**What is an operating system?An operating system (OS) is the program that, after being initially loaded into the computer by a boot program, manages all of the other application programs in a computer. The application programs make use of the operating system by making requests for services through a defined application program interface (**[**API**](https://www.techtarget.com/searchapparchitecture/definition/application-program-interface-API)**). In addition, users can interact directly with the operating system through a user interface, such as a command-line interface (CLI) or a graphical UI (GUI).A system call is a way for programs to interact with the operating system. A computer program makes a system call when it makes a request to the operating system's kernel. System call provides the services of the operating system to the user programs via**

***Application Program Interface(API);- 1.List of operating system;-*** [***Batch Operating System***](https://www.geeksforgeeks.org/batch-processing-operating-system/) ***2.***[***Multi-Programming System***](https://www.geeksforgeeks.org/multiprogramming-in-operating-system/) ***3.Multi-Processing System 4.Multi-Tasking Operating System 5.***[***Time-Sharing Operating System***](https://www.geeksforgeeks.org/time-sharing-operating-system/) ***6.Distributed Operating System 7.***[***Network Operating System***](https://www.geeksforgeeks.org/what-is-a-network-operating-system/) ***8.***[***Real-Time Operating System***](https://www.geeksforgeeks.org/real-time-operating-system-rtos/)***.***

**What is the process of the computer system?The three stages of computing are input, processing and output. A computer works through these stages by 'running' a program. A program is a set of step-by-step instructions which tells the computer exactly what to do with input in order to produce the required output.What is the definition of a thread?A thread is, fundamentally, a clear, well-defined theme, subject, topic or focus. A thread may exist in human communication, such as a topical group chat or email exchange. A thread also exists as the processing focus of a software program, such as an operating system (OS) or application.What do you mean by scheduling in operating system?Scheduling ;--is a method that is used to distribute valuable computing resources, usually processor time, bandwidth and memory, to the various processes, threads, data flows and applications that need them.**

## What is an Operating System? types of OS (Operating System

**An Operating System (OS);- is a software that acts as an interface between computer hardware components and the user. Every computer system must have at least one operating system to run other programs. Applications like Browsers, MS Office, Notepad Games, etc., need some environment to run and perform its tasks. types of OS (Operating System):-1.Batch Operating System2.Multitasking/Time Sharing OS3.Multiprocessing OS4.Real Time OS5.Distributed OS6.Network OS7.Mobile OS.**

### 1.Batch Operating System;-Some computer processes are very lengthy and time-consuming. To speed the same process, a job with a similar type of needs are batched together and run as a group.The user of a batch operating system never directly interacts with the computer. In this type of OS, every user prepares his or her job on an offline device like a punch card and submit it to the computer operator.2.Multi-Tasking/Time-sharing Operating systems;-Time-sharing operating system enables people located at a different terminal(shell) to use a single computer system at the same time. The processor time (CPU) which is shared among multiple users is termed as time sharing.3.Real time OS;-A [real time operating system](https://www.guru99.com/real-time-operating-system.html) time interval to process and respond to inputs is very small. Examples: Military Software Systems, Space Software Systems are the Real time OS example.4.Distributed Operating System;-Distributed systems use many processors located in different machines to provide very fast computation to its users.5.Network Operating System;-Network Operating System runs on a server. It provides the capability to serve to manage data, user, groups, security, application, and other networking functions.6.Mobile OS;-Mobile operating systems are those OS which is especially that are designed to power smartphones, tablets, and wearables devices.Some most famous mobile operating systems are Android and iOS, but others include BlackBerry, Web, and watchOS.

# States of a Process in Operating Systems;--A process has several stages that it passes through from beginning to end. There must be a minimum of five states. Even though during execution, the process could be in one of these states, the names of the states are not standardized. Each [process](https://www.geeksforgeeks.org/introduction-of-process-management/)goes through several stages throughout its life cycle.Process States in Operating SystemThe states of a [process](https://www.geeksforgeeks.org/introduction-of-process-management/)are as follows:

**New (Create): In this step, the process is about to be created but not yet created. It is the program that is present in secondary memory that will be picked up by OS to create the process.Ready: New -> Ready to run. After the creation of a process, the process enters the ready state i.e. the process is loaded into the main memory. The process here is ready to run and is waiting to get the CPU time for its execution. Processes that are ready for execution by the CPU are maintained in a queue called ready queue for ready processes.Run: The process is chosen from the ready queue by the CPU for execution and the instructions within the process are executed by any one of the available CPU cores.Blocked or Wait: Whenever the process requests access to I/O or needs input from the user or needs access to a critical region(the lock for which is already acquired) it enters the blocked or waits for the state. The process continues to wait in the main memory and does not require CPU. Once the I/O operation is completed the process goes to the ready state.Terminated or Completed: Process is killed as well as**[**PCB**](https://www.geeksforgeeks.org/process-table-and-process-control-block-pcb/)**is deleted. The resources allocated to the process will be released or deallocated.Suspend Ready: Process that was initially in the ready state but was swapped out of main memory(refer to Virtual Memory topic) and placed onto external storage by the scheduler is said to be in suspend ready state. The process will transition back to a ready state whenever the process is again brought onto the main memory.Suspend wait or suspend blocked: Similar to suspend ready but uses the process which was performing I/O operation and lack of main memory caused them to move to secondary memory. When work is finished it may go to suspend ready.**

**Q Services of Operating System;-1.Program execution2.Input Output Operations**

**3.Communication between Process 4.File Management**

### Program Execution;-It is the Operating System that manages how a program is going to be executed. It loads the program into the memory after which it is executed. The order in which they are executed depends on the CPU Scheduling Algorithms. A few are FCFS, SJF, etc. When the program is in execution, the Operating System also handles deadlock i.e. no two processes come for execution at the same time. The Operating System is responsible for the smooth execution of both user and system programs. The Operating System utilizes various resources available for the efficient running of all types of functionalities.Input Output Operations;-Operating System manages the input-output operations and establishes communication between the user and device drivers. Device drivers are software that is associated with hardware that is being managed by the OS so that the sync between the devices works properly. It also provides access to input-output devices to a program when needed. ****Communication between Processes;-****The Operating system manages the communication between processes. Communication between processes includes data transfer among them. If the processes are not on the same computer but connected through a computer network, then also their communication is managed by the Operating System itself.

### File Management;-The operating system helps in managing files also. If a program needs access to a file, it is the operating system that grants access. These permissions include read-only, read-write, etc. It also provides a platform for the user to create, and delete files.

## A virtual machine (VM);-is a virtual environment that functions as a virtual computer system with its own CPU, memory, network interface, and storage, created on a physical hardware system (located off- or on-premises). Software called a [hypervisor](https://www.redhat.com/en/topics/virtualization/what-is-a-hypervisor) separates the machine’s resources from the hardware and provisions them appropriately so they can be used by the VM. How do VMs work?[Virtualization technology](https://www.redhat.com/en/topics/virtualization) allows you to share a system with many virtual environments. The hypervisor manages the hardware and separates the physical resources from the virtual environments. Resources are partitioned as needed from the physical environment to the VMs.

#### Type 1;-A type 1 hypervisor is on bare metal. VM resources are scheduled directly to the hardware by the hypervisor. KVM is an example of a type 1 hypervisor. KVM was merged into the [Linux® kernel](https://www.redhat.com/en/topics/linux/what-is-the-linux-kernel)in 2007, so if you’re using a modern version of [Linux](https://www.redhat.com/en/topics/linux/what-is-linux), you already have access to KVM. Type 2;-A type 2 hypervisor is hosted. VM resources are scheduled against a host operating system, which is then executed against the hardware. VMware Workstation and Oracle VirtualBox are examples of type 2 hypervisors. Server consolidation is a top reason to use VMs. Most operating system and application deployments only use a small amount of the physical resources available Why use a VM?when deployed to bare metal. By virtualizing your servers, you can place many virtual servers onto each physical server to improve hardware utilization.

**What do you mean by scheduling;-Scheduling is the process of arranging, controlling and optimizing work and workloads in a production process or manufacturing process. Scheduling is used to allocate plant and machinery resources, plan human resources, plan production processes and purchase materials.Types of Scheduling**

**1.Long-termScheduling;--Long term scheduling is performed when a new process is created. It is shown in the figure below. If the number of ready processes in the ready queue becomes very high, then there is a overhead on the operating system (i.e., processor) for maintaining long lists, context switching and dispatching increases. Therefore, allow only limited number of processes in to the ready queue. The "long-term scheduler" managers this. Long-term scheduler determines which programs are admitted into the system for processing. Once when admit a process or job, it becomes process and is added to the queue for the short-term scheduler**

**Medium-termScheduling;--2.Medium-term scheduling is a part of the swapping function. When part of the main memory gets freed, the operating system looks at the list of suspend ready processes, decides which one is to be swapped in (depending on priority, memory and other resources required, etc). This scheduler works in close conjunction with the long-term scheduler. It will perform the swapping-in function among the swapped-out processes. Medium-term scheduler executes some what more frequently.**

**Short-termScheduling Short-term scheduler;-- is also called as dispatcher. Short-term scheduler is invoked whenever an event occurs, that may lead to the interruption of the current running process. For example clock interrupts, I/O interrupts, operating system calls, signals, etc. Short-term scheduler executes most frequently. It selects from among the processes that are ready to execute and allocates the CPU to one of them. It must select a new process for the CPU frequently. It must be very fast.**

**Scheduling Criteria;;--Scheduling criteria is also called as scheduling methodology. Key to multiprogramming is scheduling.Different CPU scheduling algorithm have different properties .The criteria used for comapring these algorithms include the following:**

| **S. No** | **Parallel System** | **Distributed System** |
| --- | --- | --- |
| **1.** | **Parallel systems are the systems that can process the data simultaneously, and increase the computational speed of a computer system.** | **In these systems, applications are running on multiple computers linked by communication lines.** |
| **2.** | **Parallel systems work with the simultaneous use of multiple computer resources which can include a single computer with multiple processors.** | **The distributed system consists of a number of computers that are connected and managed so that they share the job processing load among various computers distributed over the network.** |
| **3.** | **Tasks are performed with a more speedy process.** | **Tasks are performed with a less speedy process.** |
| **4.** | **These systems are multiprocessor systems.** | **In Distributed Systems, each processor has its own memory.** |
| **5.** | **It is also known as a tightly coupled system.** | **Distributed systems are also known as loosely coupled systems.** |
| **6.** | **These systems have close communication with more than one processor.** | **These systems communicate with one another through various communication lines, such as high-speed buses or telephone lines.** |
| **7.** | **These systems share a memory, clock, and peripheral devices** | **These systems do not share memory or clock in contrast to parallel systems.** |
| **8.** | **In this, all processors share a single master clock for synchronization.** | **In this there is no global clock in distributed computing, it uses various synchronization algorithms.** |

**What is First Come First Serve Scheduling?(fcfs)The First come first serve scheduling algorithm is non-preemptive in nature i.e, if a process is already running, then it is not interrupted by another process until the currently running process is executed completely.Buying a movie ticket from the ticket counter is a perfect real-life example of a first come first serve (FCFS) algorithm. The person who comes first and stands in the queue gets to buy the ticket first. Similarly in the FCFS scheduling algorithm, the process that arrives first gets executed first. When we run a program, we create a particular instance of the program called a process. There might be a condition where more than one process is created at a given time and the CPU has to serve all the processes. There are various process scheduling algorithms that decide which process has to be executed at a given time by considering various factors. FCFS or First come first serve is one such algorithm that schedules the processes. What is the Shortest Job First Scheduling in the Operating System?**

**Shortest Job First (SJF);- algorithm is also known as Shortest Job Next (SJN) or Shortest Process Next (SPN). It is a CPU processes scheduling algorithm that sorts and executes the process with the smallest execution time first and then the subsequent processes with the increased execution time. Both preemptive and non-preemptive scheduling strategies are possible in the SJF scheduling algorithm. In SJF, there is a significant amount of reduction in the average waiting time for other processes that are waiting to be executed.However, it can be quite challenging to estimate the burst time required for a process, making it difficult to apply this technique to the operating system scheduling process.The burst time for a process can only be approximated or predicted. To get the most out of the SJF algorithm, our approximations must be correct. Numerous methods can be used to predict a process's CPU burst time.1. Static Techniques:Process Size: We approximate the burst time of the upcoming process by using the burst time of an older process having a similar size.Process Type: Depending on the type of the process, we can estimate the burst time. There are several different sorts of processes, , for example,, Background Processes, User Processes, Operating System Processes, etc.**

# Components of Operating System;--1.Process Management 2.I/O Device Management 3.File Management 4.Network Management;Process Management;-A process is program or a fraction of a program that is loaded in main memory. A process needs certain resources including CPU time, Memory, Files, and I/O devices to accomplish its task. The process management component manages the multiple processes running simultaneously on the Operating System.I/O Device Management;--One of the purposes of an operating system is to hide the peculiarities of specific hardware devices from the user. I/O Device Management provides an abstract level of H/W devices and keep the details from applications to ensure proper use of devices, to prevent errors, and to provide users with convenient and efficient programming environment.File Management;--File management is one of the most visible services of an operating system. Computers can store information in several different physical forms; magnetic tape, disk, and drum are the most common forms.A file is defined as a set of correlated information and it is defined by the creator of the file. Mostly files represent data, source and object forms, and programs. Data files can be of any type like alphabetic, numeric, and alphanumeric.Network Management;--The definition of network management is often broad, as network management involves several different components. Network management is the process of managing and administering a computer network. A computer network is a collection of various types of computers connected with each other.Network management comprises fault analysis, maintaining the quality of service, provisioning of networks, and performance management.

### What is a Thread?  A thread is a single sequence stream within a process. Because threads have some of the properties of processes, they are sometimes called lightweight processes. Threads are a popular way to improve the application through parallelism. For example, in a browser, multiple tabs can be different threads. MS Word uses multiple threads, one thread to format the text, another thread to process inputs, etc. What are the differences between process and thread?A thread has its own program counter (PC), a register set, and a stack space. Threads are not independent of one another, like processes. As a result, threads share with other threads their code section, data section, and OS resources like open files and signals.

### What is virtual memory?Virtual memory creates an illusion that each user has one or more contiguous address spaces, each beginning at address zero. The sizes of such virtual address spaces are generally very high. The idea of virtual memory is to use disk space to extend the RAM. Running processes don’t need to care whether the memory is from RAM or disk. The illusion of such a large amount of memory is created by subdividing the virtual memory into smaller pieces, which can be loaded into physical memory whenever they are needed by a process.

**Explain the main purpose of an operating system?An operating system acts as an intermediary between the user of a computer and computer hardware. The purpose of an operating system is to provide an environment in which a user can execute programs conveniently and efficiently. An operating system is a software that manages computer hardware. The hardware must provide appropriate mechanisms to ensure the correct operation of the computer system and to prevent user programs from interfering with the proper operation of the system.**

### What is a kernel?A kernel is the central component of an operating system that manages the operations of computers and hardware. It basically manages operations of memory and CPU time. It is a core component of an operating system. Kernel acts as a bridge between applications and data processing performed at the hardware level using inter-process communication and system calls.What is the time-sharing system?Time-sharing is a logical extension of multiprogramming. The CPU performs many tasks by switches that are so frequent that the user can interact with each program while it is running. A time-shared operating system allows multiple users to share computers simultaneously.

### Describe the objective of multi-programming.Multi-programming increases CPU utilization by organizing jobs (code and data) so that the CPU always has one to execute. The main objective of multi-programming is to keep multiple jobs in the main memory. If one job gets occupied with IO, the CPU can be assigned to other jobs.

### 

**Ques 5. What is a Thread? Explain the differences between a Process and a Thread.**

**Resource allocation: A process has its own memory space and system resources, such as file descriptors, sockets, and environment variables. Each process is allocated its own CPU time and other system resources, which makes processes more independent and robust than threads. In contrast, a thread shares the same memory space and system resources as the other threads within the same process.Scheduling: A process is scheduled and managed independently by the operating system scheduler, whereas threads are scheduled and managed within the process by a user-level scheduler. Since each process has its own memory space, there is less chance of memory corruption, while threads may interfere with each other’s memory if they are not synchronized correctly.  
Communication: Processes communicate with each other using inter-process communication (IPC) mechanisms such as pipes, sockets, and shared memory. Threads, on the other hand, can communicate directly with other threads within the same process by sharing memory.Overhead: Creating and managing a process is more expensive in terms of system resources than creating and managing a thread. This is because each process requires its own memory space, file descriptors, and other system resources, whereas threads share these resources with other threads in the same process.**

### What do you mean by process synchronization?Process synchronization is basically a way to coordinate processes that use shared resources or data. It is very much essential to ensure synchronized execution of cooperating processes so that will maintain data consistency. Its main purpose is to share resources without any interference using mutual exclusion. There are two types of process synchronization:1) Independent Process 2)Cooperative Process

**What do you mean by FCFS?FCFS (First Come First Serve) is a type of OS scheduling algorithm that executes processes in the same order in which processes arrive. In simple words, the process that arrives first will be executed first. It is non-preemptive in nature. FCFS scheduling may cause the problem of starvation if the burst time of the first process is the longest among all the jobs. Burst time here means the time that is required in milliseconds by the process for its execution. It is also considered the easiest and simplest OS scheduling algorithm as compared to others. Implementation of FCFS is generally managed with help of the FIFO (First In First Out) queue.**

**Paging: It is generally a memory management technique that allows OS to retrieve processes from secondary storage into main memory. It is a non-contiguous allocation technique that divides each process in the form of pages.   
Segmentation: It is generally a memory management technique that divides processes into modules and parts of different sizes. These parts and modules are known as segments that can be allocated to process. What is thrashing in OS?It is generally a situation where the CPU performs less productive work and more swapping or paging work. It spends more time swapping or paging activities rather than its execution. By evaluating the level of CPU utilization, a system can detect thrashing. It occurs when the process does not have enough pages due to which the page-fault rate is increased. It inhibits much application-level processing that causes computer performance to degrade or collapse.**

### Write difference between micro kernel and monolithic kernel?**MicroKernel**: It is a minimal OS that executes only important functions of OS. It only contains a near-minimum number of features and functions that are required to implement OS.  Example: QNX, Mac OS X, K42, etc. **Monolithic Kernel:** It is an OS architecture that supports all basic features of computer components such as resource management, memory, file, etc.  Example: Solaris, DOS, OpenVMS, Linux, etc.

### What is a deadlock in OS? What are the necessary conditions for a deadlock?Deadlock is generally a situation where a set of processes are blocked as each process is holding resources and waits to acquire resources held by another process. In this situation, two or more processes simply try to execute simultaneously and wait for each to finish their execution because they are dependent on each other. We can see a hand problem in our system whenever a deadlock occurs in a program. It is one of the common problems you can see in multiprocessing.  **Necessary Conditions for Deadlock** There are basically four necessary conditions for deadlock as given below:@Mutual Exclusion@Hold and Wait@No Pre-emption@Circular Wait or Resource Wait

**WHAT IS MEMORY? Computer memory can be defined as a collection of some data represented in the binary format. On the basis of various functions, memory can be classified into various categories. We will discuss each one of them later in detail. A computer device that is capable to store any information or data temporally or permanently is called storage device.**

**FIXED PARTITIONING The earliest and one of the simplest technique which can be used to load more than one processes into the main memory is Fixed partitioning or Contiguous memory allocation. In this technique, the main memory is divided into partitions of equal or different sizes. The operating system always resides in the first partition while the other partitions can be used to store user processes. The memory is assigned to the processes in contiguous way**

**In fixed partitioning, 1. The partitions cannot overlap. 2. A process must be contiguously present in a partition for the execution. There are various cons of using this technique.**

**1. Internal Fragmentation If the size of the process is lesser then the total size of the partition then some size of the partition get wasted and remain unused. This is wastage of the memory and called internal fragmentation. As shown in the image below, the 4 MB partition is used to load only 3 MB process and the remaining 1 MB got wasted**

**2. External Fragmentation The total unused space of various partitions cannot be used to load the processes even though there is space available but not in the contiguous form. As shown in the image below, the remaining 1 MB space of each partition cannot be used as a unit to store a 4 MB process. Despite of the fact that the sufficient space is available to load the process, process will not be loaded.**

**Demand Paging According to the concept of Virtual Memory, in order to execute some process, only a part of the process needs to be present in the main memory which means that only a few pages will only be present in the main memory at any time. However, deciding, which pages need to be kept in the main memory and which need to be kept in the secondary memory, is going to be difficult because we cannot say in advance that a process will require a particular page at particular time. Therefore, to overcome this problem, there is a concept called Demand Paging is introduced. It suggests keeping all pages of the frames in the secondary memory until they are required. In other words, it says that do not load any page in the main memory until it is required. Whenever any page is referred for the first time in the main memory, then that page will be found in the secondary memory. After that, it may or may not be present in the main memory depending upon the page replacement algorithm which will be covered later in this tutorial.**

**What is a Page Fault? If the referred page is not present in the main memory then there will be a miss and the concept is called Page miss or page fault. The CPU has to access the missed page from the secondary memory. If the number of page fault is very high then the effective access time of the system will become very high.**

**SEGMENTATION:- In Operating Systems, Segmentation is a memory management technique in which, the memory is divided into the variable size parts. Each part is known as segment which can be allocated to a process. The details about each segment are stored in a table called as segment table. Segment table is stored in one (or many) of the segments.**

### Page Replacement Algorithms: **1. First In First Out (FIFO):**This is the simplest page replacement algorithm. In this algorithm, the operating system keeps track of all pages in the memory in a queue, the oldest page is in the front of the queue. When a page needs to be replaced page in the front of the queue is selected for removal. **Example 1:**Consider page reference string 1, 3, 0, 3, 5, 6, 3 with 3 page frames.Find the number of page faults.2. Optimal Page replacement: In this algorithm, pages are replaced which would not be used for the longest duration of time in the future. Example-2: Consider the page references 7, 0, 1, 2, 0, 3, 0, 4, 2, 3, 0, 3, 2, 3 with 4 page frame. Find number of page fault.

**In Least Recently Used (LRU);- algorithm is a Greedy algorithm where the page to be replaced is least recently used. The idea is based on locality of reference, the least recently used page is not likely   
Let say the page reference string 7 0 1 2 0 3 0 4 2 3 0 3 2 . Initially we have 4 page slots empty.   
Initially all slots are empty, so when 7 0 1 2 are allocated to the empty slots —> 4 Page faults   
0 is already there so —> 0 Page fault.   
when 3 came it will take the place of 7 because it is least recently used —>1 Page fault   
0 is already in memory so —> 0 Page fault.   
4 will takes place of 1 —> 1 Page Fault   
Now for the further page reference string —> 0 Page fault because they are already available in the memory.**

**Storage Management is defined as it refers to the management of the data storage equipment’s that are used to store the user/computer generated data. Hence it is a tool or set of processes used by an administrator to keep your data and storage equipment’s safe. Storage management is a process for users to optimize the use of storage devices and to protect the integrity of data for any media on which it resides and the category of storage management generally contain the different type of subcategories covering aspects such as security, virtualization and more, as well as different types of provisioning or automation, which is generally made up the entire storage management software market.**

| **Parameter** | **PREEMPTIVE SCHEDULING** | **NON-PREEMPTIVE SCHEDULING** |
| --- | --- | --- |
| **Basic** | **In this resources(CPU Cycle) are allocated to a process for a limited time.** | **Once resources(CPU Cycle) are allocated to a process, the process holds it till it completes its burst time or switches to waiting state.** |
| **Interrupt** | **Process can be interrupted in between.** | **Process can not be interrupted until it terminates itself or its time is up.** |
| **Starvation** | **If a process having high priority frequently arrives in the ready queue, a low priority process may starve.** | **If a process with a long burst time is running CPU, then later coming process with less CPU burst time may starve.** |
| **Overhead** | **It has overheads of scheduling the processes.** | **It does not have overheads.** |
| **Flexibility** | **flexible** | **rigid** |
| **Cost** | **cost associated** | **no cost associated** |
| **CPU Utilization** | **In preemptive scheduling, CPU utilization is high.** | **It is low in non preemptive scheduling.** |
| **Waiting Time** | **Preemptive scheduling waiting time is less.** | **Non-preemptive scheduling waiting time is high.** |

**1 Explain in brief race condition? In this tutorial we are going to learn about Race Condition in Operating Systems.Today, we are going to learn about the most important concept in Operating Systems. The Race Condition is a condition which usually occurs in Multi Threading concept which occurs in an Operating System.The Race Condition usually occurs at the case of Tube Light which has multiple switches. This Tube Light with multiple switches is the biggest example for the Race Condition which has occurred in Operating Systems.The Race Condition also occurs in the case of processes also. If we do not take care of this Race Condition well then we might get stuck in a Deadlock too.Race condition Two types of process categories Independent process:- when one process executing it does not affect other executing process. Cooperating process: - it affect other executing process,it means share some data or variables etc.**

**Critical Section Problem in OS (Operating System);-Critical Section is the part of a program which tries to access shared resources. That resource may be any resource in a computer like a memory location, Data structure, CPU or any IO device.The critical section cannot be executed by more than one process at the same time; operating system faces the difficulties in allowing and disallowing the processes from entering the critical section.The critical section problem is used to design a set of protocols which can ensure that the Race condition among the processes will never arise.In order to synchronize the cooperative processes, our main task is to solve the critical section problem. We need to provide a solution in such a way that the following conditions can be satisfied.Requirements of Synchronization mechanismsPrimary**

1. **Mutual Exclusion;-Our solution must provide mutual exclusion. By Mutual Exclusion, we mean that if one process is executing inside critical section then the other process must not enter in the critical section.Progress;-Progress means that if one process doesn't need to execute into critical section then it should not stop other processes to get into the critical section.**

### Secondary

1. **Bounded Waiting;-We should be able to predict the waiting time for every process to get into the critical section. The process must not be endlessly waiting for getting into the critical section.**
2. **Architectural Neutrality;-Our mechanism must be architectural natural. It means that if our solution is working fine on one architecture then it should also run on the other ones as well.**

**Semaphores are just normal variables used to coordinate the activities of multiple processes in a computer system. They are used to enforce mutual exclusion, avoid race conditions, and implement synchronization between processes.**

**The process of using Semaphores provides two operations: wait (P) and signal (V). The wait operation decrements the value of the semaphore, and the signal operation increments the value of the semaphore. When the value of the semaphore is zero, any process that performs a wait operation will be blocked until another process performs a signal operation. Semaphores are of two types:**

1. **Binary Semaphore –   
   This is also known as a mutex lock. It can have only two values – 0 and 1. Its value is initialized to 1. It is used to implement the solution of critical section problems with multiple processes.**
2. **Counting Semaphore –   
   Its value can range over an unrestricted domain. It is used to control access to a resource that has multiple instances.**

**Synchronization**

**In this article, we will see a number of classical problems of**[synchronization](https://www.geeksforgeeks.org/process-synchronization-set-1/)**as examples of a large class of concurrency-control problems. In our solutions to the problems, we use**[semaphores](https://www.geeksforgeeks.org/semaphores-operating-system/)**for synchronization, since that is the traditional way to present such solutions. However, actual implementations of these solutions could use**[mutex](https://www.geeksforgeeks.org/mutex-vs-semaphore/)**locks instead of binary semaphores.**

## Synchronization Problems

**These problems are used for testing nearly every newly proposed synchronization scheme. The following problems of synchronization are considered as classical problems:**

**1. Bounded-buffer (or Producer-Consumer) Problem,  
2. Dining-Philosophers Problem,  
3. Readers and Writers Problem,  
4. Sleeping Barber Problem**

**These are summarized, for detailed explanation, you can view the linked articles for each.**

### **Bounded-Buffer (or Producer-Consumer) Problem**

**The**[Bounded Buffer problem](https://www.geeksforgeeks.org/producer-consumer-solution-using-semaphores-java/)**is also called the producer-consumer problem. This problem is generalized in terms of the Producer-Consumer problem. The solution to this problem is, to create two counting semaphores “full” and “empty” to keep track of the current number of full and empty buffers respectively. Producers produce a product and consumers consume the product, but both use of one of the containers each time.**

**Storage Management :- refers to the processes that help make data storage easier through software or techniques. It tries to improve and maximize the efficiency of data storage resources. Storage management processes can deal with local or external storage such as USBs, SDDs, HDD, the Cloud, etc., Storage management techniques or software can be divided into the following four subsets: 1. Performance, 2. Availability, 3. Recoverability, and 4. Capacity.**

**File concept: • A file is a large collection of related information that is recorded on secondary storage such as magnetic disks, magnetic tapes and optical disks. • In general, a file is a sequence of bits, bytes, lines or records whose meaning is defined by the files creator and user. • Computer store information in storage media such as disk, tape drives, and optical disks. The operating system provides a logical view of the information stored in the disk. This logical storage unit is a “file”. • The information stored in files are non-volatile, means they are not lost during power failures. A file is named collection of related information that is stored on physical storage. • Every file has a structure defined by its owner or creator and depends on the file type. • Also, file is a method of data collection that is used as a medium for giving input and receiving output from the program. • Every File has a logical location where they are located for storage and retrievalFile Structure: • File Structures is the Organization of Data in Secondary Storage Device in such a way that, it minimizes the file access time and the file storage space. • A File Structure is a combination of representations for data in files and operations for accessing the data. • A File Structure allows applications to read, write and modify the data. • A File Structure should be according to a required format that the operating system can understand. • A File has a certain defined structure according to its type like- ¬ A text file is a sequence of characters organized into lines. ¬ A source file is a sequence of procedures and functions. ¬ An object file is a sequence of bytes organized into blocks that are understandable by the machine. • When operating system defines different file structures, it also contains the code to support these file structure. • Unix, MS-DOS support minimum number of file structure whereas LINUX, MS-Windows supports for large number of file structure.**

**File Type: • File type refers to the ability of the operating system to distinguish different types of file such as text files, source files, picture file, audio file, binary files etc. • In short- depending on file type, operating system knows owner or creator of file. That is if file type has extension .doc then OS recognizes its owner is MS-Word, if file type is .cpp then OS recognizes it is CPP source program etc. • Operating system like MS-Windows and LINUX operating systems support many types of files. • Operating system like MS-DOS and UNIX have the following types of files − 1) Ordinary files: • These are the files that contain user information. • These may have text, databases or executable program. • The user can apply various operations on such files like add, modify, delete or even remove the entire file. 2) Directory files: • These files contain list of file names and other information related to these files.**

**File Operations: As you know that files are used to store the required information for its later uses. There are many file operations that can be perform by the computer system such asCreate operation Delete operation Open operation Close operation Read operation Write operation Append operation Seek operation Get attribute operation Set attribute operation Rename operation Let's see all these operations in details1) File Create Operation: ¬ This operation creates new file with no data. ¬ Without creating any file, there is no any operation can be performed, therefore file create operation is the first operation of the file. 2) File Open Operation: ¬ This operation opens the file before using it. 3) File Close Operation: ¬ This operation closes the opened file. ¬ When all the accesses are finished and the attributes and the disk addresses are no longer needed then file must be closed to free up the internal space. 4) File Delete Operation: ¬ This operation deletes the file when it is no longer needed. ¬ File delete operation free up the disk space that file holds. ¬ After deleting the file, it doesn't exist. 5) File Read Operation: ¬ This read operation is used to just read the data that are stored in the**

**Directory structure: (File directory) • Collection of files referred as file directory. That is, in directory number of different files can be stored. • The directory contains information about the files, including attributes, location and ownership. • The directory is itself a file, accessible by various file management routines and is managed by the operating system. • Following diagram shows directory structure of four different user**

**File allocation methods: • The process of allocating or assigning memory space for files on disk space is called as ‘File allocation”. • Files are allocated on disk spaces by operating system. • Operating systems uses following three main ways to allocate disk space for files.- 1. Contiguous Allocation method. 2. Linked Allocation method. 3. Indexed Allocation method. Let’s see these types in details1) Contiguous Allocation method.- • In contiguous file allocation, OS allocates contiguous address space (memory blocks) for each file on disk whenever file is just created. (i.e. pre-allocation of memory blocks) • In this method, OS assigns disk address for every file in linear order. • In this scheme, file allocation table is maintained for each file and that shows the starting block and the length of the file Linked Allocation method- • In linked allocation, allocation is done on an individual block basis where each block contains a pointer to the next block in the chain. As like linked list data structure. • In this case, each file has a list of links (pointers) to disk blocks. • In this scheme, file allocation table is also maintained for each file is that shows the starting block and the ending block of the file. 3) Indexed Allocation method: • The indexed allocation solves the problems of contiguous and linked allocation. • In indexed allocation, each file has its own index block which stores the addresses of disk space occupied by the file. An index block is having all pointers to files. • In this allocation, directory contains the addresses of such index blocks of files. • In this case, the file allocation table contains a separate one-level index for each file. The index has one entry for each block allocated to the file.**

**Free space management: • We know that, as file created, file system (i.e. OS) allocates some memory blocks to file. And when file is deleted then file system free or release allocated memory blocks such that it would be used for another file in future. Thus, file system is responsible to keep track of all the allocated and free blocks present on the hard disk. • Managing such free spaces which were found after deleting file is crucial task and it is called as ‘Free space management’. The free space list can be implemented using following methods: (Free space management Approach) 1) Bitmap or Bit vector 2) Linked list 3) Grouping Let’s see these in details1) Bitmap or Bit vector- • In this approach, the free space list is implemented as a bit map vector. • It contains the number of bits where each bit represents each block. • If the block is empty then the bit is 1 otherwise it is 0. Initially all the blocks are empty therefore each bit in the bit map vector contains 1. Linked list: • It is another approach for free space management. In this approach all free blocks are linked together and keeping a pointer in the cache which points to the first free block. • Therefore, all the free blocks on the disks will be linked together with a pointer. Whenever a block gets allocated, its previous free block will be linked to its next free block. Grouping: • This is another free space management technique. In this approach, the addresses of all free blocks are maintained in a group such that in future free block will easily available for allocation.**

**Disk Management: • Concept: We know that, in a computer there are different processes requests for I/O operation (Disk request) at a time. In such request some processes demands for write operation and some processes demands for read operation. Such multiple Read/Write operations request needs to be managed & it is called as ‘Disk management’. Disk management is done by disk controller i.e. by operating system. • Basic of Disk (Secondary Memory): • We know that, secondary storage devices are those devices whose memory is nonvolatile i.e. stored data will not lost even if the system is turned offDisk Structure: In modern computers, most of the secondary storage is in the form of magnetic disks. Hence, knowing the structure of a magnetic disk is necessary to understand how the data in the disk is accessed by the computer. • Tracks of the same distance from center forms a cylinder. • A read-write head is used to read data from a sector of the magnetic disk. • Read-Write(R-W) head moves over the rotating hard disk. This Read-Write head performs all the read and write operations on the disk and hence, position of the R-W head is a major concern. • To perform a read or write operation on a memory location, we need to place the R-W head over that position.**

**Disk Scheduling: • Disk scheduling is done by operating systems to schedule I/O requests arriving for the disk. Disk scheduling is also known as I/O scheduling. Importance of disk scheduling: 1) Multiple I/O requests may arrive by different processes and only one I/O request can be served at a time by the disk controller. Thus other I/O requests need to wait in the waiting queue and need to be scheduled. 2) Two or more request may be far from each other so can result in greater disk arm movement. 3) We know that, hard drives are one of the slowest parts of the computer system and thus need to be accessed in an efficient manner & that’s why disk scheduling is necessary. Disk scheduling Algorithms: • Disk scheduling algorithms are useful to the disk controller to schedule multiple I/O requests. We have to learn following disk scheduling algorithms1) First Come First Serve (FCFS) algorithm 2) Shortest Seek Time First (SSTF) algorithm 3) Elevator (SCAN) 4) Circular SCAN (CSCAN) Let’s see these algorithms in details1) First Come First Serve (FCFS) algorithm: • FCFS is the simplest of all the Disk Scheduling Algorithms. • In FCFS algorithm, the I/O requests are granted in the FIFO order as they arrive in the disk queue. 2) Shortest Seek Time First (SSTF) algorithm: • In SSTF (Shortest Seek Time First), requests having shortest seek time are executed first. • So, the seek time of every request is calculated in advance in the queue and then they are scheduled according to their calculated seek time. • As a result, the request near the disk arm will get executed first. SSTF is certainly an improvement over FCFS as it decreases the average response time and increases the throughput (amount of time done in specific time unit) of system. 3) SCAN algorithm: (Elevator algorithm) • In SCAN algorithm the disk arm moves into a particular direction and services the requests coming in its path and after reaching the end of disk, it reverses its direction and again services the request arriving in its path. • So, this algorithm works as an elevator and hence also known as ‘elevator’ algorithm. As a result, the requests at the midrange are serviced more and those arriving behind the disk arm will have to wait. 4) CSCAN algorithm: (Circular SCAN algorithm) • We know that, in SCAN algorithm, the disk arm again scans the path that has been scanned, after reversing its direction. So, it may be possible that too many requests are waiting at the other end and also generates overheads. These situations are avoided in CSCAN algorithm. • In this algorithm, the disk arm (R/W head) first goes to one end of disk while doing so it process all requests that comes under this movement. After reaching at one end, it then goes towards other remaining end. After reaching other end it again start scanning & completes remaining requests towards the previous end.**

**Disk Formatting: • Disk formatting is a process to configure the data-storage devices such as hard-drive, floppy disk and flash drive when we are going to use them for the very first time. • Disk formatting is usually required when new operating system is going to be used by the user. • It is also done when there is space issue and we require additional space for the storage of more data in the drives. When we format the disk then the existing files within the disk is also erased. • We can perform disk formatting on both magnetic platter hard-drives and solid-state drives. • When we are going to use hard-drive for initial use it is necessary to search for virus and bad sectors within the drive. Disk formatting has ability to erase the bad applications, viruses and also repairs bad sectors of drive**

**Boot Blocks: • Boot blocks are those memory blocks located on hard disk that helps to start computer initially i.e. after power ON. • The boot blocks have initial program stored on it that helps a computer to start running. And this initial program is known as ‘bootstrap’. • Bootstrap program initializes all aspects of the system, from CPU registers to device controllers and the contents of the main memory, and then starts the operating system.Bad Blocks: • Bad Block is an area of storing devices that is no longer useful for the storage of data because it is completely damaged or corrupted. • We know that to data access the data from the disk, disk moves fastly and in this case it will have small tolerances, this causes to failure of blocks and such block is called ‘Bad block’**

**) Deadlock Detection and Recovery: • This approach let the processes fall in deadlock and then periodically check whether deadlock occur in the system or not. If it occurs then it applies some of the recovery methods on the system to resolve deadlock. • In this approach, The OS doesn't apply any mechanism to avoid or prevent the deadlocks. Therefore the system considers that the deadlock will definitely occur. In order to get rid of deadlocks, The OS periodically checks the system for any deadlock. In case, it finds any of the deadlock then the OS will recover the system using some recovery techniques. • The main task of the OS is detecting the deadlocks. And we know that, OS can detect the deadlocks with the help ofResource allocation graph.• A deadlock happens in operating system when two or more processes needs some resource to complete their execution but that resource is already held by the otherprocess. • Deadlock is a situation that occurs in OS when any process enters a waiting state because another process is holding the demanded resource. Deadlock is a common problem in multiprocessing where several processes share a specific type of mutually exclusive resource known as a ‘soft lock’ or ‘software.’ Deadlock Characterization: (Necessary Conditions for deadlock) • We know that a deadlock happens in operating system when two or more processes need some resource to complete their execution that is held by the other process. • A deadlock occurs if the four conditions (Mutual exclusion, hold and wait, no preemption, circular wait) are true. They are given as follows − 1) Mutual Exclusion: • There should be a resource that can only be held by one process at a time. In the diagram below, there is a single instance of Resource 1 and it is held by Process1 only.**

**Multiprogramming in an operating system as the name suggests multi means more than one and programming means the execution of the program. when more than one program can execute in an operating system then this is termed a multiprogramming**[**operating system**](https://www.geeksforgeeks.org/operating-systems/)**.Advantages of threads;We can execute multiple tasks of an application at a time.Reduces the complexity of a big applications.Helps to improve the performance of an application drastically.Utilizes the max resources of multiprocessor systems.Better user interface in case of GUI based applications.Reduces the development time of an application.swapping is a memory management scheme in which any process can be temporarily swapped from main memory to secondary memory so that the main memory can be made available for other processes. It is used to improve main memory utilization. In secondary memory, the place where the swapped-out process is stored is called swap space.**

### Monolithic Structure of Operating System;-The monolithic operating system is a very basic operating system in which file management, memory management, device management, and process management are directly controlled within the kernel. The kernel can access all the resources present in the system. In monolithic systems, each component of the operating system is contained within the kernel. Operating systems that use monolithic architecture were first time used in the 1970s. Monolithic kernelA monolithic kernel is an operating system architecture where the entire operating system is working in kernel space. The monolithic model differs from other operating system architectures, such as the microkernel architecture, in that it alone defines a high-level virtual interface over computer hardware.Advantages of Monolithic KernelHere are the following advantages of a monolithic kernel, such as:The execution of the monolithic kernel is quite fast as the services such as memory management, file management, process scheduling, etc., are implemented under the same address space.A process runs completely in single address space in the monolithic kernel.The monolithic kernel is a static single binary file.Disadvantages of Monolithic Kernel;-Here are some disadvantages of the monolithic kernel, such as:If any service fails in the monolithic kernel, it leads to the failure of the entire system.The entire operating system needs to be modified by the user to add any new service.Features of Monolithic System;-The monolithic operating system provides the following features to the users, such as:****Simple structure:**** This type of operating system has a simple structure. All the components needed for processing are embedded into the kernel.****Works for smaller tasks:**** It works better for performing smaller tasks as it can handle limited resources.

**Advantages of Paging in Operating System;-Efficient use of memory – Paging allows the operating system to divide memory into smaller pieces, called pages, which can be allocated to different programs and processes as needed. This means that memory can be used more efficiently, reducing the amount of wasted space and improving overall performance.Protection and security – Paging also provides a layer of protection and security for the operating system and the programs running on it. Each page can be assigned its own protection level, preventing unauthorized access and ensuring that sensitive data is kept secure.Reduced fragmentation – Fragmentation occurs when files or programs are stored in scattered pieces across the hard drive, making it difficult for the operating system to access them quickly. Paging helps reduce fragmentation by allocating memory in a more organized and efficient manner.Flexibility and scalability – Paging also provides flexibility and scalability, allowing the operating system to allocate memory resources to different programs and processes as needed. This means that as the demands on the computer change, the operating system can adjust the allocation of memory accordingly, ensuring optimal performance.Virtual memory – Finally, paging enables the use of virtual memory, which is a technique that allows the computer to use more memory than is physically available. This is accomplished by temporarily storing data on the hard drive and swapping it in and out of memory as needed. Virtual memory helps ensure that the computer can run more programs and handle more data without running out of memory.**

# Memory Allocation in Operating System;-Allocating space to software applications is referred to as memory allocation. Memory is a sizable collection of bytes. Contiguous and non-contiguous memory allocation are the two basic types of memory allocation. Contiguous memory allocation enables the tasks to be finished in a single memory region. Contrarily, non-contiguous memory allocation distributes the procedure throughout many memory locations in various memory sections.An operating system memory allocation method is contiguous memory allocation. What, however, is memory allocation? A software or process requires memory space in order to be run. As a result, a process must be given a specific amount of memory that corresponds to its needs. Memory allocation is the term for this procedure.

# READERS WRITERS PROBLEM;-The readers-writers problem is a classical problem of process synchronization, it relates to a data set such as a file that is shared between more than one process at a time. Among these various processes, some are Readers - which can only read the data set; they do not perform any updates, some are Writers - can both read and write in the data sets.The readers-writers problem is used for managing synchronization among various reader and writer process so that there are no problems with the data sets, i.e. no inconsistency is generated.Let's understand with an example - If two or more than two readers want to access the file at the same point in time there will be no problem.