Operating System

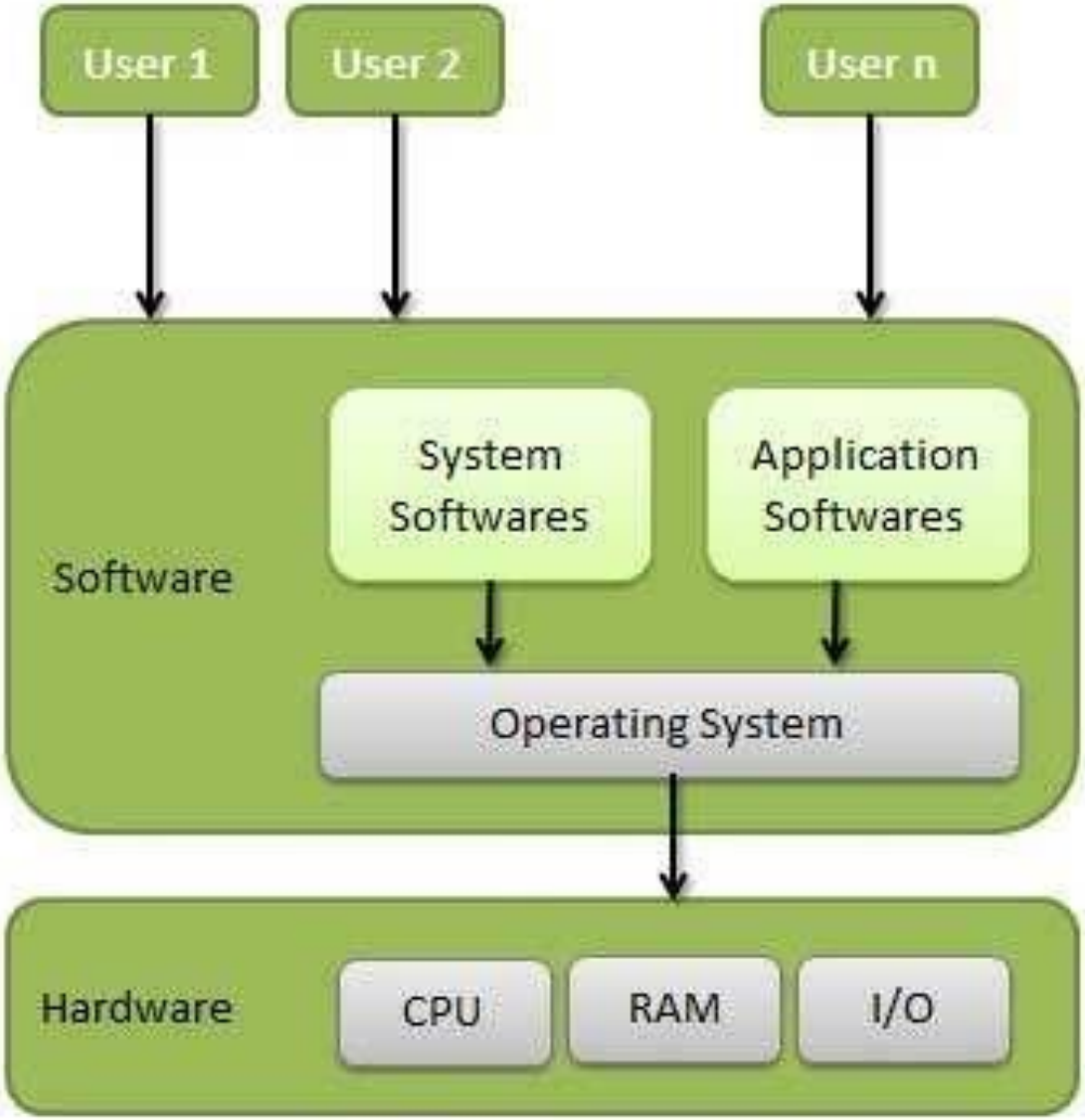
By Bhavika Parmar

Unit : 1
Operating System Concepts

What is OS?

- An **Operating System** (OS) is an interface between a computer user and computer hardware. An operating system is a software which performs all the basic tasks like file management, memory management, process management, handling input and output, and controlling peripheral devices such as disk drives and printers.
- An operating system is software that enables applications to interact with a computer's hardware. The software that contains the core components of the operating system is called the **kernel**.

- The primary purposes of an **Operating System** are to enable applications (spftwares) to interact with a computer's hardware and to manage a system's hardware and software resources.
- Some popular Operating Systems include Linux Operating System, Windows Operating System, VMS, OS/400, AIX, z/OS, etc. Today, Operating systems is found almost in every device like mobile phones, personal computers, mainframe computers, automobiles, TV, Toys etc.


User 1
User 2
User n
Software
System Softwares
Application Softwares
Operating System
Hardware
CPU
RAM
I/O

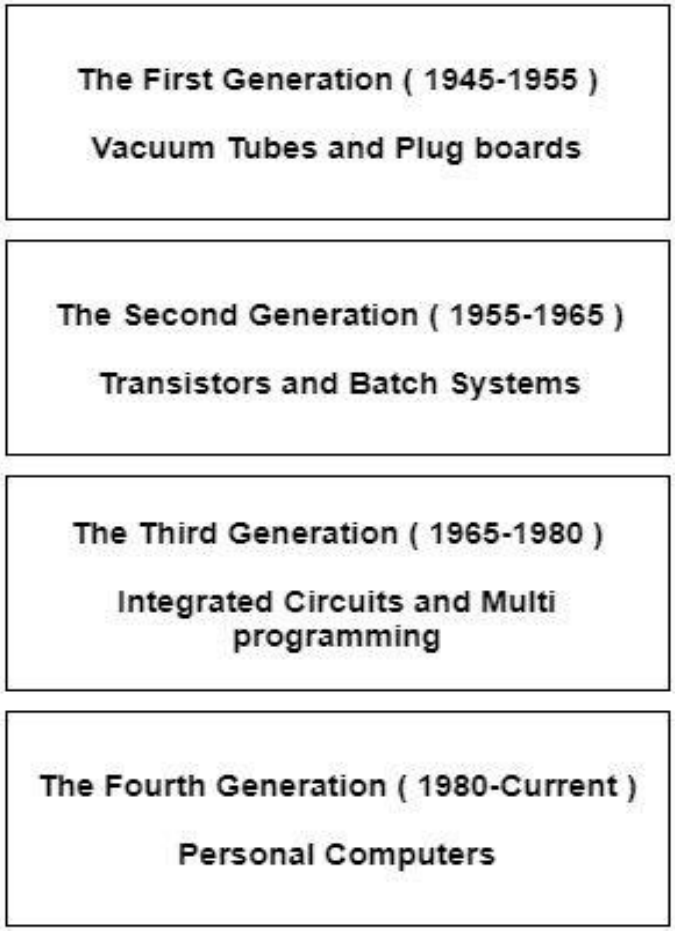
What is the main function of OS?

- OS serves as an interface between computer hardware and the user. The main function of an operating system (OS) is to control files and directories. Managing files on a computer is the operating system's responsibility. Creating, opening, closing, and deleting files all fall under this.

What is Evolution?

- Operating systems have progressed from slow and expensive systems to today's technology, which has exponentially increased computing power at comparatively modest costs. So let's have a detailed look at the evolution of operating systems.

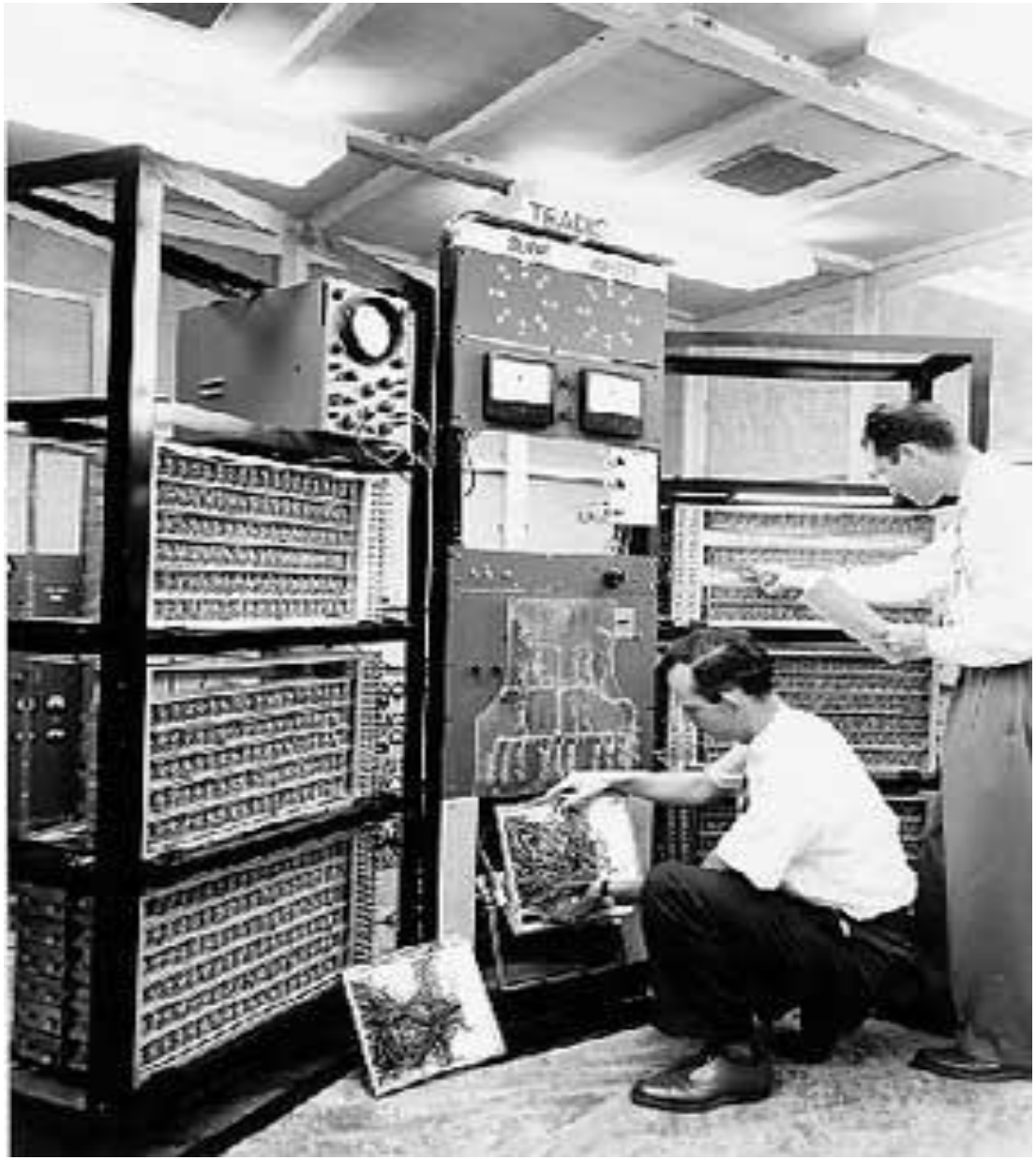
The operating system can be classified into four generations, as follows:



OPERATING SYSTEM GENERATIONS

The First Generation (1945 - 1955): Vacuum Tubes and Plugboards

- Digital computers were not constructed until the second world war. However, the mechanical relays were very slow and were later replaced with vacuum tubes. These machines were enormous but were still very slow.
- These early computers were designed, built and maintained by a single group of people. Programming languages were unknown and there were no operating systems so all the programming was done in machine language. All the problems were simple numerical calculations.
- By the 1950's punch cards were introduced and this improved the computer system. Instead of using plugboards, programs were written on cards and read into the system.

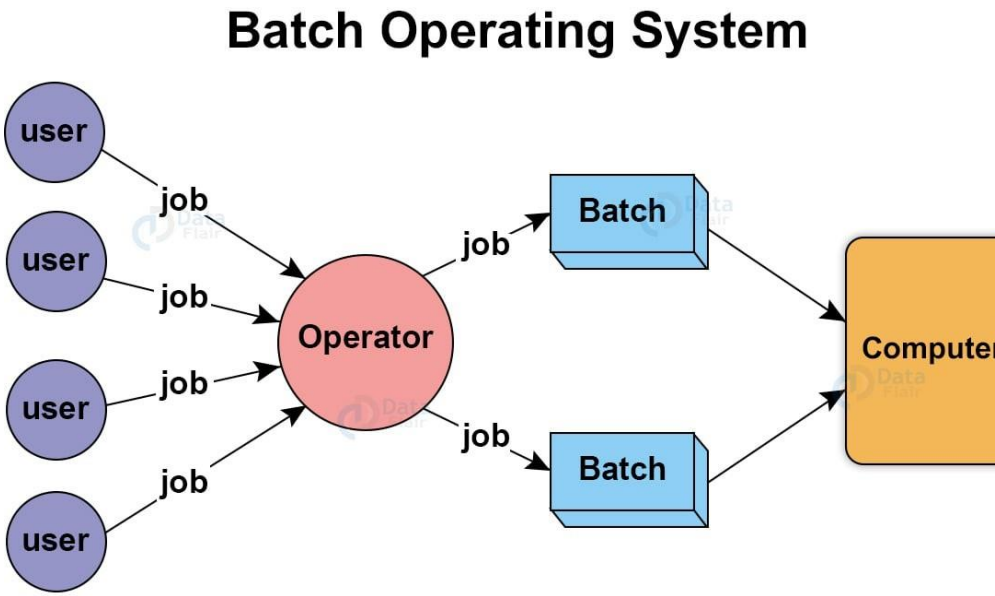

TRADIC

The Second Generation (1955 - 1965): Transistors and Batch Systems

- The second generation of computer hardware was most notably characterised by transistors replacing vacuum tubes as the hardware component technology. The first operating system GMOS was developed by the IBM computer. GMOS was based on single stream batch processing system, because it collects all similar jobs in groups or batches and then submits the jobs to the operating system using a punch card to complete all jobs in a machine. Operating system is cleaned after completing one job and then continues to read and initiates the next job in punch card.

- Researchers began to experiment with multiprogramming and multiprocessing in their computing services called the time-sharing system. A noteworthy example is the Compatible Time Sharing System (CTSS), developed at MIT during the early 1960s.

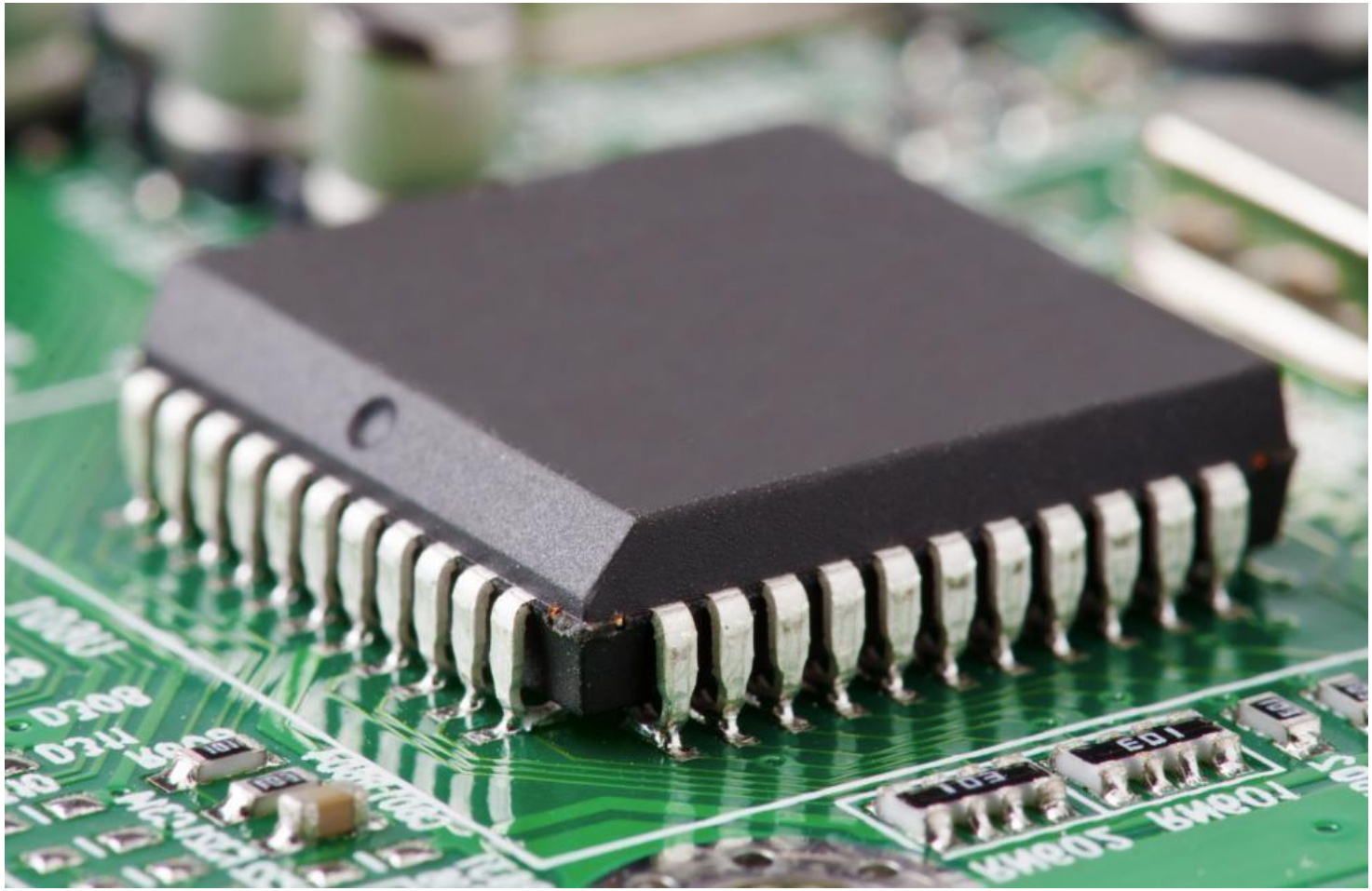
- The Batch System was introduced to reduce the wasted time in the computer. A tray full of jobs was collected in the input room and read into the magnetic tape. After that, the tape was rewound and mounted on a tape drive. Then the batch operating system was loaded in which read the first job from the tape and ran it. The output was written on the second tape.



- Batch operating system is a type of operating system which does not communicate or interact with the computer directly. It is helpful when computer processes are very lengthy and time-consuming. It speeds up the process by combining together similar types of jobs and runs them as a group.
- There is an operator in this kind of operating system which is used to club similar jobs having the same requirement and is responsible to sort jobs with similar needs. Eg: BankSystem, Payroll System, etc

The Third Generation (1965 - 1980): Integrated Circuits and Multiprogramming

- There were two types of computer systems i.e the scientific and the commercial computers. These were combined by IBM in the System/360. This used integrated circuits and provided a major price and performance advantage over the second generation systems.
- Another progress which leads to developing of personal computers in fourth generation is a new development of minicomputers



Mutiprogramming

- This type of operating system enables people who are in two different shells to use the computer system at the same time.The processing time is termed time sharing as it is shared among multiple users.
- Each task is allotted some time (known as Quantum) to execute its work. After the allotted time interval is over the operating system switches over to the next task
- **Eg: UNIX, Multics**

The Fourth Generation (1980 - Present): Personal Computers

- The fourth generation is characterised by the appearance of the personal computer and the workstation. The component technology of the third generation, was replaced by very large scale integration (VLSI). Many Operating Systems which we are using today like Windows, Linux, MacOS etc developed in the fourth generation.
- These were chips containing thousands of transistors on a square centimeter of silicon. Because of these, microcomputers were much cheaper than minicomputers and that made it possible for a single individual to own one of them.

- The advent of personal computers also led to the growth of networks. This created network operating systems and distributed operating systems. The users were aware of a network while using a network operating system and could log in to remote machines and copy files from one machine to another.

Characteristics of Operating System

- **Memory Management:** The operating system manages memory. It has complete knowledge of primary memory; which part of the memory is used by which program. Whenever a program requests, it allocates memory.
- **Processor Management:** It allocates the program to the processor (CPU) and also deallocates it when a program runs out of the CPU needs.
- **Device Management:** The operating system keeps the information about all devices. It is also called the I/O controller, and the operating system also decides which devices are used to which program, when, and for how long.

- **Security:** It prevents unauthorized access to any program. It uses passwords and other technologies.
- **Reliability:** It is very reliable because no any virus and harmful code can be detected in it.
- **File Management:** It allocates and deallocates resources and decides which program to allocate resources.
- **Easy to use:** It can be easily used as it also has a GUI interface.

Needs of Operating System

- More than one program runs at a time in a computer, and all of them require your computer's CPU and memory. The operating system manages resources for all those programs. That is why the operating system is required.
- Multitasking is a very critical feature of the OS. With its help, we can run many programs simultaneously.
- The operating system provides a platform to run any application program in the computer. Due to which we can do our work with the help of that application.
- It helps the user in file management. Through this, the user can save the data according to his needs.

- You use your mouse to open the application and click on the menu. All this is possible due to the modern operating system. This operating system allows you to do this with the help of GUI (Graphical user interface).

- The operating system creates a communication link between the user and the computer, allowing the user to run any application program and obtain the required output properly.

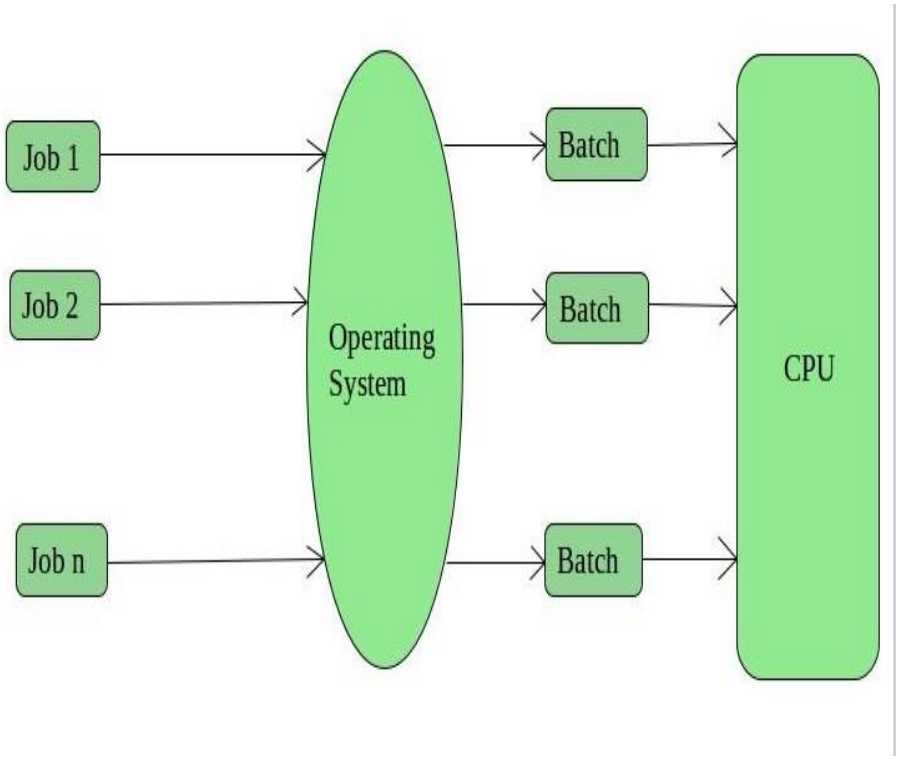
- It is almost impossible for a user to use a computer system without an operating system. Many processes run simultaneously when a program is executed, which is not easy for a person to manage.

Types of OS

- Batch operating system
- Time sharing operating system
- Distributed operating system
- Network operating system

1. Batch Operating System

- This type of operating system does not interact with the computer directly. There is an operator which takes similar jobs having the same requirement and group them into batches. It is the responsibility of the operator to sort jobs with similar needs.



Advantages of Batch Operating System:

- It is very difficult to guess or know the time required for any job to complete. Processors of the batch systems know how long the job would be when it is in queue
- Multiple users can share the batch systems
- The idle time for the batch system is very less
- It is easy to manage large work repeatedly in batch systems

Disadvantages of Batch Operating System:

- The computer operators should be well known with batch systems
- Batch systems are hard to debug
- It is sometimes costly
- The other jobs will have to wait for an unknown time if any job fails

2. Time-Sharing Operating Systems –

- Each task is given some time to execute so that all the tasks work smoothly. Each user gets the time of CPU as they use a single system. These systems are also known as Multitasking Systems. The task can be from a single user or different users also. The time that each task gets to execute is called quantum. After this time interval is over OS switches over to the next task.

Advantages of Time-Sharing OS:

- Each task gets an equal opportunity
- Fewer chances of duplication of software
- CPU idle time can be reduced

Disadvantages of Time-Sharing OS:

- One must have to take care of the security and integrity of user programs and data
- Data communication problem

3. Network Operating System –

- these systems run on a server and provide the capability to manage data, users, groups, security, applications, and other networking functions. These types of operating systems allow shared access of files, printers, security, applications, and other networking functions over a small private network.
- these computers are popularly known as **tightly coupled systems**.

Advantages of Network Operating System:

- Highly stable centralized servers
- Security concerns are handled through servers
- New technologies and hardware up-gradation are easily integrated into the system
- Server access is possible remotely from different locations and types of systems

Disadvantages of Network Operating System:

- Servers are costly
- User has to depend on a central location for most operations
- Maintenance and updates are required regularly

Examples of Network Operating System are:

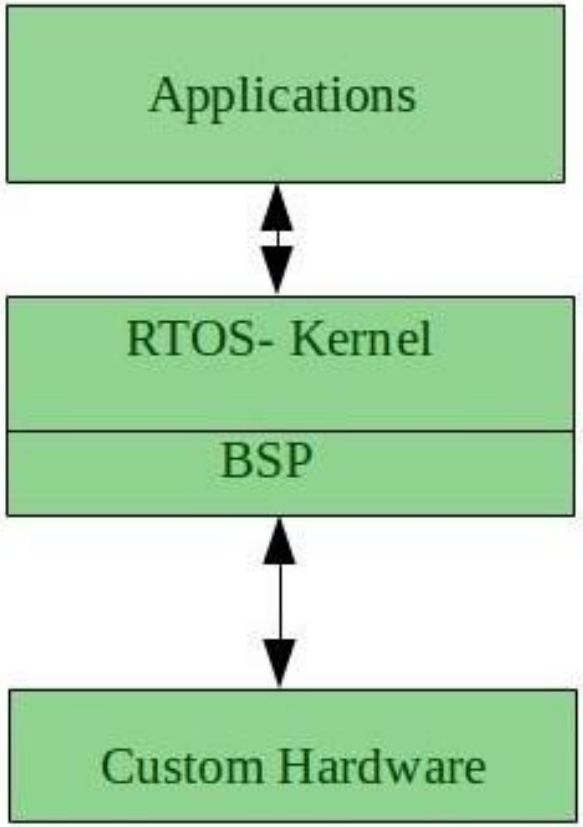
Microsoft Windows Server 2003, Microsoft Windows Server 2008

4. Real-Time Operating System –

- These types of OSs serve real-time systems. The time interval required to process and respond to inputs is very small. This time interval is called **response time**.
- **Real-time systems** are used when there are time requirements that are very strict like missile systems, air traffic control systems, robots, etc.

Two types of Real-Time Operating System which are as follows:

- **Hard Real-Time Systems:**
  These OSs are meant for applications where time constraints are very strict and even the shortest possible delay is not acceptable. These systems are built for saving life like automatic parachutes or airbags which are required to be readily available in case of any accident. Virtual memory is rarely found in these systems.
- **Soft Real-Time Systems:**
  These OSs are for applications where for time-constraint is less strict.


Applications
RTOS- Kernel
BSP
Custom Hardware

BSP:

- A board support package (BSP) is **essential code for a given computer hardware device that will make that device work with the computer's OS (operating system)**. The BSP contains a small program called a boot loader or boot manager that places the OS and device drivers into memory.

Advantages of RTOS:

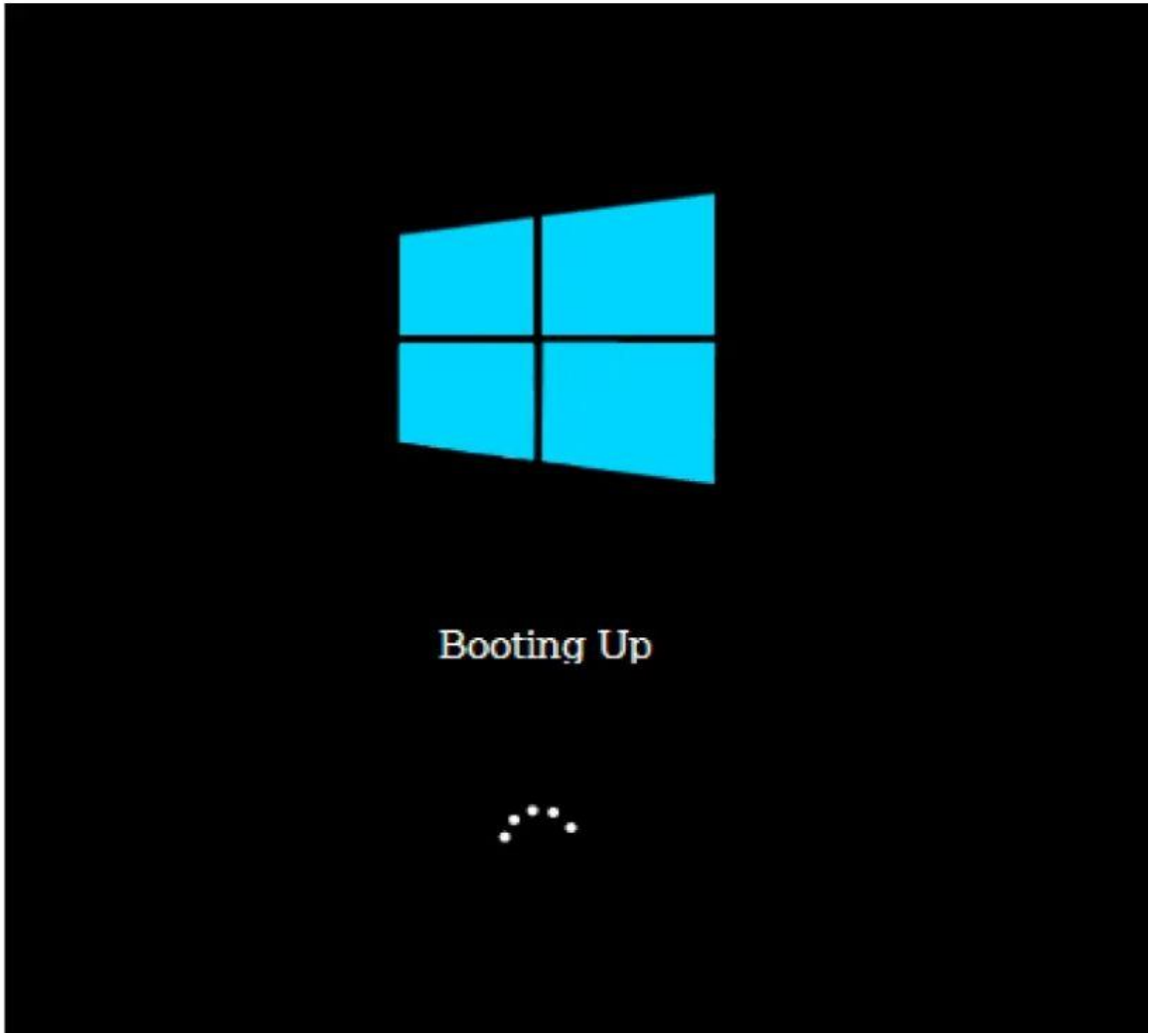
- **Error Free:** These types of systems are error-free.
- **Memory Allocation:** Memory allocation is best managed in these types of systems.
- **Maximum Consumption:** Maximum utilization of devices and system, thus more output from all the resources

Disadvantages of RTOS:

- **Limited Tasks:** Very few tasks run at the same time and their concentration is very less on few applications to avoid errors.
- **Use heavy system resources:** Sometimes the system resources are not so good and they are expensive as well.
- **Complex Algorithms:** The algorithms are very complex and difficult for the designer to write on.
- **Examples of Real-Time Operating Systems are:** Scientific experiments, medical imaging systems, industrial control systems, weapon systems, robots, air traffic control systems, etc.

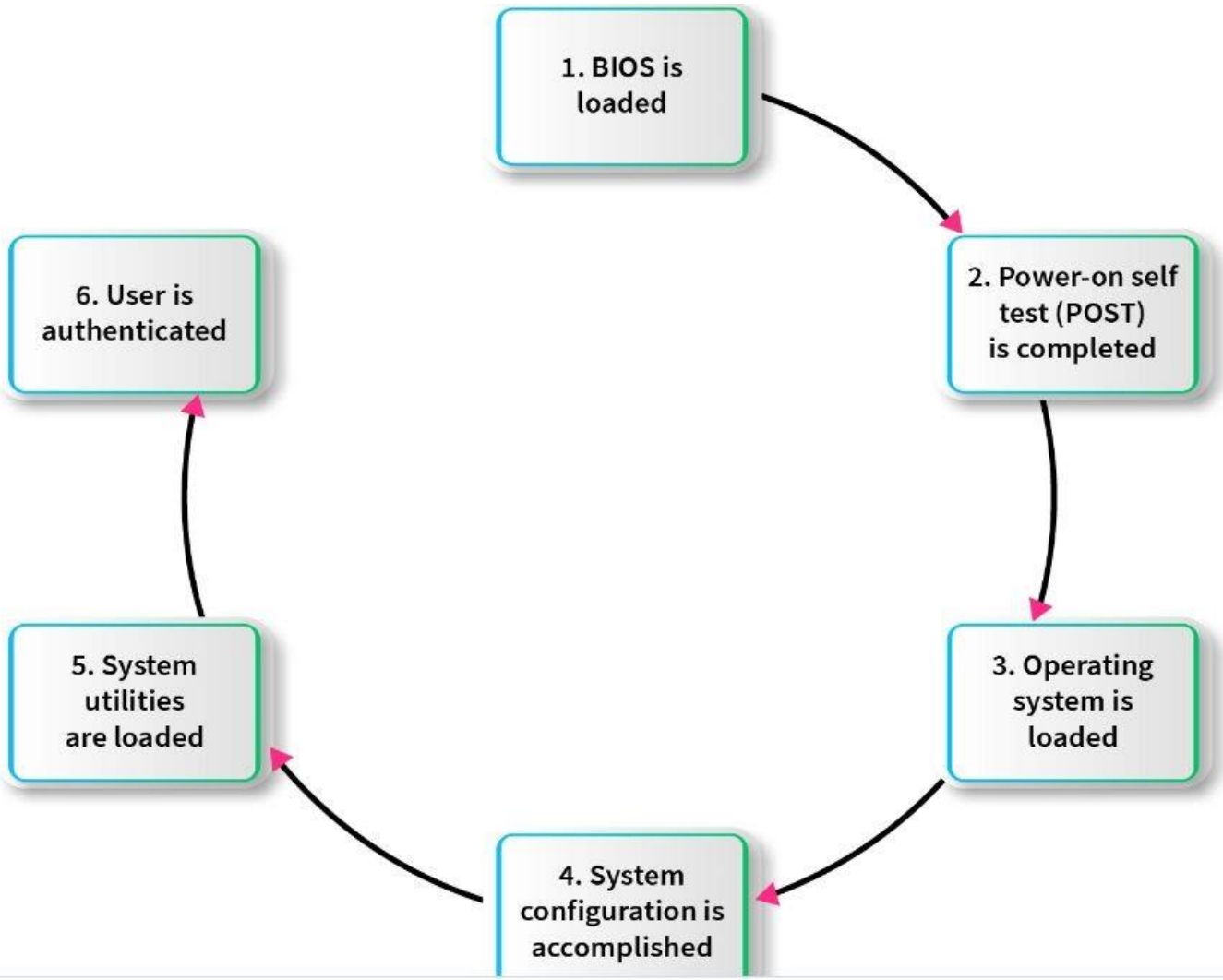
Booting

- Booting is basically the process of starting the computer. When the CPU is first switched on it has nothing inside the Memory. In order to start the Computer, load the Operating System into the Main Memory and then Computer is ready to take commands from the User. Learn the types of booting.
- Common keys for accessing the Boot Menu are Esc, F2, F10 or F12, depending on the manufacturer of the computer or motherboard.


Booting Up

What happens in the Process of Booting?

- Booting happens when you start the computer. This happens when we turned ON the power or the computer restarts. The system BIOS (Basic Input/Output System) makes the peripheral devices active. Further, it requires that the boot device loads the operating system into the main memory.


1. BIOS is loaded
2. Power-on self test (POST) is completed
3. Operating system is loaded
4. System configuration is accomplished
5. System utilities are loaded
6. User is authenticated

Boot Devices

- Booting can be done either through hardware (pressing the start button) or by giving software commands. Therefore, a boot device is a device that loads the operating system. Moreover, it contains the instructions and files which start the computer. Examples are the hard drive, floppy disk drive, CD drive, etc. Among them, the hard drive is the most used one.

- The BIOS chip tells it to look in a fixed place, usually on the lowest-numbered hard disk (the boot disk) for a special program called a boot loader. The boot loader is pulled into memory and started. The boot loader's job is to start the real operating system.
- POST (Power On Self-Test) The Power On Self-Test happens each time you turn your computer on. It initializes the various hardware devices. It is an important process so as to ensure that all the devices operate smoothly without any conflicts.

- The POST first checks the bios and then tests the RAM. If there is no problem with this then POST continues to check the CPU, hardware devices such as the Video Card, the secondary storage devices such as the Hard Drive, Floppy Drives, Zip Drive or CD/DVD Drives. If some errors found, then an error message is displayed on the screen or a number of beeps are heard. These beeps are known as POST beep codes

Types of Booting

- There are two types of booting:

Cold Booting

- A cold boot is also called **a hard boot.** It is the process when we first start the computer. In other words, when the computer is started from its initial state by pressing the power button it is called cold boot. The instructions are read from the ROM and the operating system is loaded in the main memory.

Warm Booting

- Warm Boot is also called **soft boot**. It refers to when we restart the computer. Here, the computer does not start from the initial state. When the system gets stuck sometimes it is required to restart it while it is ON. Therefore, in this condition the warm boot takes place. Restart button or CTRL+ALT+DELETE keys are used for warm boot.

Steps of Booting

1. The Startup

- It is the first step that involves switching the power ON. It supplies electricity to the main components like BIOS and processor.

2. BIOS: Power On Self Test

- It is an initial test performed by the BIOS. Further, this test performs an initial check on the input/output devices, computer’s main memory, disk drives, etc. Moreover, if any error occurs, the system produces a beep sound.

3. Loading of OS

- In this step, the operating system is loaded into the main memory. The operating system starts working and executes all the initial files and instructions.

4. System Configuration

- In this step, the drivers are loaded into the main memory. Drivers are programs that help in the functioning of the peripheral devices.

5. Loading System Utilities

- System utilities are basic functioning programs, for example, volume control, antivirus, etc. In this step, system utilities are loaded into the memory.

6. User Authentication

- If any password has been set up in the computer system, the system checks for user authentication. Once the user enters the login Id and password correctly the system finally starts.

Failure during boot:

- If the computer cannot boot, we will get a boot failure error. This error indicates that the computer is not passing POST or a device in the computer, such as the hard drive or memory, has failed. You may also hear a beep code to identify that hardware is failing during the POST. An error message or blue screen may show on the screen as operating system files cannot be loaded, due to not being found or being corrupt.

Functions of Operating Systems

Security

- The operating system uses a password protection to protect user data it also prevents unauthorized access to programs and user data, but for external functionality we need to install malware software to protect the system.

Control over system performance

- The operating system monitors overall system setup to help in improving the performance and it also records the response time between service requests and system response so that it has a complete view of the system.

Job Accounting

- Operating systems always keep track of time and resources that are used by various tasks and users, this information can be used to track resource usage for a particular user or a group of users.

Error detecting aids

- Operating systems constantly monitor the system which helps us to detect errors and also avoid the malfunctioning of computer systems.

Coordination between other software and users

- Operating systems help in coordinate and assign interpreters, compilers, assemblers, and other software to the various users of the computer systems.

Memory Management

- It keeps track of primary memory.
- Memory addresses that have already been allocated and the memory addresses of the memory that has not yet been used.
- It allocates the memory to a process when the process requests it and deallocates the memory when the process has terminated.
- Multiprogramming.

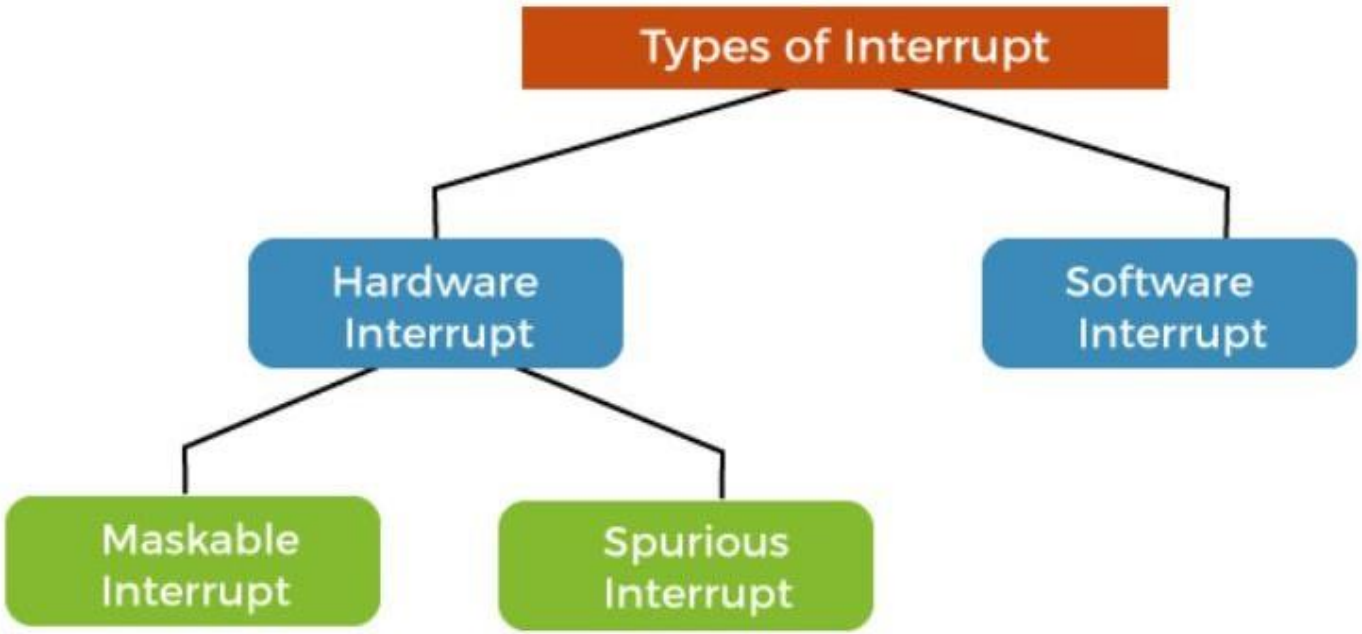
- Processor Management
- Keeps track of the status of processes.
- The program to track the status is known as traffic controller.
- It allocates the CPU and deallocates the processor when it is not required.
- Device Management
- Keeping track of all devices connected to the system.
- The OS designates a program that is responsible for every device which is called the Input/output controller.
- File management
- It keeps track of where information is stored, user access settings and status of every file and more.

INTERRUPT

- An interrupt is a signal emitted by hardware or software when a process or an event needs immediate attention. It alerts the processor to a high-priority process requiring interruption of the current working process. In I/O devices, one of the bus control lines is dedicated for this purpose and is called the ***Interrupt Service Routine*** (ISR).
- Bus control lines – carries control signals from the processor to other components.

- When a device raises an interrupt at the process, the processor first completes the execution of an instruction. Then it loads the ***Program Counter*** (PC) with the address of the first instruction of the ISR. Before loading the program counter with the address, the address of the interrupted instruction is moved to a temporary location. Therefore, after handling the interrupt, the processor can continue with the process.
- An operating system usually has some code that is called an ***interrupt handler***. The interrupt handler prioritizes the interrupts and saves them in a queue if more than one is waiting to be handled. The operating system has another little program called a ***scheduler*** that figures out which program to control next.

Types of Interrupt



- 1. Hardware Interrupts
- A hardware interrupt is a condition related to the state of the hardware that may be signaled by an external hardware device, e.g., an interrupt request (IRQ) line on a PC, or detected by devices embedded in processor logic to communicate that the device needs attention from the operating system. For example, pressing a keyboard key or moving a mouse triggers hardware interrupts that cause the processor to read the keystroke or mouse position.

- A IRQ value is an assigned location where the computer can expect a particular device to interrupt it when the device sends the computer signals about its operation.
- For example, when a printer has finished printing it sends as interrupt signals to computer.

Maskable Interrupts:

- Processors typically have an internal interrupt mask register which allows selective enabling and disabling of hardware interrupts. Each interrupt signal is associated with a bit in the mask register; on some systems, the interrupt is enabled when the bit is set and disabled when the bit is clear, while on others, a set bit disables the interrupt. When the interrupt is disabled, the associated interrupt signal will be ignored by the processor. Signals which are affected by the mask are called ***maskable interrupts***.

Spurious Interrupts:

- Also known as a *phantom interrupt* or *ghost interrupt*, a spurious interrupt is a type of hardware interrupt for which no source can be found. These interrupts are difficult to identify if a system misbehaves.

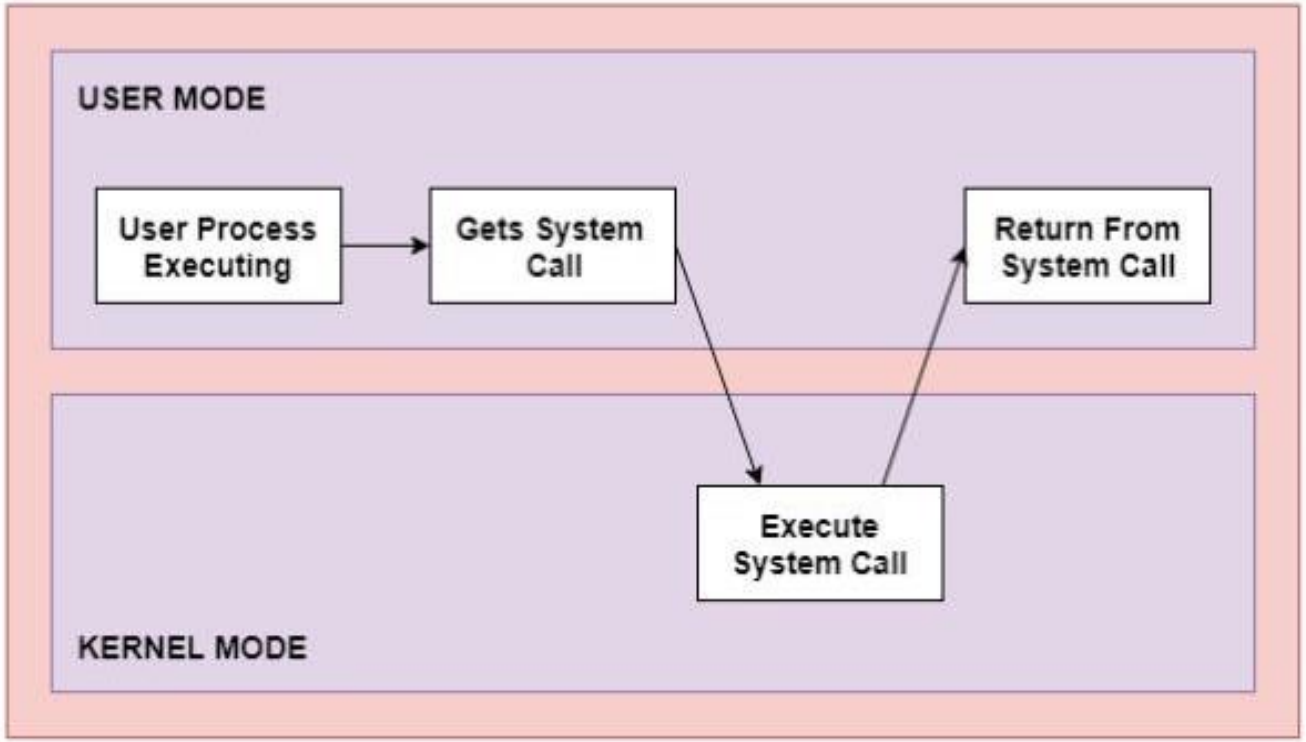
Software Interrupts

- The processor requests a software interrupt upon executing particular instructions or when certain conditions are met. Every software interrupt signal is associated with a particular interrupt handler.
- A software interrupt may be intentionally caused by executing a special instruction that invokes an interrupt when executed by design. Such instructions function similarly to subroutine calls and are used for various purposes, such as requesting operating system services and interacting with device drivers.

SYSTEM CALLS

- A system call is a way for a user program to interface with the operating system. The program requests several services, and the OS responds by invoking a series of system calls to satisfy the request. A system call can be written in assembly language or a high-level language like **C** or **Pascal**. System calls are predefined functions that the operating system may directly invoke if a high-level language is used.

- There are two modes in a computer: the user mode and the kernel mode. Most programs in the operating system are executed in the user mode. When a program is executed in the user mode and requires access to any kind of service or a hardware resource, it asks the kernel for resource via a system call. When the program makes a system call, the mode changes from user mode to kernel mode. After completing the task, the mode changes from the kernel mode back to the user mode. We also call this mode change context switching


USER MODE
User Process Executing
Gets System Call
Return From System Call
Execute System Call
KERNEL MODE

In general, system calls are required in the following situations −

- If a file system requires the creation or deletion of files. Reading and writing from files also require a system call.
- Creation and management of new processes.
- Network connections also require system calls. This includes sending and receiving packets.
- Access to a hardware devices such as a printer, scanner etc. requires a system call.

- A buffer is **a storage device used to compensate for a difference in rate of data flow, or time of occurrence of events, when transmitting data from one device to another**.

Types of System Calls

- Process Control

These system calls deal with processes such as process creation, process termination etc.

- File Management

These system calls are responsible for file manipulation such as creating a file, reading a file, writing into a file etc.

Device Management

- These system calls are responsible for device manipulation such as reading from device buffers, writing into device buffers etc.

Information Maintenance

- These system calls handle information and its transfer between the operating system and the user program.

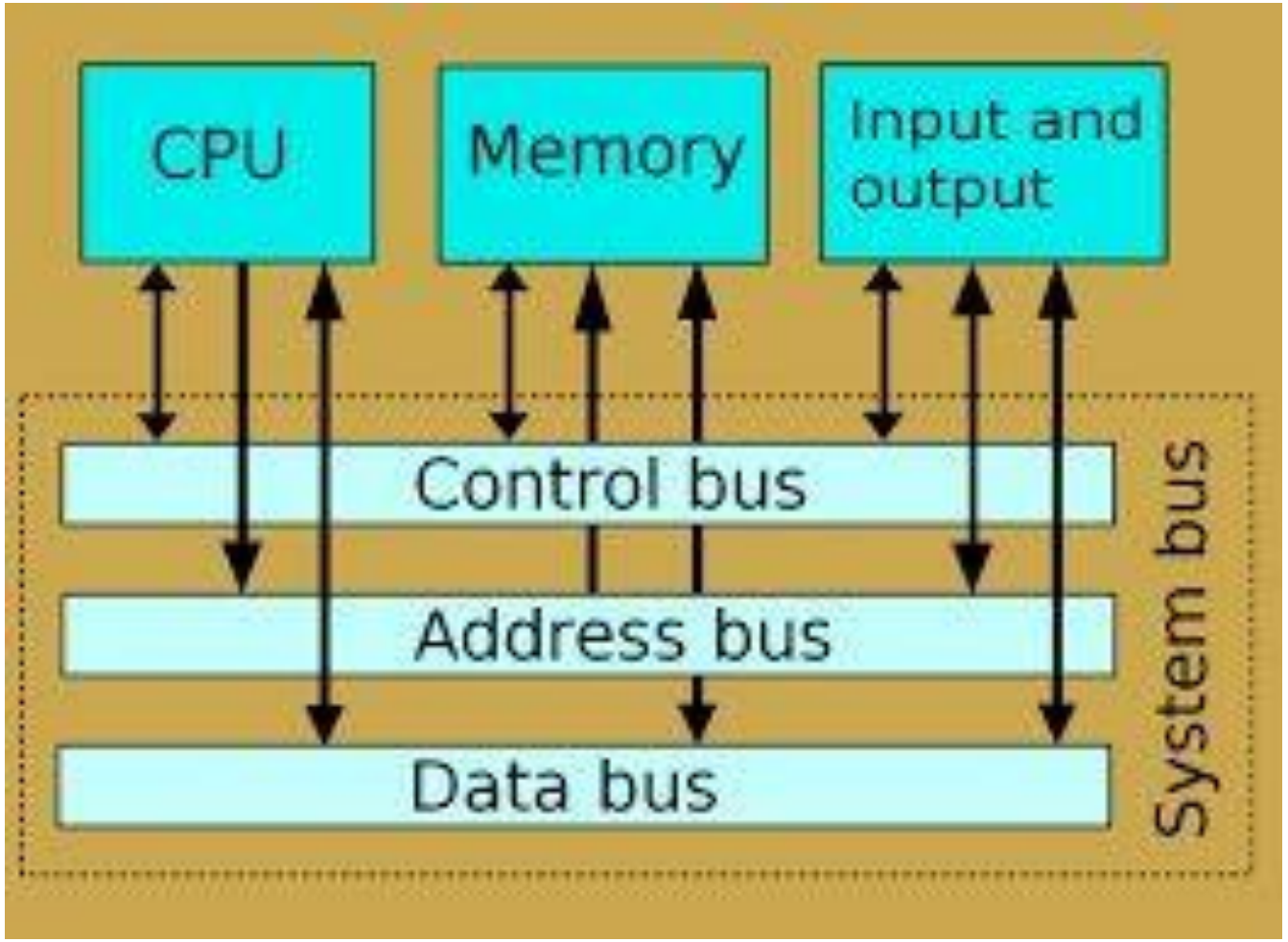
Communication

- These system calls are useful for interprocess communication. They also deal with creating and deleting a communication connection.

Data bus and Address bus

- There are three major components in a computer system. They are the Central Processing Unit (CPU), Memory and Input, Output devices. The CPU is the major unit that process data. Memory holds data required for processing. The input and output devices allow the users to communicate with the computer. The mechanism for each of the components to communicate with each other is the bus architecture.

The system bus is divided into three types called address bus, data bus and control bus



What is Address Bus

- The address bus helps to identify the particular location in the memory. Assume that the CPU needs to read data from memory. Then, the address bus helps to identify that specific location. Furthermore, each IO device has a unique ID, and it is the address of that component. The address bus helps to transfer memory addresses of data and IO.

- The width of the address bus determines the amount of memory the system can address. When there are ‘n’ address lines, it can directly address 2^n memory locations. For example, an 8085 microprocessor has address bus of 16 bits. Therefore, it can access $2^{16} = 65536$ different memory locations.

What is Data Bus

- Data bus helps to transfer data between various components. It includes the related hardware components such as wires and optical fiber. The data bus consists of 8, 32,64, etc. separate lines. The number of lines refers to the width of the data bus. This bus width helps to determine the data transferring rate. Therefore, the data bus width determines the system performance, but it is expensive to increase the number of lines.
- The other bus besides the address bus and data bus is the control bus. It is bidirectional and transmits control signals from one component to another.

What is Control Bus

- The control bus is bidirectional; it transmits command signals from the CPU and response signals from the hardware. It helps the CPU synchronize its command signals to the computer's components and slower external devices. As a result, the control bus consists of control lines that each send a specific signal, like read, write, and interrupt. The control lines that make up a control bus differ between processors, but most include system clock lines, status lines, and byte enable lines.

- For example, a computer’s CPU will use the data bus to transmit information to and from the central memory. The control bus allows the CPU to determine whether and when the system is sending or receiving this data. This is because a control bus has a control line for read and one for write that determine the direction the information flows (memory to CPU or CPU to memory). If the CPU needs to write some data to the central memory, it will send a signal on (assert) the control bus’s write control line. Sending a signal on the read control line allows the CPU to receive data from memory.

THANK YOU