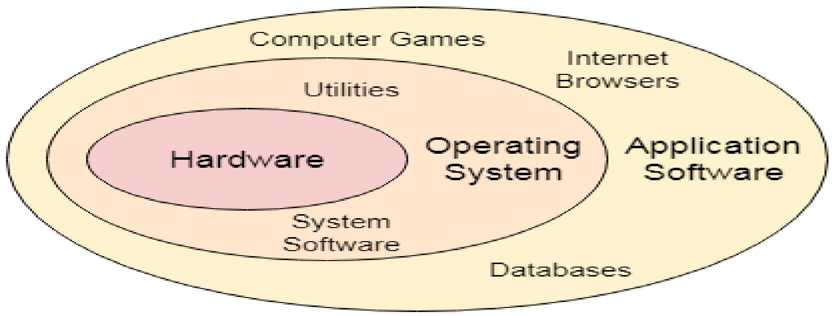
Unit-1st (Introduction)

OS Definition and functions

In the Computer System (comprises of Hardware and software), Hardware can only understand machine code (in the form of 0 and 1) which doesn't make any sense to a naive user.We need a system which can act as an intermediary and manage all the processes and resources present in the system.

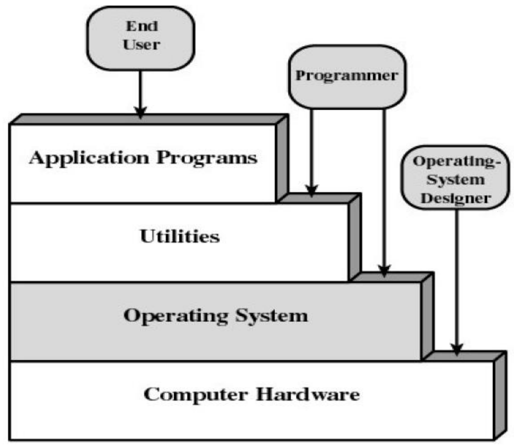


An **Operating System** can be defined as an **interface between user and hardware**. It is responsible for the execution of all the processes, Resource Allocation, CPU management, File Management and many other tasks.The purpose of an operating system is to provide an environment in which a user can execute programs in convenient and efficient manner.

Operating System: Objectives and Functions

OS as a user / computer interface– Usability- The reason for an operatingsystem to exist is to make computers more convenient to use. An OS aims to wrapthe underneath hardware resources and provides services to end users in asystematic way. These services may be divided into two types: services directlyavailable for end users through all kinds of I/O devices, such as mouse, keyboard,monitor, printer, and so on; and services for application programs, which in turnprovides services for end users.

If we look on these services as interfaces among different components of acomputer system, then the following hierarchical architecture may be obtained:



OS as resource manager - Efficiency - It is not the OS itself but the hardware thatmakes all kinds of services possible and available to application programs. An OSmerely exploits the hardware to provide easily accessible interfaces. Exploitationmeans management upon the hardware resources, and thus also imposes controlupon or manages the entities that use the services so that the resources are usedefficiently. It involves process scheduling, memory management, I/O devicemanagement, etc

```
                              +------------------+
+-----------+                 |                  |
|           |                 |    Controlled    |
| Controller |--------->|    components    |
|           |                 |                  |
+-----------+                 +------------------+
```

Evolution of OS

Maintainability - It does not suffice to simply consider anoperating system an invariable unit. An OS may evolve while time elapses due tothe following reasons:

• **Hardware upgrades or new types of hardware:** With hardware technologiesdevelopment, the OS also needs to upgrade so as to utilize the new mechanismsintroduced by new hardware.For example, Pentium IV extended instruction set of Pentium III for multimediaapplications and Internet transmission. An OS designed for the previous versionsof Intel x86 series will have to be upgraded to be able to accommodate these newinstructions.

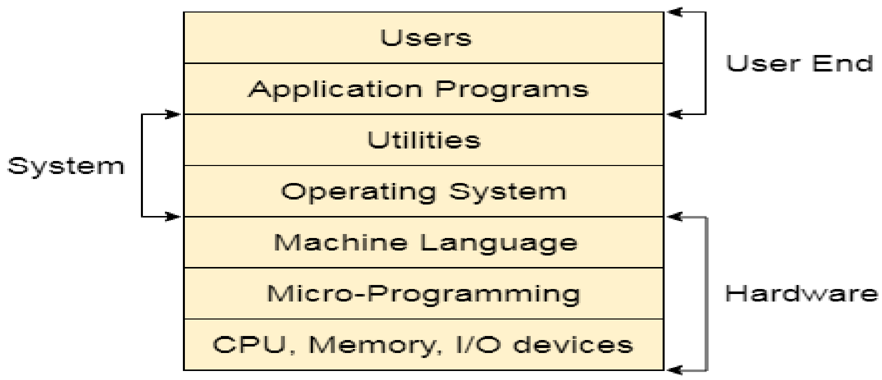
• **New services:** An OS may also expand to include more services in response touser demand.

• **Fixes:** No software is perfect, and any program may contain more or less bugs ordefects, thus fixes should be made from time to time.Microsoft Windows is a vivid example of this kind.

Structure of a Computer System

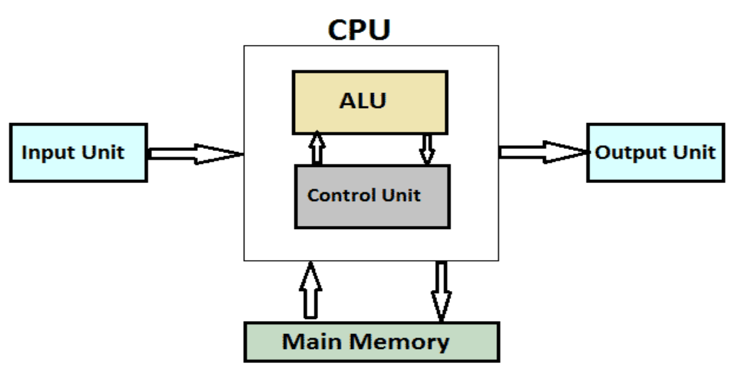
A Computer System consists of:

- Users (people who are using the computer)
- Application Programs (Compilers, Databases, Games, Video player, Browsers, etc.)
- System Programs (Shells, Editors, Compilers, etc.)
- Operating System (A special program which acts as an interface between user and hardware)
- Hardware (CPU, Disks, Memory, etc)



A computer system can be divided roughly into four components: **Hardware, Operating System, Application Programs and Users**.

1. **Hardware:** It includes:

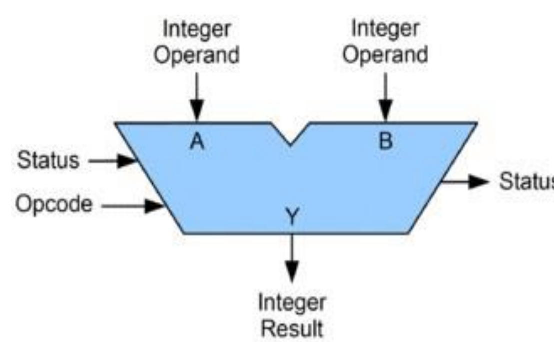


A). CPU (Central Processing Unit): -The electronic circuitrywithin a computer that carriesout the instructions of acomputer program byperforming the basic arithmetic,logical, control and input/output(I/O) operations specified by theinstructions. It is often referredas Microprocessor or brain of acomputer.

i) ALU (Arithmetic and Logical Unit): - An arithmetic-logic unit (ALU) is the part of a computer processor (CPU) that carries out arithmetic and logic operations on the operands in computer instruction words. In some processors, the ALU is divided into two units, an arithmetic unit (AU) and a logic unit (LU).

Operand: The quantity on whichan operation is to be done.

Opcode: Specifies the desiredarithmetic or logic operation tobe Performed by the ALU.

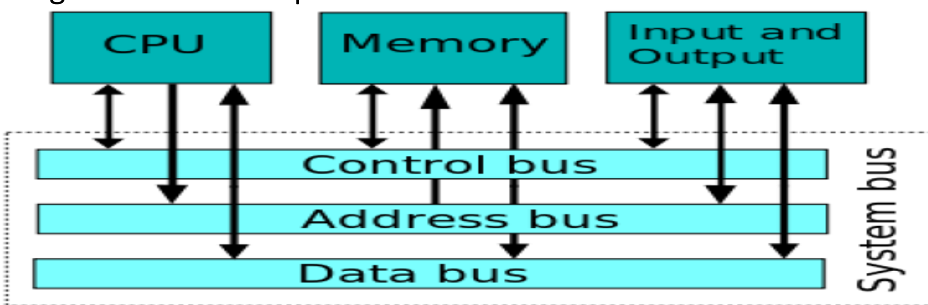


Status: These are variousindividual signals that conveySupplemental information aboutthe result of an ALU operation.

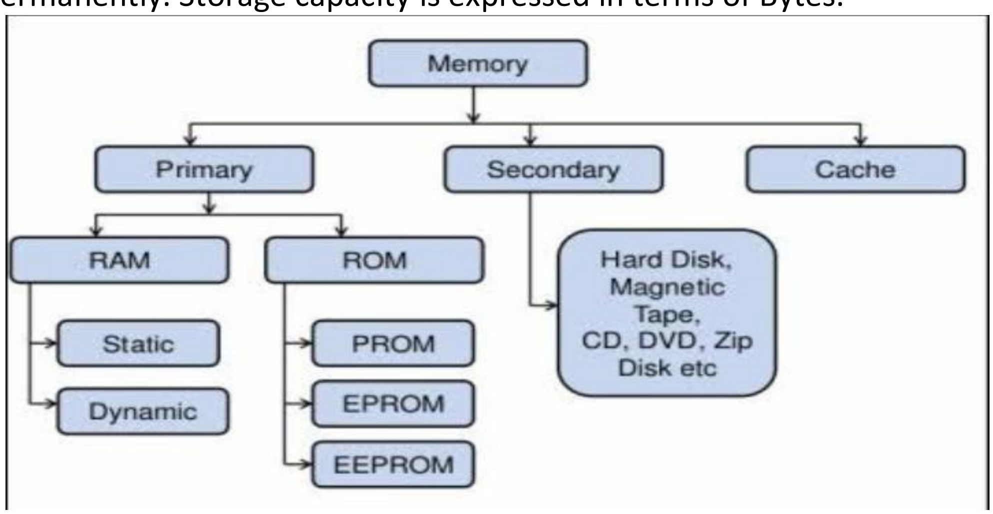
ii) CU (Control Unit): -The **control unit** (CU) is a component of a computer's central processingunit (CPU) that directs the operation of the processor. It tells the computer'smemory, arithmetic/logic unit and input and output devices on how to respond to aprogram's instructions.It directs the operation of the other units by providing timing and control signals.Most computer resources are managed by the CU. It directs the flow of databetween the CPU and the other devices.

Register: It is one of a small set of data holding places that are part of thecomputer processor. A register may hold an instruction, a storage address, or anykind of data (such as a bit sequence or individual characters).

Bus: A **Bus** is a communication system that transfers data between componentsinside a computer, or between computers. This expression covers all relatedhardware components (wire, optical fiber, etc.) and software, includingcommunication protocols.



iii) MU (Memory Unit): - A Computer memory is any physical device capable of storing information Temporarily or permanently. Storage capacity is expressed in terms of Bytes.



i) Primary Memory: Primary memory is computer memory that a processor orcomputer accesses first or directly. It allows a processor to access runningexecution applications and services that are temporarily stored in a specificmemory location.Primary memory is also known as primary storage or main memory.

RAM (Random Access Memory): A type of data storage used in computers that isgenerally located on the motherboard. This type of memory is volatile and allinformation that was stored in RAM is lost when the computer is turned off.Volatile memory is temporary memory while ROM (read-only memory) is nonvolatileand holds data permanently when the power is turned off.

SRAM	DRAM
Stores data till the power is supplied	Stores data only for few milliseconds even when power is supplied
Uses an arry of 6 transistors for each memory cell	Uses a single transistor and capaitor for each memory cell
Does not refreshes the memory Cell	Needs to refresh the memory cell after each reading of the capacitor
Data access is faster	Data acces is slower
Consume more power	Consume less power
Low density/less memory per chip	High density/more memory per chip
Cost per bit is high	Cost per bit is low

ROM (Read Only Memory): A type of non-volatile memory used in computersand other electronic devices. Data stored in ROM can only be modified slowly,with difficulty, or not at all, so it is mainly used to store firmware (software that isclosely tied to specific hardware, and unlikely to need frequent updates) orapplication software in plug-in cartridges. Eg- BIOS Chip.

Memory Type	Category	Erasure	Write Mechanism	Volatility
Random-access memory (RAM)	Read-write memory	Electrically, byte-level	Electrically	Volatile
Read-only memory (ROM)	Read-only memory	Not possible	Masks	Nonvolatile
Programmable ROM (PROM)			Electrically	
Erasable PROM (EPROM)	Read-mostly memory	UV light, chip-level		
Electrically Erasable PROM (EEPROM)		Electrically, byte-level		
Flash memory		Electrically, block-level		

ii) Secondary Memory: Secondary memory is where programs and data are kepton a long-term basis. Common secondary storage devices are the hard disk andoptical disks. Eg- Hard Disks, Floppy Disk, CD/DVD Disks.

iii) Cache Memory: Cache memory is a small-sized type of volatile computermemory that provides high-speed data access to a processor and stores frequentlyused computer programs, applications and data.

Registers

Register are used to quickly accept, store, and transfer data and instructions thatare being used immediately by the CPU. These registers are the top of the memoryhierarchy, and are the fastest way for the system to manipulate data. Registers areused by the CPU for Performing the Operations.

Register perform following functions:

1) Fetch: The Fetch Operation is used for taking the instructions those are givenby the user and the Instructions those are stored into the Main Memory will befetch by using Registers.

2) Decode: The Decode Operation is used for interpreting the Instructions meansthe Instructions are decoded means the CPU will find out which Operation is to beperformed on the Instructions.

3) Execute: The Execute Operation is performed by the CPU. And Results thoseare produced by the CPU are then Stored into the Memory and after that they aredisplayed on the user Screen.

Types of Registers:

1. Memory Address Register (MAR): This register holds the address of memorywhere CPU wants to read or write data. When CPU wants to store some data in thememory or reads the data from the memory, it places the address of the requiredmemory location in the MAR.

2. Memory Buffer Register (MBR): This register holds the contents of data orinstruction read from, or written in memory. The contents of instruction placed inthis register are transferred to the Instruction Register, while the contents of dataare transferred to the accumulator or I/O register. In other words you can say thatthis register is used to store data/instruction coming from the memory or going tothe memory.

3. I/O Address Register (I/O AR): I/O Address register is used to specify theaddress of a particular I/O device.

4. I/O Buffer Register (I/O BR): I/O Buffer Register is used for exchanging databetween the I/O module and the processor.

5. Program Counter (PC): Program Counter register is also known as InstructionPointer Register. This register is used to store the address of the next instruction tobe fetched for execution. When the instruction is fetched, the value of IP isincremented. Thus this register always points or holds the address of nextinstruction to be fetched.

6. Instruction Register (IR): Once an instruction is fetched from main memory, it isstored in the Instruction Register. The control unit takes instruction from thisregister, decodes and executes it by sending signals to the appropriate componentof computer to carry out the task.

7. Accumulator Register (AC): The accumulator register is located inside theALU, It is used during arithmetic & logical operations of ALU. The control unitstores data values fetched from main memory in the accumulator for arithmetic orlogical operation. This register holds the initial data to be operated upon, theintermediate results, and the final result of operation. The final result is transferredto main memory through MBR.

8. Flag Register (FR): The Flag register is used to indicate occurrence of a certaincondition during an operation of the CPU. It is a special purpose register with sizeone byte or two bytes. Each bit of the flag register constitutes a flag (or alarm),such that the bit value indicates if a specified condition was encountered whileexecuting an instruction. For example, if zero value is put into an arithmeticregister (accumulator) as a result of an arithmetic operation or a comparison, thenthe zero flag will be raised by the CPU. Thus, the subsequent instruction can checkthis flag and when a zero flag is "ON" it can take, an appropriate route in thealgorithm.

9. Data Register (DR): A register used in microcomputers to temporarily store databeing transmitted to or from a peripheral device.

What does an Operating system do?

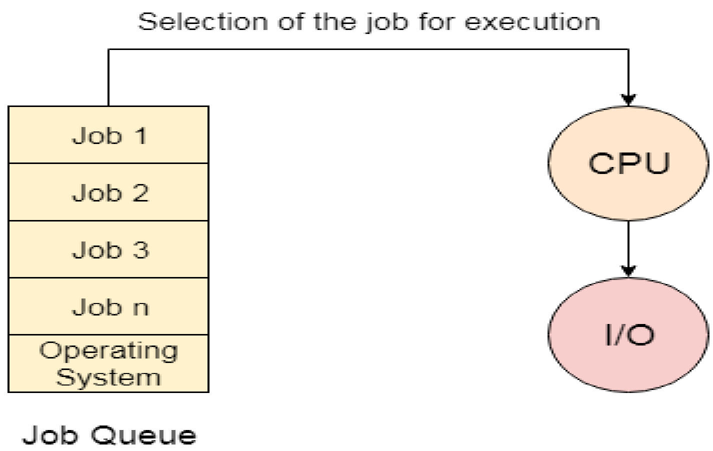
1. Process Management
2. Process Synchronization
3. Memory Management
4. CPU Scheduling
5. File Management
6. Security

Types of OS

There are many types of operating system exist in the current scenario:

1. Batch Operating System: -In the era of 1970s, the Batch processing was very popular. The Jobs were executed in batches. People were used to have a single computer which was called mainframe.In Batch operating system, access is given to more than one person; they submit their respective jobs to the system for the execution.

The system put all of the jobs in a queue on the basis of first come first serve and then executes the jobs one by one. The users collect their respective output when all the jobs get executed.



Disadvantages of Batch OS

- **Starvation: -** Batch processing suffers from starvation. If there are five jobs J1, J2, J3, J4, J4 and J5 present in the batch. If the execution time of J1 is very high then other four jobs will never be going to get executed or they will have to wait for a very high time. Hence the other processes get starved.
- **Not Interactive: -**Batch Processing is not suitable for the jobs which are dependent on the user's input. If a job requires the input of two numbers from the console then it will never be going to get it in the batch processing scenario since the user is not present at the time of execution.

2. Interactive Operating System: - In an Interactive operating system, there is a direct interaction between the user and the computer. Mostly, all personal computers use Interactive operating systems. In this kind of operating system, the user enters some command in the system and the system works according to it.

3. Time Sharing Operating System: -It is similar to the multiprogramming system with some additional extensions and also known as Multitasking OS. In a Time sharing OS, the system is capable of handling multiple jobs simultaneously and here the processing time is shared among all the users.
With Time sharing OS, users at different locations or terminals can access the same computer at the same time. Here, the CPU uses the switching mechanism that helps it to switch from one job to another so that each job gets equal and processing time.

4. Real Time Operating System: -In Real Time systems, each job carries a certain deadline within which the Job is supposed to be completed, otherwise the huge loss will be there or even if the result is produced then it will be completely useless.

The Application of a Real Time system exists in the case of military applications, if you want to drop a missile then the missile is supposed to be dropped with certain precision.

5. Multiprocessor Operating System: - Multiprocessor system means, there are more than one processor which work parallel to perform the required operations.It allows the multiple processors, and they are connected with physical memory, computer buses, clocks, and peripheral devices.The main objective of using a multiprocessor operating system is to increase the execution speed of the system and consume high computing power.

Advantages

The advantages of multiprocessor systems are as follows −

- If there are multiple processors working at the same time, more processes can be executed parallel at the same time. Therefore the throughput of the system will increase.
- Multiprocessor systems are more reliable. Due to the fact that there are more than one processor, in case of failure of any one processor will not make the system come to a halt. Although the system will become slow if it happens but still it will work.
- Electricity consumption of a multiprocessor system is less than the single processor system. This is because, in single processor systems, many processes have to be executed by only one processor so there is a lot of load on it. But in case of multiple processor systems, there are many processors to execute the processes so the load on each processor will be comparatively less so electricity consumed will also be less.

6. Multiuser Operating System: -A multi-user operating system is an operating system that permits several users to access a single system running to a single operating system. These systems are frequently quite complex, and they must manage the tasks that the various users connected to them require. Users will usually sit at terminals or computers connected to the system via a network and other system machines like printers. A multi-user operating system varies from a connected single-user operating system in that each user accesses the same operating system from different machines.

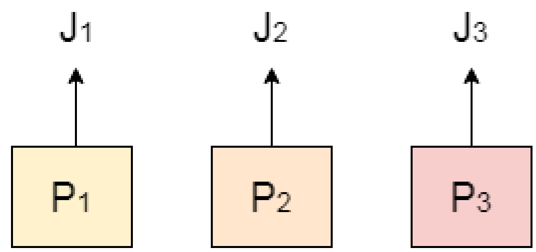
Types of Multi-User Operating System: -

Distributed System: - A distributed system is also known as distributed computing. It is a collection of multiple components distributed over multiple computers that interact, coordinate, and seem like a single coherent system to the end-user. With the aid of the network, the end-user would be able to interact with or operate them**.**

Time-Sliced Systems: -It's a system in which each user's job gets a specific amount of CPU time. In other words, each work is assigned to a specific time period. These time slices look too small to the user's eyes. An internal component known as the 'Scheduler' decides to run the next job. This scheduler determines and executes the job that must perform based on the priority cycle.

Multiprocessor System:-Multiple processors are used in this system, which helps to improve overall performance. If one of the processors in this system fails, the other processor is responsible for completing its assigned task.

7. **Multiprocessing Operating System: -** In Multiprocessing, Parallel computing is achieved. There are more than one processors present in the system which can execute more than one process at the same time. This will increase the throughput of the system.



Multi Processing

8. **Multithreaded System: -** Multithreading is the ability of a program or an operating system process to manage its use by more than one user at a time and to even manage multiple requests by the same user without having to have multiple copies of the programming running in the computer. Each user request for a program or system service (and here a user can also be another program) is kept track of as a thread with a separate identity. As programs work on behalf of the initial request for that thread and are interrupted by other requests, the status of work on behalf of that thread is kept track of until the work is completed.

Some important functions performed by an operating System

Memory Management: - Memory management refers to management of Primary Memory or MainMemory. Main memory is a large array of words or bytes where each word or bytehas its own address.Main memory provides a fast storage that can be accessed directly by the CPU.For a program to be executed, it must in the main memory. An Operating Systemdoes the following activities for memory management −

•Keeps tracks of primary memory, i.e., what part of it are in use by whom,what part is not in use.

•In multiprogramming, the OS decides which process will get memory whenand how much.

•Allocates the memory when a process requests it to do so.

•De-allocates the memory when a process no longer needs it or has beenterminated.

Processor Management:-In multiprogramming environment, the OS decides which process gets theprocessor when and for how much time. This function is called processscheduling. An Operating System does the following activities for processormanagement −

•Keeps tracks of processor and status of process. The program responsiblefor this task is known as traffic controller.

•Allocates the processor (CPU) to a process.

•De-allocates processor when a process is no longer required.

Device Management:-An Operating System manages device communication via their respective drivers.It does the following activities for device management −

•Keeps tracks of all devices. Program responsible for this task is known asthe I/O controller.

•Decides which process gets the device when and for how much time.

•Allocates the device in the efficient way.

•De-allocates devices.

File Management: -A file system is normally organized into directories for easy navigation and usage.These directories may contain files and other directions.An Operating System does the following activities for file management −

•Keeps track of information, location, uses, status etc. The collectivefacilities are often known as file system.

•Decides who gets the resources.

•Allocates the resources.

•De-allocates the resources.

Security:-

By means of password and similar other techniques, it prevents unauthorizedaccess to programs and data.

Operating System Services: - Operating Systems provide a set of essential services that support the execution of application programs and ensure efficient and safe use of hardware resources.

1. Program Execution: -

- Loads a program into memory and runs it.
- Provides runtime environment and resources.

2. I/O Operations

- Handles input and output devices.
- Provides device drivers and manages communication with hardware.

3. File System Manipulation

- Creates, deletes, reads, writes, and manages files and directories.
- Controls access permissions.

4. Communication Services

- Enables communication between processes via:
    - **Message passing**
    - **Shared memory**
- Used in both distributed and local systems.

5. Error Detection

- Continuously monitors the system for potential errors.
- Ensures correct and consistent computing.

6. Resource Allocation

- Manages CPU cycles, memory, file storage, and I/O devices.
- Ensures fair distribution among programs and users.

7. Security and Protection

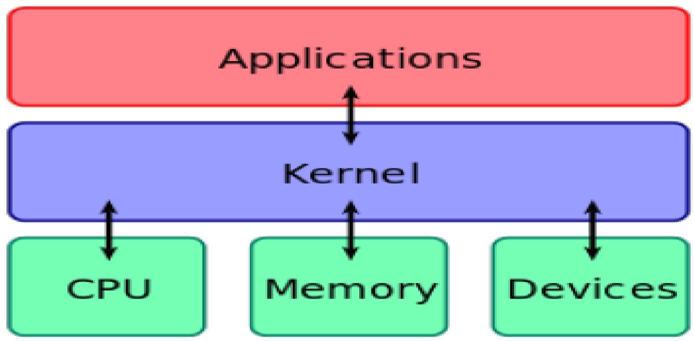
- Controls access to system resources.
- Ensures authentication, authorization, and data integrity.

8. User Interface

- Provides interfaces for users:
    - Command-Line Interface (CLI)
    - Graphical User Interface (GUI)
    - Touch interfaces

Kernel (Core of a computer's Operating System)

The kernel is a computer programs that isthe core of a computer's operatingsystem, with complete control over everything in the system. On most systems, it is one of the first programsloaded on start-up (after the bootloader).It handles the rest of start-up as well as input/output requests from software, translating them into data-processing instructions for the central processing unit. It handles memory and peripherals like keyboards, monitors, printers, and speakers.



A kernel connects the application software to the hardware of a computer.The critical code of the kernel is usually loaded into a protected area of memory,which prevents it from being overwritten by applications or other, more minorparts of the operating system.

The kernel performs its tasks, such as running processes and handling interrupts,in kernel space. In contrast, everything a user does is in user space: writing text ina text editor, running programs in a GUI, etc.

This separation prevents user data and kernel data from interfering with each otherand causing instability and slowness.The kernel's interface is a low-level abstraction layer. When a process makes

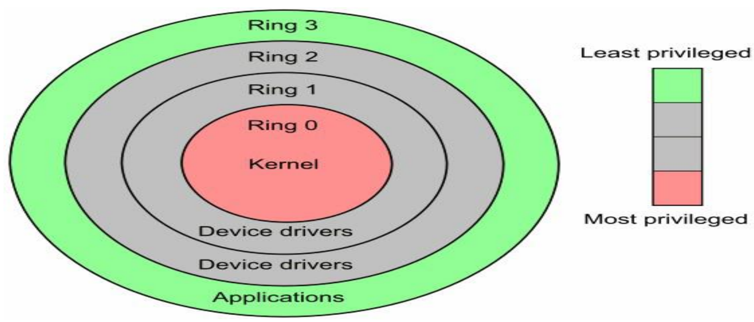
Requests of the kernel, it is called a system call. Kernel designs differ in how theymanage these system calls and resources.A monolithic kernel runs all the operating system instructions in the same addressspace, for speed.A microkernel runs most processes in user space, for modularity.

Functions of Kernel: The kernel's primaryfunction is to mediate accessto the computer's resources,including:

The **central processing unit**this central component of a computer system is responsible for running or executing programs. The kernel takes responsibility for deciding at any time which of the many running programs should be allocated to the processor or processors (each of which can usually run only one program at a time).

Random-access memory (RAM): It is used to store both program instructions and data.Typically, both need to be present in memory in order for a program to execute.Often multiple programs will want access to memory, frequently demanding morememory than the computer has available. The kernel is responsible for decidingwhich memory each process can use, and determining what to do when not enoughmemory is available.

Input/output (I/O) devices: - Itincludes such peripherals as keyboards, mice, disk drives, printers,USB devices, network adapters, and display devices. The kernel allocates requestsfrom applications to perform I/O to an appropriate device and provides convenientmethods for using the device (typically abstracted to the point where theapplication does not need to know implementation details of the device).Kernels also usually provide methods for synchronization and communicationbetween processes called inter-process communication (IPC).



Reentrant Kernel

In kernel mode, a reentrant kernel allows processes (or, more precisely, their corresponding kernel threads) to give up the CPU. They have no effect on other processes entering kernel mode. Multiple processor systems may be scheduled together in the case of single-processor systems.

Example:

A disc read is an example of this. When a user program requests a disc read, the scheduler will delegate the CPU to another process (kernel thread) until the disc controller issues an interrupt indicating that the data is accessible and our thread can be resumed. This process can still access I/O, such as user input, which requires kernel functions. The system remains responsive, and the amount of CPU time wasted as a result of IO delays is reduced. The original function (whatever requested data) would be blocked in a non-reentrant kernel until the disc read was completed.

If a computer program or routine can be safely called again before its previous invocation has been completed, it is said to be reentrant (i.e. it can be safely executed concurrently). A computer program or routine that is reentrant:

- There cannot be any non-constant data that is static (or global).
- The address must not be returned to static (or global) non-constant data.
- It must only work with the data provided by the caller.

- Locks should not be used to protect singleton resources, a variable that is only referenced once
- It must not change its own code (unless executing in its own unique thread storage).
- Non-reentrant computer programs or routines must not be called.

Non-reentrant functions can still be executed by reentrant kernels if locks are used to ensure that only one process can run the non-reentrant function. Even though the current process is operating in kernel mode, hardware interrupts can suspend it (allowing things like Ctrl+c to cease execution).

Difference between Microkernel and Monolithic Kernel

The microkernel and monolithic kernels are two types of kernels in the operating system. The kernel is the main part of the OS. As a result, the kernel's important code is stored in different memory spaces. The kernel is a crucial component because it maintains the proper functioning of the complete system. It manages hardware and processes, files handling, and several other functions.

In this article, you will learn about the microkernel and monolithic kernel. But before discussing the differences, you must know about the microkernel and monolithic kernel.

What is Microkernel?

The microkernel is a type of kernel that permits the customization of the OS. It is privileged and provides low-level address space management as well as **Inter-Process Communication (IPC)**. Furthermore, OS functions like the virtual memory manager, file system, and CPU scheduler are built on top of the microkernel. Every service has its address space to make them secure. Moreover, every application has its address space. As a result, there is protection between applications, OS Services, and the kernel.

When an application requests a service from the OS services, the OS services communicate with one another in order to provide the requested service to the application. Inter-Process Communication (IPC) can assist in establishing this communication. Overall, microkernel-based operating systems offer a high level of extensibility. It is also possible to customize the operating system's services to meet the needs of the application.

Advantages

1. These are modular, and several modules may be modified, reloaded, replaced without modifying the kernel.

2. The architecture of the microkernel is small and isolated, but it may work better.

3. The Microkernel system is a versatile technique in which the APIs implemented by several servers may coexist.

4. The system can be expanded more easily because it may be added to the system application without interrupting the kernel.

5. It adds new features without recompiling.

6. When compared to monolithic systems, there are fewer system crashes.

Disadvantages

1. A context switch is required in the microkernel when the drivers are run as processes.

2. The microkernel system performance might be variable and cause issues.

3. Microkernel services are more expensive than in a traditional monolithic system.

What is Monolithic Kernel?

The monolithic kernel manages the system's resources between the system application and the system hardware. Unlike the microkernel, user and kernel services are run in the same address space. It increases the kernel size and also increases the size of the OS.

The monolithic kernel offers CPU scheduling, device management, file management, memory management, process management, and other OS services via the system calls. All of these components, including file management and memory management, are located within the kernel. The user and kernel services use the same address space, resulting in a fast-executing operating system. One drawback of this kernel is that if anyone process or service of the system fails, the complete system crashes. The entire operating system must be modified to add a new service to a monolithic kernel.

Advantages

1. The monolithic kernel runs quickly because of memory management, file management, process scheduling, etc.

2. All of the components may interact directly with each other's and also with the kernel.

3. It is a single huge process that executes completely within a single address space.

4. Its structures are easy and simple. The kernel contains all of the components required for processing.

Disadvantages

1. If the user needs to add a new service, the user requires modifying the complete operating system.

2. It isn't easy to port code written in the monolithic operating system.

3. If any of the services fails, the entire system fails.