogramming, The main advantage that virtual memory provides is, a running process does not need to be entirely in memory.
- Programs can be larger than the available physical memory. Virtual Memory provides an abstraction of main memory, eliminating concerns about storage limitations.
- A memory hierarchy, consisting of a computer system's memory and a disk, enables a process to operate with only some portions of its address space in RAM to allow more processes to be in memory.




It maps memory addresses used by a program, called virtual addresses, into physical addresses in computer memory.

- All memory references within a process are logical addresses that are dynamically translated into physical addresses at run time. This means that a process can be swapped in and out of the main memory such that it occupies different places in the main memory at different times during the course of execution.
- A process may be broken into a number of pieces and these pieces need not be continuously located in the main memory during execution. The combination of dynamic run-time address translation and the use of a page or segment table permits this.

If these characteristics are present then, it is not necessary that all the pages or segments are present in the main memory during execution. This means that the required pages need to be loaded into memory whenever required. Virtual memory is implemented using Demand Paging or Demand Segmentation.

Data in primary memory can be accessed faster than secondary memory but still, access times of primary memory are generally in a few microseconds, whereas the CPU is capable of performing operations in nanoseconds. Due to the time lag between accessing data and acting on data performance of the system decreases as the CPU is not utilized properly, it may remain idle for some time. In order to minimize this time gap new segment of memory is Introduced known as Cache Memory.

It is based on principle of locality of reference, which refers to the observation that program tries to access a relatively small portion of their address space at any given time, and repeatedly tries to access some portion of the memory. For ex: In fees department of your college, transactions are accessed frequently to check on the dues.

**Key Features of Cache Memory**

1. **Speed:** Faster than the main memory (RAM), which helps the CPU retrieve data more quickly.
2. **Proximity:** Located very close to the CPU, often on the CPU chip itself, reducing data access time.
3. **Function:** Temporarily holds data and instructions that the CPU is likely to use again soon, minimizing the need to access the slower main memory.

**Role of Cache Memory**

The role of cache memory is explained below,

- Cache memory plays a crucial role in computer systems.
- It provide faster access.
- It acts buffer between CPU and main memory(RAM).
- Primary role of it is to reduce average time taken to access data, thereby improving overall system performance.

**page replacement**

Page replacement is a process of swapping out an existing page from the frame of a main memory and replacing it with the required page.

Page replacement is required when-

All the frames of main memory are already occupied.

Thus, a page has to be replaced to create a room for the required page.

**Page Replacement Algorithms-**

Page replacement algorithms help to decide which page must be swapped out from the main memory to create a room for the incoming page.

FIFO Page Replacement Algorithm

LRU Page Replacement Algorithm

Optimal Page Replacement Algorithm

**FIFO**

As the name suggests, this algorithm works on the principle of "**First in First out**".

It replaces the oldest page that has been present in the main memory for the longest time.

It is implemented by keeping track of all the pages in a queue.

**Table:**

**LRU Page Replacement Algorithm-**

As the name suggests, this algorithm works on the principle of "Least Recently Used".

It replaces the page that has not been referenced by the CPU for the longest time.

**Optimal Page Replacement Algorithm-**

This algorithm replaces the page that will not be referred by the CPU in future for the longest time.

It is practically impossible to implement this algorithm.

This is because the pages that will not be used in future for the longest time can not be predicted.

However, it is the best known algorithm and gives the least number of page faults.

Hence, it is used as a performance measure criterion for other algorithms.

A multicore processor is an integrated circuit that has two or more processor cores attached for enhanced performance and reduced power consumption. These processors also enable more efficient simultaneous processing of multiple tasks, such as with parallel processing and multithreading.

Why is multi-core important?

The primary benefit of high core count processors is that they are able to complete multiple tasks at the same time, and they can switch from one program to the next with ease. A multi-core processor has two or more cores, each of which reads and executes program instructions

Benefits

Multicore technology allows a system to perform more tasks and improve overall system performance.

It's ideal for applications that require extensive computations on separate sets of data, like video encoding and 3D rendering.

The differences between job scheduling and process scheduling:

Job scheduling is basically the process in which a process is selected to be brought into a queue while process scheduling is the process of a process manager scheduling a certain process and allocating it to the CPU on the basis of a strategy.

Job scheduling:

It is the mechanism of the selection of a process that has to be brought into the ready queue.

It is also called long-term scheduling.

It is done by the job scheduler or long-term scheduler.

The transfer of process occurs from a new state to a ready state.

In job scheduling, there is more control over multiprogramming.

Process scheduling:

It is the mechanism of the selection of the process that has to be executed next and the allocation of the CPU to that process.

It is also called short-term scheduling.

It is done by the process scheduler or CPU scheduler or short-term scheduler.

The transfer of process occurs from a ready state to a running state.

In process scheduling, there is less control over multiprogramming.

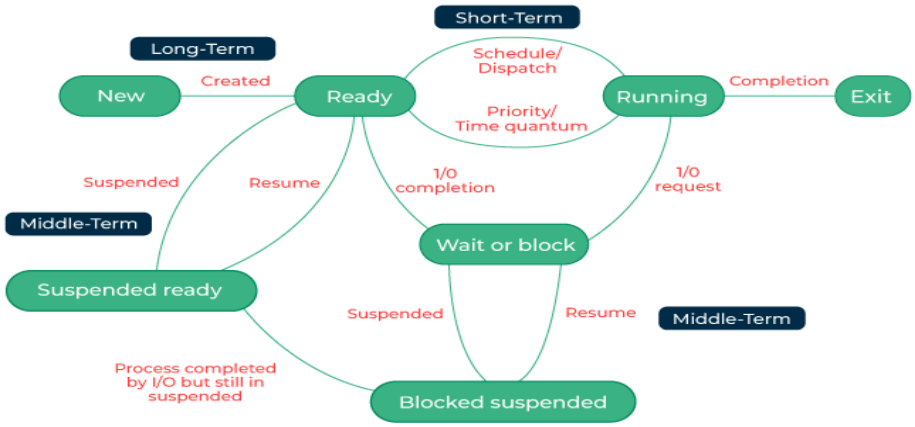
**process scheduling**

A process is the instance of a computer program in execution.

- Scheduling is important in operating systems with multiprogramming as multiple processes might be eligible for running at a time.
- One of the key responsibilities of an Operating System (OS) is to decide which programs will execute on the CPU.
- Process Schedulers are fundamental components of operating systems responsible for deciding the order in which processes are executed by the CPU. In simpler terms, they manage how the CPU allocates its time among multiple tasks or processes that are competing for its attention.

What is Process Scheduling?

Process scheduling is the activity of the process manager that handles the removal of the running process from the CPU and the selection of another process based on a particular strategy. Throughout its lifetime, a process moves between various **scheduling queues**, such as the ready queue, waiting queue, or devices queue.




Scheduling falls into one of two categories:

- **Non-Preemptive:** In this case, a process's resource cannot be taken before the process has finished running. When a running process finishes and transitions to a waiting state, resources are switched.
- **Preemptive:** In this case, the OS can switch a process from running state to ready state. This switching happens because the CPU may give other processes priority and substitute the currently active process for the higher priority process.

ypes of Process Schedulers

There are three types of process schedulers:

**1. Long Term or Job Scheduler**

Long Term Scheduler loads a process from disk to main memory for execution. The new process to the 'Ready State'.

- It mainly moves processes from Job Queue to Ready Queue.
- It controls the Degree of Multi-programming, i.e., the number of processes present in a ready state or in main memory at any point in time.
- It is important that the long-term scheduler make a careful selection of both I/O and CPU-bound processes. I/O-bound tasks are which use much of their time in input and output operations while CPU-bound processes are which spend their time on the CPU. The job scheduler increases efficiency by maintaining a balance between the two.
- In some systems, the long-term scheduler might not even exist. For example, in time-sharing systems like Microsoft Windows, there is usually no long-term scheduler. Instead, every new process is directly added to memory for the short-term scheduler to handle.
- Slowest among the three (that is why called long term).

**2. Short-Term or CPU Scheduler**

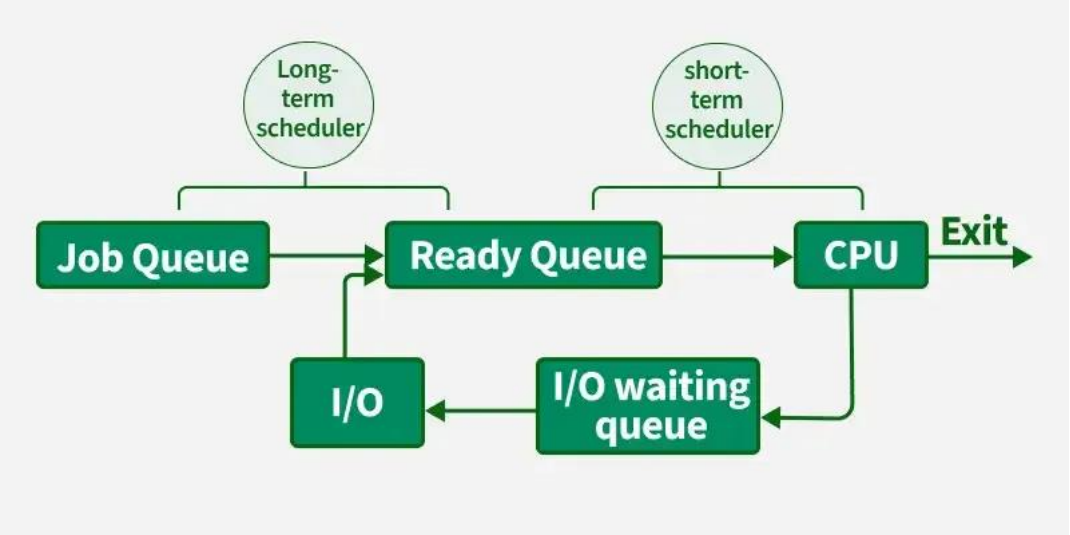
CPU Scheduler is responsible for selecting one process from the ready state for running (or assigning CPU to it).

- STS (Short Term Scheduler) must select a new process for the CPU frequently to avoid starvation.
- The CPU scheduler uses different scheduling algorithms to balance the allocation of CPU time.
- It picks a process from ready queue.
- Its main objective is to make the best use of CPU.
- It mainly calls dispatcher.
- Fastest among the three (that is why called Short Term).

The dispatcher is responsible for loading the process selected by the Short-term scheduler on the CPU (Ready to Running State). Context switching is done by the dispatcher only. A dispatcher does the following work:

- Saving context (process control block) of previously running process if not finished.
- Switching system mode to user mode.
- Jumping to the proper location in the newly loaded program.

Time taken by dispatcher is called dispatch latency or process context switch time.

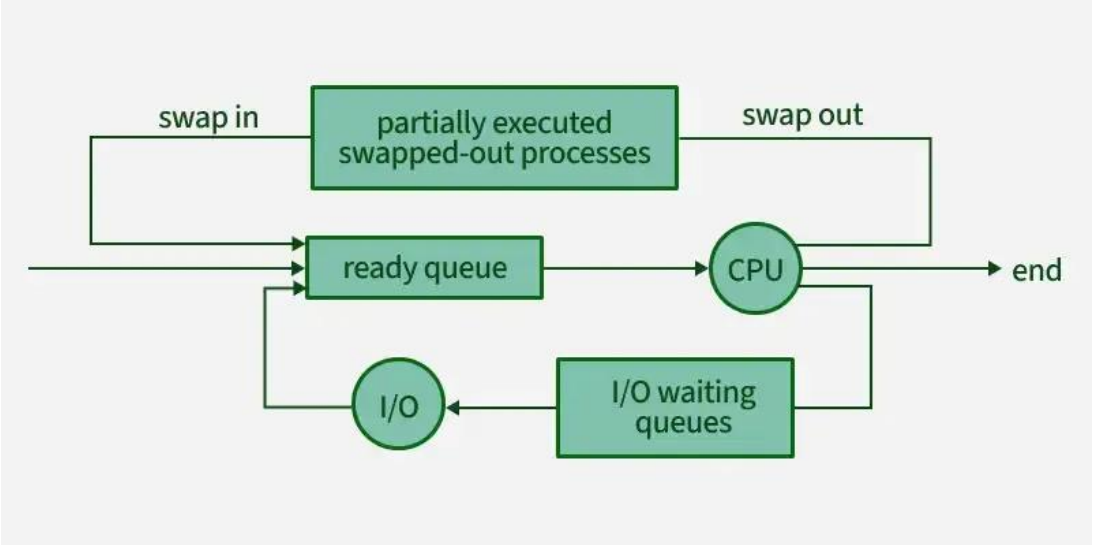



*Short-Term Scheduler*

### 3. Medium-Term Scheduler

Medium Term Scheduler (MTS) is responsible for moving a process from memory to disk (or swapping).

- It reduces the degree of multiprogramming (Number of processes present in main memory).
- A running process may become suspended if it makes an I/O request. A suspended processes cannot make any progress towards completion. In this condition, to remove the process from memory and make space for other processes, the suspended process is moved to the secondary storage. This process is called swapping, and the process is said to be swapped out or rolled out. Swapping may be necessary to improve the process mix (of CPU bound and IO bound)
- When needed, it brings process back into memory and pick up right where it left off.
- It is faster than long term and slower than short term.




What is Preemptive Scheduling?

The operating system can interrupt or preempt a running process to allocate CPU time to another process, typically based on priority or time-sharing policies. Mainly a process is switched from the running state to the ready state.

- **Completion Time:** Time at which process completes its execution.
- **Turn Around Time:** Time Difference between completion time and arrival time. **Turn Around Time = Completion Time – Arrival Time**
- **Waiting Time(W.T):** Time Difference between turn around time and burst time. **Waiting Time = Turn Around Time – Burst Time**

Algorithms based on preemptive scheduling are Round Robin (RR) , Shortest Remaining Time First (SRTF) , Priority (preemptive version)

| Process ID | Arrival Time | Burst Time |
|---|---|---|
| 1 | 0 | 5 |
| 2 | 1 | 6 |
| 3 | 2 | 3 |
| 4 | 3 | 1 |
| 5 | 4 | 5 |
| 6 | 6 | 4 |

in the above table, there are six processes named as P1, P2, P3, P4, P5 and P6. Their arrival time and burst time are given below in the table. The time quantum of the system is 4 units.

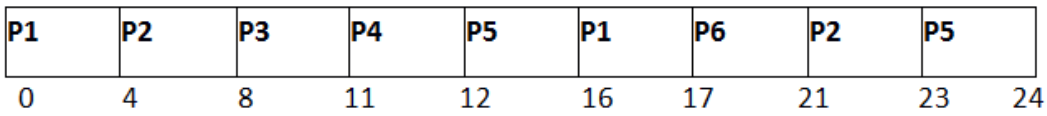
Ready Queue:

Initially, at time 0, process P1 arrives which will be scheduled for the time slice 4 units.

| P1 |
|---|
| 5 |

Meanwhile the execution of P1, four more processes P2, P3, P4 and P5 arrives in the ready queue. P1 has not completed yet, it needs another 1 unit of time hence it will also be added back to the ready queue.

| P2 | P3 | P4 | P5 | P1 |
|---|---|---|---|---|
| 6 | 3 | 1 | 5 | 1 |



| Process ID | Arrival Time | Burst Time | Completion Time | Turn Around Time | Waiting Time |
|---|---|---|---|---|---|
| 1 | 0 | 5 | 17 | 17 | 12 |
| 2 | 1 | 6 | 23 | 22 | 16 |
| 3 | 2 | 3 | 11 | 9 | 6 |
| 4 | 3 | 1 | 12 | 9 | 8 |
| 5 | 4 | 5 | 24 | 20 | 15 |
| 6 | 6 | 4 | 21 | 15 | 11 |

Avg Waiting Time = (12+16+6+8+15+11)/6 = 76/6 units

First Come First Serve Algorithm

The First Come First Serve Algorithm can be executed in Pre Emptive and Non Pre Emptive manner.
he characteristics of FCFS CPU Process Scheduling are:

1. Implementation is simple.
2. Does not cause any causalities while using
3. It adopts a non pre emptive and pre emptive strategy.
4. It runs each procedure in the order that they are received.
5. Arrival time is used as a selection criterion for procedures.

## Advantages of FCFS CPU Process Scheduling

The advantages of FCFS CPU Process Scheduling are:

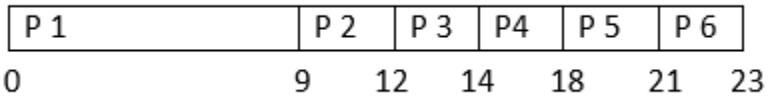
1. In order to allocate processes, it uses the First In First Out queue.
2. The FCFS CPU Scheduling Process is straight forward and easy to implement.

3. In the FCFS situation pre emptive scheduling, there is no chance of process starving.
4. As there is no consideration of process priority, it is an equitable algorithm.

**FCFS – NON PREEMPTIVE APPROACH**

| S. No | Process ID | Process Name | Arrival Time | Burst Time |
|---|---|---|---|---|
| 1 | P 1 | A | 0 | 9 |
| 2 | P 2 | B | 1 | 3 |
| 3 | P 3 | C | 1 | 2 |
| 4 | P 4 | D | 1 | 4 |
| 5 | P 5 | E | 2 | 3 |
| 6 | P 6 | F | 3 | 2 |

Gantt chart:




Turn Around Time = Completion Time - Arrival Time

Waiting Time = Turn Around Time - Burst Time

| S. No | Process ID | Arrival Time | Burst Time | Completion Time | Turn Around Time | Waiting Time |
|---|---|---|---|---|---|---|
| 1 | P 1 | 0 | 9 | 9 | 9 | 0 |
| 2 | P 2 | 1 | 3 | 12 | 11 | 8 |
| 3 | P 3 | 1 | 2 | 14 | 13 | 11 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 4 | P 4 | 1 | 4 | 18 | 17 | 13 |
| 5 | P 5 | 2 | 3 | 21 | 19 | 16 |
| 6 | P 6 | 3 | 2 | 23 | 20 | 18 |

**The Average Completion Time is:**

Average CT = ( 9 + 12 + 14 + 18 + 21 + 23 ) / 6

Average CT = 97 / 6

Average CT = 16.16667

**The Average Waiting Time is:**

Average WT = ( 0 + 8 + 11 + 13 + 16 + 18 ) /6

Average WT = 66 / 6

Average WT = 11

**The Average Turn Around Time is:**

Average TAT = ( 9 + 11 + 13 + 17 + 19 +20 ) / 6

Average TAT = 89 / 6

Average TAT = 14.83334