



# OPERATING SYSTEM

Ashirbad Swain

## Operating System Notes

Operating system is a system software. It acts as an intermediary/interface between hardware and the user.

It provides an environment to the user so that, the user can perform its task in convenient and efficient way.

It responsible for the execution of all the process, Resource allocation, CPU management, File management and many other tasks.

An operating system is a collection of software that manages computer hardware and provides services for programs. Specifically, it hides hardware complexity, manages computational resources, and provides isolation and protection.

There are 3 key elements of an operating system, which are:

1. Abstractions (process, thread, file, socket, memory)
2. Mechanisms (create, schedule, open, write, allocate)
3. Policies (LRU, EDF)

There are 2 operating system design principles, which are:

1. Separation of mechanism and policy by implementing flexible mechanism to support policies
2. Optimize for common case: where will the OS be used? What will the user want to execute on that machine? What are the workload requirements?

The 3 types of OS commonly used are,

1. Monolithic OS: where the entire OS is working in kernel space and is alone in supervisor mode.
2. Modular OS: in which some part of the system core will be located in independent files called modules that can be added to the system at run time.
3. Micro OS: where the kernel is broken down into separate processes, known as servers. Some of the servers run in kernel space and some run-in user-space.

Some of the important topics every developer should know,

1. Process & Process Management
2. Threads & Concurrency
3. Scheduling
4. Memory Management
5. Inter-Process Communication
6. I/O Management
7. Virtualization
8. Distributed File Systems
9. Distributed Shared Memory
10. Cloud computing

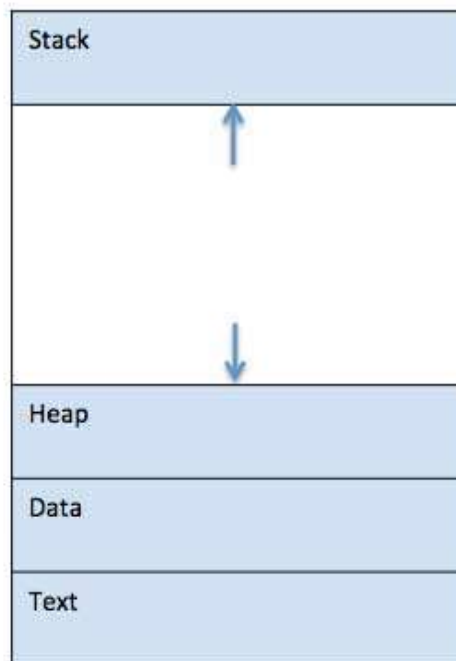
## Process & Process Management

A process basically a program in execution.

When we write a computer program in a text file (independent of the programming language or OS) and then execute the program, it becomes a process within the system which performs all the instructions defined in the program. Once the program is loaded into memory, and it becomes a process, the operating system creates a Process Control Block (PCB) for it. The PCB includes an integer identifying the process (PID), as well as several other pieces of information needed to keep track of it.

When a program is loaded into the memory and it becomes a process, it can be divided into four sections — stack, heap, text and data.

The following image shows a simplified layout of a process inside main memory.



Stack: The process stack contains the temporary data such as method/function parameters, return address and local variables.

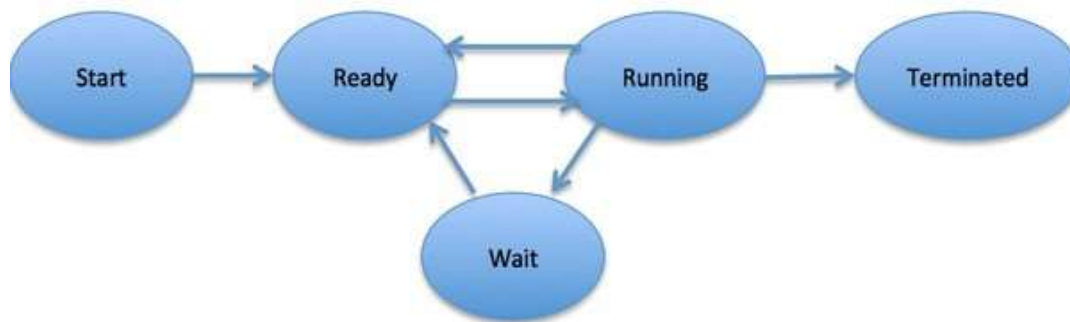
Heap: This is dynamically allocated memory to a process during its runtime.

Text: This includes the current activity represented by the value of program counter and the contents of the processor's register.

Data: This section contains the global and static variables.

### Life Cycle of Process

When a process executes, it passes through different states and the states are,



#### Start:

This is the initial state when a process is first started/created.

#### Ready:

The process is waiting to be assigned to a processor. Ready processes are waiting to have the processor allocated to them by the operating system so that they can run. Process may come into this state after Start state or while running it by but interrupted by the scheduler to assign CPU to some other process.

#### Running:

Once the process has been assigned to a processor by the OS scheduler, the process state is set to running and the processor executes its instructions.

#### Waiting:

Process moves into the waiting state if it needs to wait for a resource, such as waiting for user input, or waiting for a file to become available.

#### Terminated or Exit:

Once the process finishes its execution, or it is terminated by the operating system, it is moved to the terminated state where it waits to be removed from main memory.

## PCB (Process Control Block)

A Process Control Block is a data structure maintained by the Operating System for every process. The PCB is identified by an integer process ID (PID).

A PCB keeps all the information needed to keep track of a process as listed below:

Process ID
State
Pointer
Priority
Program counter
CPU registers
I/O information
Accounting information
etc....

### Process State

The current state of the process i.e., whether it is ready, running, waiting or whatever.

### Process Privileges

This is required to allow/disallow access to system resources.

### Process ID

Unique identification for each of the process in the operating system.

### Pointer

A pointer to parent process.

### Program Counter

Program counter is a pointer to the address of the next instruction to be executed for this process.

### CPU Registers

Various CPU registers where process need to be stored for execution for running state.

### CPU Scheduling Information

Process Priority and other scheduling information which is required to schedule the process.

### Memory Management Information

This includes the information of page table, memory limits, segment table depending on memory used by the operating system.

### Accounting Information

This includes the amount of CPU used for process execution, time limits, execution ID etc.

### IO Status Information

This includes a list of I/O devices allocated to the process.

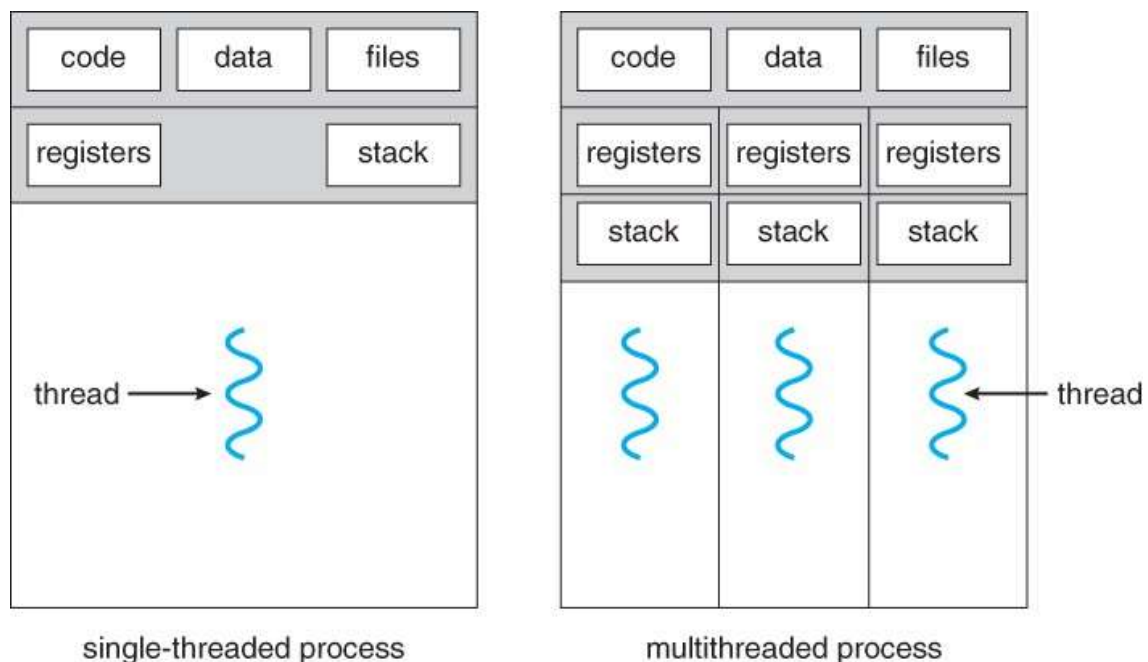
## Threads & Concurrency

A thread is a flow of execution through the process code, with its own program counter that keeps track of which instruction to execute next, system registers which hold its current working variables, and a stack which contains the execution history.

A thread shares with its peer threads few information like code segment, data segment and open files. When one thread alters a code segment memory item, all other threads see that.

A thread is also called a lightweight process. Threads provide a way to improve application performance through parallelism. Threads represent a software approach to improving performance of operating system by reducing the overhead thread is equivalent to a classical process.

Each thread belongs to exactly one process and no thread can exist outside a process. Each thread represents a separate flow of control. Threads have been successfully used in implementing network servers and web server. They also provide a suitable foundation for parallel execution of applications on shared memory multiprocessors.



### Advantages of Thread

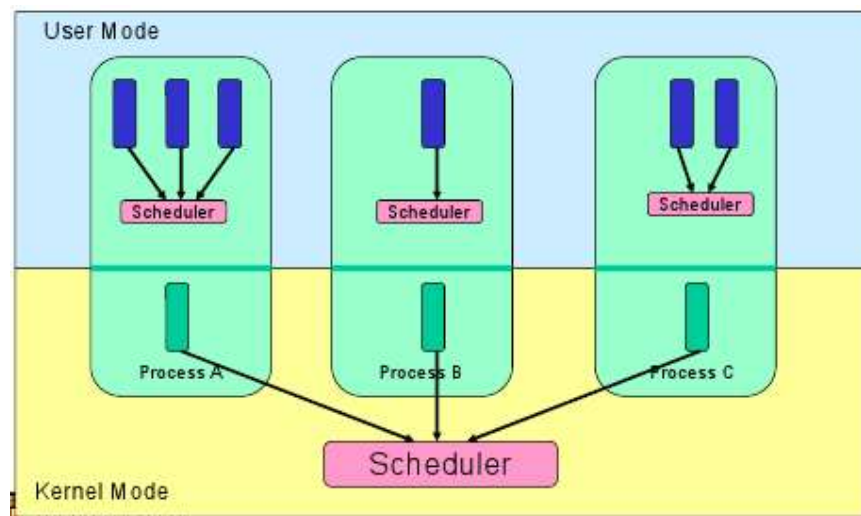
- Threads minimize the context switching time.
- Use of threads provides concurrency within a process.
- Efficient communication.
- It is more economical to create and context switch threads.
- Threads allow utilization of multiprocessor architecture to a greater scale and efficiency.

Threads are implemented in following 2 ways.

1. User Level Threads: user managed threads
2. Kernel Level Threads: operating system managed threads acting on kernel, an operating system core.

### User Level Threads

In this case, the thread management kernel is not aware of the existence of threads. The thread library contains code for creating and destroying threads, for passing message and data between threads, for scheduling thread execution and for saving and restoring thread contexts. The application starts with a single thread.



### Advantages

- Thread switching does not require kernel mode privileges.
- User Level thread can run on any operating system.
- Scheduling can be application specific in the user level thread.
- User level threads are fast to create and manage.

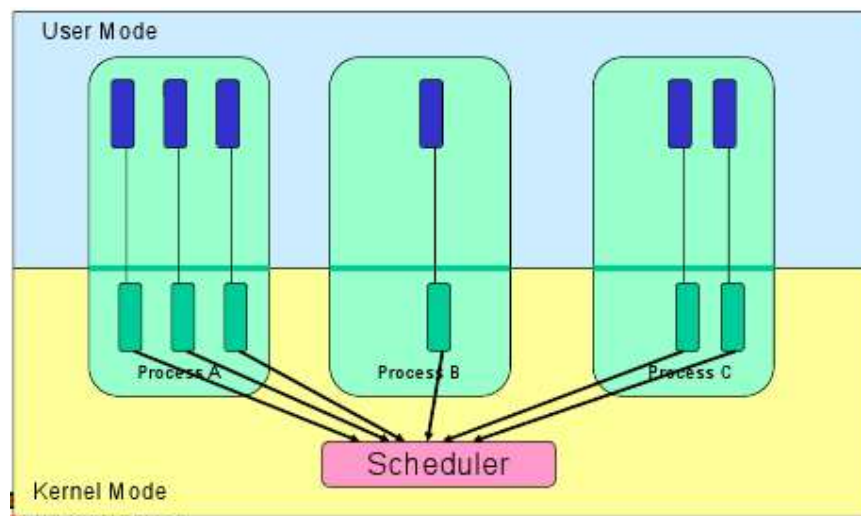
### Disadvantages

- In a typical operating system, most system calls are blocking.
- Multithreaded application cannot take advantage of multiprocessing.

## Kernel Level Threads

In this case, thread management is done by the Kernel. There is no thread management code in the application area. Kernel threads are supported directly by the operating system. Any application can be programmed to be multithreaded. All of the threads within an application are supported within a single process.

The Kernel maintains context information for the process as a whole and for individuals' threads within the process. Scheduling by the Kernel is done on a thread basis. The Kernel performs thread creation, scheduling and management in Kernel space. Kernel threads are generally slower to create and manage than the user threads.



### Advantages

- Kernel can simultaneously schedule multiple threads from the same process on multiple processes.
- If one thread in a process is blocked, the kernel can schedule another thread of the same process.
- Kernel routines themselves can be multithreaded.

### Disadvantages

- Kernel threads are generally slower to create and manage than the user threads.
- Transfer of control from one thread to another within the same process require a mode switch to the kernel.



## Scheduling

At its core this is a really simple concept... scheduling is the means in which work is allocated to resources for completion. "Work" can be a multitude of different tasks such as virtual computation elements like threads, and processes, which then need to be scheduled onto hardware resources such as processors, network links, and expansion cards.

Scheduling is primarily implemented to ensure that computer resources are kept busy (load balancing), and users are able to share system resources (i.e., CPU time, disk drives) affectively. The scheduler is a module in the operating system which selects the next jobs to be loaded into the system and the next process to execute.

Important scheduling concepts to understand —

- Schedulers can be broken down into long-term, medium-term and short-term depending on the lifecycle for which the functions need to be performed.
- Schedulers dictate what processes to run on a system, the degree of concurrency to be supported at any one time, and how the split between I/O intensive and CPU intensive processes is to be handled.
- An I/O-bound process is one that spends more of its time doing I/O than it spends doing computations. A CPU-bound process, in contrast, generates I/O requests infrequently, using more of its time doing computations.
- A dispatcher is a scheduling module that gives control of the CPU to the process selected by the short-term scheduler. It receives control in kernel mode as the result of an interrupt or system call.
- Several algorithms exist for implementing scheduling (i.e. distributing resources among components which simultaneously and asynchronously request them).

The main function of scheduling in the operating system is to provide a mechanism for sharing CPU time amongst threads and processes.

## Memory Management

Memory management is the functionality of an operating system which handles or manages primary memory and moves processes back and forth between main memory and disk during execution. Memory management keeps track of each and every memory location, regardless of whether it is allocated to some process or it is free. It checks how much memory is to be allocated to processes. It decides which process will get memory at what time. It tracks whenever some memory gets freed or unallocated and correspondingly it updates the status.

### Quick refresher:

- A hard disk or "hard drive" is a physical set of magnetic discs that can hold several gigabytes of data. Disk space refers to how much space is available on the hard disk for storing files (i.e. save a document, or install a new program).
- Memory refers to the random-access memory (RAM) inside the computer. Physically it is a set of small chips known as memory modules which are used to store actively running programs on the computer. For example, when the computer boots up, the operating system's interface and other start-up processes are loaded into main memory... another example is when you open a program like Microsoft Word.

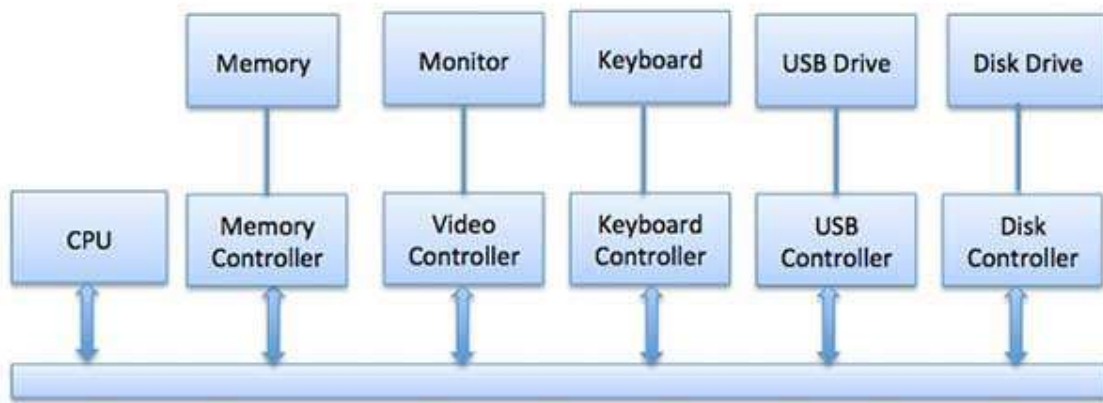
In the OS, memory management keeps track of every single memory location (both virtual and physical), whether allocated to a process or free. It checks how much memory is to be allocated to each process and decides when the process will get the memory. RAM can be accessed hundreds of times faster than the hard drive, which is why active processes are loaded into RAM, and this functionality is vital to an operating system.

## I/O Device Management

One of the important jobs of an Operating System is to manage various I/O devices including mouse, keyboards, touch pad, disk drives, display adapters, USB devices, Bit-mapped screen, LED, Analog-to-digital converter, On/off switch, network connections, audio I/O, printers etc.

An I/O system is required to take an application I/O request and send it to the physical device, then take whatever response comes back from the device and send it to the application. I/O devices can be divided into two categories:

- Block devices — A block device is one with which the driver communicates by sending entire blocks of data. For example, hard disks, USB cameras, Disk-On-Key etc.
- Character Devices — A character device is one with which the driver communicates by sending and receiving single characters (bytes, octets). For example, serial ports, parallel ports, sounds cards etc.



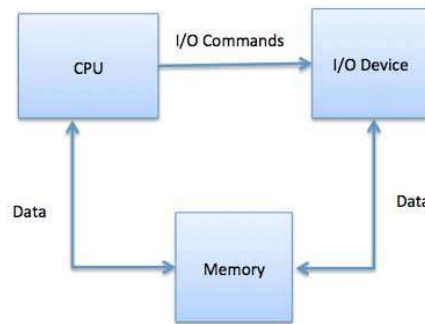
The CPU must have a way to pass information to and from an I/O device. There are three approaches available to communicate with the CPU and Device.

### 1. Special Instruction I/O

This uses CPU instructions that are specifically made for controlling I/O devices. These instructions typically allow data to be sent to an I/O device or read from an I/O device.

## 2. Memory-mapped I/O

When using memory-mapped I/O, the same address space is shared by memory and I/O devices. The device is connected directly to certain main memory locations so that I/O device can transfer block of data to/from memory without going through CPU.



While using memory mapped IO, OS allocates buffer in memory and informs I/O device to use that buffer to send data to the CPU. I/O device operates asynchronously with CPU, interrupts CPU when finished.

The advantage to this method is that every instruction which can access memory can be used to manipulate an I/O device. Memory mapped IO is used for most high-speed I/O devices like disks, communication interfaces.

## 3. Direct memory access (DMA)

Direct Memory Access (DMA) means CPU grants I/O module authority to read from or write to memory without involvement. DMA module itself controls exchange of data between main memory and the I/O device. CPU is only involved at the beginning and end of the transfer and interrupted only after entire block has been transferred.

## Virtualization

Virtualization is technology that allows you to create multiple simulated environments or dedicated resources from a single, physical hardware system. Software called a hypervisor connects directly to that hardware and allows you to split a system into separate, distinct, and secure environments known as virtual machines (VMs). These VMs rely on the hypervisor's ability to separate the machine's resources from the hardware and distribute them appropriately.

The original, physical machine equipped with the hypervisor is called the host, while the many VMs that use its resources are called guests. These guests treat computing resources — like CPU, memory, and storage — as a hanger of resources that can easily be relocated. Operators can control virtual instances of CPU, memory, storage, and other resources, so guests receive the resources they need when they need them.

Ideally, all related VMs are managed through a single web-based virtualization management console, which speeds things up. Virtualization lets you dictate how much processing power, storage, and memory to give VMs, and environments are better protected since VMs are separated from their supporting hardware and each other. Simply put, virtualization creates the environments and resources you need from underused hardware.

### Types of Virtualizations:

1. **Data Virtualization:** Data that's spread all over can be consolidated into a single source. Data virtualization allows companies to treat data as a dynamic supply — providing processing capabilities that can bring together data from multiple sources, easily accommodate new data sources, and transform data according to user needs. Data virtualization tools sits in front of multiple data sources and allows them to be treated as single source, delivering the needed data — in the required form — at the right time to any application or user.
2. **Desktop Virtualization:** Easily confused with operating system virtualization — which allows you to deploy multiple operating systems on a single machine — desktop virtualization allows a central administrator (or automated administration tool) to deploy simulated desktop environments to hundreds of physical machines at once. Unlike traditional desktop environments that are physically installed, configured, and updated on each machine, desktop virtualization allows admins to perform mass configurations, updates, and security checks on all virtual desktops.

3. **Server Virtualization:** Servers are computers designed to process a high volume of specific tasks really well so other computers — like laptops and desktops — can do a variety of other tasks. Virtualizing a server lets it to do more of those specific functions and involves partitioning it so that the components can be used to serve multiple functions.
4. **Operating System Virtualization:** Operating system virtualization happens at the kernel — the central task managers of operating systems. It's a useful way to run Linux and Windows environments side-by-side. Enterprises can also push virtual operating systems to computers, which:  
(1) Reduces bulk hardware costs, since the computers don't require such high out-of-the-box capabilities,  
(2) Increases security, since all virtual instances can be monitored and isolated, and  
(3) Limits time spent on IT services like software updates.
5. **Network Functions Virtualization:** Network functions virtualization (NFV) separates a network's key functions (like directory services, file sharing, and IP configuration) so they can be distributed among environments. Once software functions are independent of the physical machines they once lived on, specific functions can be packaged together into a new network and assigned to an environment. Virtualizing networks reduces the number of physical components — like switches, routers, servers, cables, and hubs — that are needed to create multiple, independent networks, and it's particularly popular in the telecommunications industry.

### Distributed File Systems

A distributed file system is a client/server-based application that allows clients to access and process data stored on the server as if it were on their own computer. When a user accesses a file on the server, the server sends the user a copy of the file, which is cached on the user's computer while the data is being processed and is then returned to the server.

Ideally, a distributed file system organizes file and directory services of individual servers into a global directory in such a way that remote data access is not location-specific but is identical from any client. All files are accessible to all users of the global file system and organization is hierarchical and directory-based.

### Distributed Shared Memory

Distributed Shared Memory (DSM) is a resource management component of a distributed operating system that implements the shared memory model in distributed systems, which have no physically shared memory. The shared memory provides a virtual address space that is shared among all computers in a distributed system.

In DSM, data is accessed from a shared space similar to the way that virtual memory is accessed. Data moves between secondary and main memory, as well as, between the distributed main memories of different nodes. Ownership of pages in memory starts out in some pre-defined state but changes during the course of normal operation. Ownership changes take place when data moves from one node to another due to an access by a particular process.

### Operating System Interview Questions

#### What is the main purpose of an operating system?

There are two main purposes of an operating system:

- It is designed to make sure that a computer system performs well by managing its computational activities.
- It provides an environment for the development and execution of programs.

#### What are the different operating systems?

- Batch operating systems
- Distributed operating systems
- Timesharing operating systems
- Multi-programmed operating systems
- Real-time operating systems

#### What is a socket?

A socket is used to make connection between two applications. Endpoints of the connection are called socket.

#### What is a real time system?

Real-time system is used in the case when rigid-time requirements have been placed on the operation of a processor. It contains a well-defined and fixed time constraints.

### What is kernel?

Kernel is the core and most important part of a computer operating system which provides basic services for all parts of the OS.

### What is monolithic kernel?

A monolithic kernel is a kernel which includes all operating system code in a single executable image.

### What do you mean by a process?

An executing program is known as a process. There are two types of processes:

- Operating System Processes
- User Processes

### What is the difference between micro kernel and macro kernel?

Micro kernel: micro kernel is the kernel which runs minimal performance affecting services for operating system. In micro kernel operating system all other operations are performed by processor.

Macro Kernel: Macro Kernel is a combination of micro and monolithic kernel.

### What is the concept of reentrancy?

It is a very useful memory saving technique that is used for multi-programmed time-sharing systems. It provides functionality that multiple users can share a single copy of program during the same period.

It has two key aspects:

- The program code cannot modify itself.
- The local data for each user process must be stored separately.

### What is the difference between process and program?

A program while running or executing is known as a process.

### What is the use of paging in OS?

Paging is used to solve the external fragmentation problem in operating system. This technique ensures that the data you need is available as quickly as possible.

### What is the concept of demand paging?

Demand paging specifies that if an area of memory is not currently being used, it is swapped to disk to make room for an application's need.



### What is the advantage of a multiprocessor system?

As many as processors are increased, you will get the considerable increment in throughput. It is cost effective also because they can share resources. So, the overall reliability increases.

### What is virtual memory?

Virtual memory is a very useful memory management technique which enables processes to execute outside of memory. This technique is especially used when an executing program cannot fit in the physical memory.

### What is thrashing?

Thrashing is a phenomenon in virtual memory scheme when the processor spends most of its time in swapping pages, rather than executing instructions.

### What is deadlock? Explain.

Deadlock is a specific situation or condition where two processes are waiting for each other to complete so that they can start. But this situation causes hang for both of them.

### What are the 4 necessary and sufficient conditions behind the deadlock?

These are the 4 conditions:

- 1) Mutual Exclusion Condition: It specifies that the resources involved are non-sharable.
- 2) Hold and Wait Condition: It specifies that there must be a process that is holding a resource already allocated to it while waiting for additional resource that are currently being held by other processes.
- 3) No-Pre-emptive Condition: Resources cannot be taken away while they are being used by processes.
- 4) Circular Wait Condition: It is an explanation of the second condition. It specifies that the processes in the system form a circular list or a chain where each process in the chain is waiting for a resource held by next process in the chain.

### What is a thread?

A thread is a basic unit of CPU utilization. It consists of a thread ID, program counter, register set and a stack.

### What is FCFS?

FCFS stands for First Come, First Served. It is a type of scheduling algorithm. In this scheme, if a process requests the CPU first, it is allocated to the CPU first. Its implementation is managed by a FIFO queue.

### What is SMP?

SMP stands for Symmetric Multiprocessing. It is the most common type of multiple processor system. In SMP, each processor runs an identical copy of the operating system, and these copies communicate with one another when required.

### What is RAID? What are the different RAID levels?

RAID stands for Redundant Array of Independent Disks. It is used to store the same data redundantly to improve the overall performance.

Following are the different RAID levels:

RAID 0 – Striped Disk Array without fault tolerance

RAID 1 – Mirroring and duplexing

RAID 2 – Memory-style error-correcting codes

RAID 3 – Bit-Interleaved Parity

RAID 4 – Block-Interleaved Parity

RAID 5 – Block-Interleaved distributed Parity

RAID 6 – P+Q Redundancy

### What is Banker's algorithm?

Banker's algorithm is used to avoid deadlock. It is the one of deadlock-avoidance method. It is named as Banker's algorithm on the banking system where bank never allocates available cash in such a manner that it can no longer satisfy the requirements of all of its customers.

### What is the difference between logical address and physical address space?

Logical address space specifies the address that is generated by CPU. On the other hand, physical address space specifies the address that is seen by the memory unit.

### What is fragmentation and its type?

Fragmentation is a phenomenon of memory wastage. It reduces the capacity and performance because space is used inefficiently.

There are two types of fragmentation:

- Internal fragmentation: It is occurred when we deal with the systems that have fixed size allocation units.
- External fragmentation: It is occurred when we deal with systems that have variable-size allocation units.

### What is spooling?

Spooling is a process in which data is temporarily gathered to be used and executed by a device, program or the system. It is associated with printing. When different applications send output to the printer at the same time, spooling keeps these all jobs into a disk file and queues them accordingly to the printer.

### What is the difference between internal commands and external commands?

Internal commands are the built-in part of the operating system while external commands are the separate file programs that are stored in a separate folder or directory.

### What is semaphore?

Semaphore is a protected variable or abstract data type that is used to lock the resource being used. The value of the semaphore indicates the status of a common resource.

There are two types of semaphore:

- Binary semaphores
- Counting semaphores

### What is binary semaphore?

Binary semaphore takes only 0 and 1 as value and used to implement mutual exclusion and synchronize concurrent processes.

### What is Belady's anomaly?

Belady's Anomaly is also called FIFO anomaly. Usually, on increasing the number of frames allocated to a process virtual memory, the process execution is faster, because fewer page faults occur. Sometimes, the reverse happens, i.e., the execution time increases even when more frames are allocated to the process. This is Belady's Anomaly. This is true for certain page reference patterns.

### What is starvation in OS?

Starvation is Resource management problem. In this problem, a waiting process does not get the resources it needs for a long time because the resources are being allocated to other processes.

### What is aging in OS?

Aging is a technique used to avoid the starvation in resource scheduling system.

### What are the advantages of multithreaded programming?

A list of advantages of multithreaded programming:

- Enhance the responsiveness to the users.
- Resource sharing within the process.
- Economical
- Completely utilize the multiprocessing architecture.

### What are overlays?

Overlays makes a process to be larger than the amount of memory allocated to it. It ensures that only important instructions and data at any given time are kept in memory.

### When does trashing occur?

Thrashing specifies an instance of high paging activity. This happens when it is spending more time paging instead of executing.

### How are server systems classified?

Server systems can be classified as either computer-server systems or file server systems. In the first case, an interface is made available for clients to send requests to perform an action. In the second case, provisions are available for clients to create, access and update files.

### What is RR scheduling algorithm?

RR (round-robin) scheduling algorithm is primarily aimed for time-sharing systems. A circular queue is a setup in such a way that the CPU scheduler goes around that queue, allocating CPU to each process for a time interval of up to around 10 to 100 milliseconds.

### What factors determine whether a detection-algorithm must be utilized in a deadlock avoidance system?

One is that it depends on how often a deadlock is likely to occur under the implementation of this algorithm. The other has to do with how many processes will be affected by deadlock when this algorithm is applied.

### How does dynamic loading aid in better memory space utilization?

With dynamic loading, a routine is not loaded until it is called. This method is especially useful when large amounts of code are needed in order to handle infrequently occurring cases such as error routines.

### What is Direct Access method?

Direct Access method is based on a disk model of a file, such that it is viewed as a numbered sequence of blocks or records. It allows arbitrary blocks to be read or written. Direct access is advantageous when accessing large amounts of information.

### What is the page size when designing an operating system?

The best paging size varies from system to system, so there is no single best when it comes to page size. There are different factors to consider in order to come up with a suitable page size, such as page table, paging time, and its effect on the overall efficiency of the operating system.

### When designing the file structure for an operating system, what attributes are considered?

Typically, the different attributes for a file structure are naming, identifier, supported file types, and location for the files, size, and level of protection.

### What is root partition?

Root partition is where the operating system kernel is located. It also contains other potentially important system files that are mounted during boot time.

### What are device drivers?

Device drivers provide a standard means of representing I/O devices that maybe manufactured by different companies. This prevents conflicts whenever such devices are incorporated in a systems unit.

### What are the primary functions of VFS?

VFS, or Virtual File System, separate file system generic operations from their implementation by defining a clean VFS interface. It is based on a file-representation structure known as vnode, which contains a numerical designator needed to support network file systems.

### What is the purpose of an I/O status information?

I/O status information provides information about which I/O devices are to be allocated for a particular process. It also shows which files are opened, and other I/O device state.

### What is multitasking?

Multitasking is the process within an operating system that allows the user to run several applications at the same time. However, only one application is active at a time for user interaction, although some applications can run "behind the scene".

### Explain pros and cons of a command line interface?

A command line interface allows the user to type in commands that can immediately provide results. Many seasoned computer users are well accustomed to using the command line because they find it quicker and simpler.

However, the main problem with a command line interface is that users have to be familiar with the commands, including the switches and parameters that come with it. This is a downside for people who are not fond of memorizing commands.

### What is caching?

Caching is the processing of utilizing a region of fast memory for a limited data and process. A cache memory is usually much efficient because of its high access speed.

### What is assembler?

An assembler acts as a translator for low-level language. Assembly codes written using mnemonic commands are translated by the Assembler into machine language.

### What are interrupts?

Interrupts are part of a hardware mechanism that sends a notification to the CPU when it wants to gain access to a particular resource. An interrupt handler receives this interrupt signal and "tells" the processor to take action based on the interrupt request.

### What is GUI?

GUI is short for Graphical User Interface. It provides users with an interface wherein actions can be performed by interacting with icons and graphical symbols. People find it easier to interact with the computer when in a GUI especially when using the mouse. Instead of having to remember and type commands, users click on buttons to perform a process.

### What is Preemptive multitasking?

Preemptive multitasking allows an operating system to switch between software programs. This, in turn, allows multiple programs to run without necessarily taking complete control over the processor and resulting in system crashes.

### What is NOS?

NOS is short for Network Operating System. It is a specialized software that will allow a computer to communicate with other devices over the network, including file/folder sharing.

### What is plumbing/piping?

It is the process of using the output of one program as an input to another. For example, instead of sending the listing of a folder or drive to the main screen, it can be piped and sent to a file, or sent to the printer to produce a hard copy.

### What are the basic functions of an operating system?

Operating system controls and coordinates the use of the hardware among the various applications programs for various uses. Operating system acts as resource allocator and manager. Also operating system is control program which controls the user programs to prevent errors and improper use of the computer. It is especially concerned with the operation and control of I/O devices.

### What is a process?

A program in execution is called a process.

Processes are of two types:

- Operating System processes
- User processes

### What is context switching?

Transferring the control from one process to other process requires saving the state of the old process and loading the saved state for new process. This task is known as context switching.

### What is process synchronization?

A situation, where several processes access and manipulate the same data concurrently and the outcome of the execution depends on the particular order in which the access takes place, is called race condition. To guard against the race condition, we need to ensure that only one process at a time can be manipulating the same data. The technique we use for this is called process synchronization.

### What is cache memory?

Cache memory is random access memory (RAM) that a computer microprocessor can access more quickly than it can access regular RAM. As the microprocessor processes data, it looks first in the cache memory and if it finds the data there (from a previous reading of data), it does not have to do the more time-consuming reading of data from larger memory.

### Difference between Compiler and Interpreter?

An interpreter reads one instruction at a time and carries out the actions implied by that instruction. It does not perform any translation. But a compiler translates the entire instructions

### What is Throughput, Turnaround time, waiting time and Response time?

Throughput - number of processes that complete their execution per time unit

Turnaround time - amount of time to execute a particular process

Waiting time - amount of time a process has been waiting in the ready queue

Response time - amount of time it takes from when a request was submitted until the first response is produced, not output (for time-sharing environment)

### What is Memory Management Unit (MMU)?

Hardware device that maps virtual to physical address. In MMU scheme, the value in the relocation register is added to every address generated by a user process at the time it is sent to memory.

The user program deals with logical addresses; it never sees the real physical addresses

### What is a trap and trapdoor?

Trapdoor is a secret undocumented entry point into a program used to grant access without normal methods of access authentication. A trap is a software interrupt, usually the result of an error condition.

### When is a system in safe state?

The set of dispatchable processes is in a safe state if there exists at least one temporal order in which all processes can be run to completion without resulting in a deadlock.

### Explain the concept of the Distributed systems?

Distributed systems work in a network. They can share the network resources, communicate with each other.

### What is cache-coherency?

In a multiprocessor system there exist several caches each may contain a copy of same variable A. Then a change in one cache should immediately be reflected in all other caches this process of maintaining the same value of a data in all the caches is called cache-coherency.

### What is a Long-Term Scheduler and Short-Term Scheduler?

Long term schedulers are the job schedulers that select processes from the job queue and load them into memory for execution. The short-term schedulers are the CPU schedulers that select a process from the ready queue and allocate the CPU to one of them.

### Explain the meaning of mutex?

Mutex is the short form for 'Mutual Exclusion object'. A mutex allows multiple threads for sharing the same resource. The resource can be file. A mutex with a unique name is created at the time of starting a program. A mutex must be locked from other threads, when any thread that needs the resource. When the data is no longer used / needed, the mutex is set to unlock.

### What is cycle stealing?

We encounter cycle stealing in the context of Direct Memory Access (DMA). Either the DMA controller can use the data bus when the CPU does not need it, or it may force the CPU to temporarily suspend operation. The latter technique is called cycle stealing. Note that cycle stealing can be done only at specific break points in an instruction cycle.

### What is Marshalling?

The process of packaging and sending interface method parameters across thread or process boundaries.

### What is a daemon?

Daemon is a program that runs in the background without user's interaction. A daemon runs in a multitasking operating system like UNIX. A daemon is initiated and controlled by special programs known as processes.



### What is pre-emptive and non-pre-emptive scheduling?

Pre-emptive scheduling: The pre-emptive scheduling is prioritized. The highest priority process should always be the process that is currently utilized.

Non-Pre-emptive scheduling: When a process enters the state of running, the state of that process is not deleted from the scheduler until it finishes its service time.

### What is busy waiting?

The repeated execution of a loop of code while waiting for an event to occur is called busy-waiting. The CPU is not engaged in any real productive activity during this period, and the process does not progress toward completion.

### What is page cannibalizing?

Page swapping or page replacements are called page cannibalizing.

### What is process migration?

It is the transfer of sufficient amount of the state of process from one machine to the target machine.

### Difference between Primary storage and secondary storage?

Primary memory is the main memory (Hard disk, RAM) where the operating system resides.

Secondary memory can be external devices like CD, floppy magnetic discs etc. secondary storage cannot be directly accessed by the CPU and is also external memory storage.

### Define compactions.

Compaction is a process in which the free space is collected in a large memory chunk to make some space available for processes.

### What are residence monitors?

Early operating systems were called residence monitors.

### What is dual-mode operation?

In order to protect the operating systems and the system programs from the malfunctioning programs the two mode operations were evolved

System mode

User mode.

### What is a device queue?

A list of processes waiting for a particular I/O device is called device queue.

### What are the different types of Real-Time Scheduling?

Hard real-time systems required to complete a critical task within a guaranteed amount of time. Soft real-time computing requires that critical processes receive priority over less fortunate ones.

### What is relative path and absolute path?

Absolute path-- Exact path from root directory.

Relative path-- Relative to the current path.

### What are the disadvantages of context switching?

Time taken for switching from one process to other is pure overhead. Because the system does no useful work while switching. So, one of the solutions is to go for threading whenever possible.

### What is a data register and address register?

Data registers - can be assigned to a variety of functions by the programmer. They can be used with any machine instruction that performs operations on data.

Address registers - contain main memory addresses of data and instructions or they contain a portion of the address that is used in the calculation of the complete addresses.

### What is DRAM?

Dynamic Ram stores the data in the form of Capacitance, and Static RAM stores the data in Voltages.

### What are local and global page replacements?

Local replacement means that an incoming page is brought in only to the relevant process' address space. Global replacement policy allows any page frame from any process to be replaced. The latter is applicable to variable partitions model only.

### Explain the concept of the batched operating systems?

In batched operating system the users give their jobs to the operator who sorts the programs according to their requirements and executes them. This is time consuming but makes the CPU busy all the time.

### What is SCSI?

SCSI - Small computer systems interface is a type of interface used for computer components such as hard drives, optical drives, scanners and tape drives. It is a competing technology to standard IDE (Integrated Drive Electronics).

### What is an Idle thread?

The special thread a dispatcher will execute when no ready thread is found.

### What is FtDisk?

It is a fault tolerance disk driver for Windows NT.

### What is Dispatcher?

Dispatcher module gives control of the CPU to the process selected by the short-term scheduler; this involves: Switching context, switching to user mode, jumping to the proper location in the user program to restart that program, dispatch latency – time it takes for the dispatcher to stop one process and start another running.

### When does the condition 'rendezvous' arise?

In message passing, it is the condition in which, both, the sender and receiver are blocked until the message is delivered.

### What is process spawning?

When the OS at the explicit request of another process creates a process, this action is called process spawning

### What are the reasons for process suspension?

- 1) swapping
- 2) interactive user request
- 3) timing
- 4) parent process request

### What are the sub-components of I/O manager in Windows NT?

- 1) Network redirector/ Server
- 2) Cache manager.
- 3) File systems
- 4) Network driver
- 5) Device driver

### What is a drawback of MVT?

- 1) ability to support multiple processors
- 2) virtual storage
- 3) source level debugging