COMPUTER ORGANIZATION-PART 2

- **I/O devices (keyboard, mouse, CRT/LCD/LED, Printers, scanners). I/O Interfacing – memory mapped & I/O Mapped I/O, Polling, interrupt driven I/O, DMA – controller. Serial communication – UART, RS 232, USB.**
- **High performance computing – pipelining, basic concepts in parallel processing, Grid and cluster computing.**

I/O DEVICES

- An input/output device, often known as an IO device, is any hardware that allows a human operator or other systems to interface with a computer.
- Input/output devices, as the name implies, are capable of delivering data (output) to and receiving data from a computer (input).
- An input/output (I/O) device is a piece of hardware that can take, output, or process data.
- It receives data as input and provides it to a computer, as well as sends computer data to storage media as a storage output.

INPUT DEVICES

- Following are some of the important input devices which are used in a computer −

    1. **Keyboard**
    2. **Mouse**
    3. **Joy Stick**
    4. **Light pen**
    5. **Track Ball**
    6. **Scanner**
    7. **Graphic Tablet**
    8. **Microphone**
    9. **Magnetic Ink Card Reader(MICR)**
    10. **Optical Character Reader(OCR)**
    11. **Bar Code Reader**
    12. **Optical Mark Reader(OMR)**

KEYBOARD
- The keyboard is the most frequent and widely used input device for entering data into a computer.
- Although there are some additional keys for performing other operations, the keyboard layout is similar to that of a typical typewriter.
- Generally, keyboards come in two sizes: 84 keys or 101/102 keys, but currently keyboards with 104 keys or 108 keys are also available for Windows and the Internet.

Types of Keys

1. **Numeric Keys**: It is used to enter numeric data or move the cursor. It usually consists of a set of 17 keys.
2. **Typing Keys**: The letter keys (A-Z) and number keys (09) are among these keys.

3. **Control Keys**: These keys control the pointer and the screen. There are four directional arrow keys on it. Home, End, Insert, Alternate (Alt), Delete, Control (Ctrl), etc., and Escape are all control keys (Esc).

4. **Special Keys**: Enter, Shift, Caps Lock, NumLk, Tab, etc., and Print Screen are among the special function keys on the keyboard.

5. **Function Keys**: The 12 keys from F1 to F12 on the topmost row of the keyboard.

MOUSE

- The most common pointing device is the mouse.
- The mouse is used to move a little cursor across the screen while clicking and dragging.
- The cursor will stop if you let go of the mouse.
- The computer is dependent on you to move the mouse; it won't move by itself.
- As a result, it's an input device.
- A mouse is an input device that lets you move the mouse on a flat surface to control the coordinates and movement of the on-screen cursor/pointer.
- The left mouse button can be used to select or move items, while the right mouse button when clicked displays extra menus.
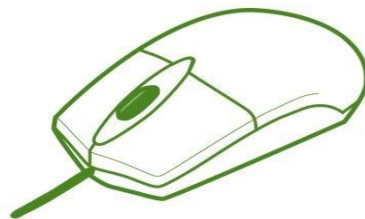
Advantages

- Easy to use

- Not very expensive
- Moves the cursor faster than the arrow keys of the keyboard.

SCANNER

- Scanner is an input device, which works more like a photocopy machine.
- It is used when some information is available on paper and it is to be transferred to the hard disk of the computer for further manipulation.
- It's employed when there's information on paper that needs to be transferred to the computer's hard disc for subsequent manipulation.
- Scanner collects images from the source and converts them to a digital format that may be saved on a disc. Before they are printed, these images can be modified.

Types of Scanners

- The one thing that is constant in the technical world is upgrading, scanners also get upgraded as the technology evolves.
- They are also available in various sizes, configurations, etc., and can be assembled as per the requirement of the end-user.
- While some are specifically built for scanning black and white documents only, others can also scan color documents.

1. Feed-in or Sheet-fed Scanners:

✓ These are simple scanners that have the limitation of scanning paper documents only as they have a feeder tray that takes in the paper kept in the tray.

✓ Thus, scanning anything in book form is not possible here.

✓ However, few advanced models do have the option of scanning several pages in sequence.

2. Flatbed Scanner:

✓ This is very commonly used and available in the markets readily.

✓ These have a flat surface for screening with a cover that must be lifted to place the material that has to be scanned.

✓ This model is viable for scanning books like magazines, educational content that is bound to be bulky.

3. Handheld scanners:

✓ These are essentially used to scan barcodes, thus useful for business transactions and general use.

✓ Other scanners like the overhead scanner, the big format scanner, slide or negative scanners are also available in the market.

OUTPUT DEVICES

- Following are some of the important output devices used in a computer.

1. **Monitors**
2. **Graphic Plotter**
3. **Printer**

MONITORS

- Monitors, commonly called as Visual Display Unit (VDU), are the main output device of a computer.
- It forms images from tiny dots, called pixels that are arranged in a rectangular form.
- The sharpness of the image depends upon the number of pixels.

Types of Monitors
There are several types of monitors; some are as follows:

1. Cathode Ray Tube (CRT) Monitors
- It is a technology used in early monitors.
- It uses a beam of electrons to create an image on the screen.
- It comprises the guns that fire a beam of electrons inside the screen.
- The electron beams repeatedly hit the surface of the screen.
- These guns are responsible for generating RGB (Red, Green, and Blue) colors, and more other colors can be generated with the help of combining these three colors.
- Today's Flat Panel Monitors replace the CRT monitors.

- There are some disadvantages of CRT −

✔ Large in Size
✔ High power consumption

2. Flat Panel Monitors
- These types of monitors are lightweight and take less space.
- They consume less power as compared to CRT monitors.
- These monitors are more effective as they do not provide harmful radiation.
- These monitors are more expensive than CRTs.
- The flat-panel monitors are used in PDA, notebook computers, and cellular phones.
- These monitors are available in various sizes like 15", 17", 18" & 19" and more.
- The display of a flat-panel monitor is made with the help of two plates of glass.
- These plates contain a substance, which is activated in many ways.

- The flat-panel display is divided into two categories −

1. **Emissive Displays** − Emissive displays are devices that convert electrical energy into light. For example, **plasma panel and LED (Light-Emitting Diodes)**.

2. **Non-Emissive Displays** − Non-emissive displays use optical effects to convert sunlight or light from some other source into graphics patterns. For example, **LCD (Liquid-Crystal Device).**

Liquid Crystal Display:
- LCD (Liquid crystal display) screen contains a substance known as liquid crystal.
- The particles of this substance are aligned in a way that the light located backside on the screens, which allow generating an image or blocking.
- Liquid crystal display offers a clear picture as compared to CRT display and emits less radiation.
- Furthermore, it consumes less power and takes less space than a CRT display.

Gas Plasma Display:
- This display uses gas plasma technology, which uses a layer of gas between 2 plates of glass.
- When voltage is applied, the gas releases ultraviolet light.
- By this ultraviolet light, the pixels on the screen glow and form an image.
- These displays are available in different sizes of up to 150 inches.
- Although it offers effective colors as compared to the LCD monitor, it is more expensive. That's why it is less used.

3. **LED MONITOR**

- It is a flat screen computer monitor, which stands for light-emitting diode display.
- It is lightweight in terms of weight and has a short depth.
- As the source of light, it uses a panel of LEDs.
- Nowadays, a wide number of electronic devices, both large and small devices such as laptop screens, mobile phones, TVs, computer monitors, tablets, and more, use LED displays.
- It is believed that James P. Mitchell invented the first LED display.
- On 18 March 1978, the first prototype of an LED display was published to the market at the SEF (Science and Engineering Fair) in Iowa.
- On 8 May 1978, it was shown again in Anaheim California, at the SEF.
- This prototype received awards from NASA and General Motors.

Advantages of LED Monitor:

- It includes a broader dimming range.
- It is a more reliable monitor.
- It is often less expensive.
- It consumes less power (20 watts), and run on a lower temperature.
- It has a more dynamic contrast ratio.

Difference between LCD and LED

The below table contains several differences between LCD and LED:

LCD	LED
It stands for **Liquid Crystal Display**.	Short for **Light-Emitting Diodes**.
LCD monitors are not a subset of LED monitors.	LED monitors are subset of LCD monitors.
It primarily uses fluorescent lights.	It mainly uses light-emitting diodes.
In LCDs, usually fluorescent lights are located at the backside of the screen.	Usually, light-emitting diodes are located around the edges or backside of the screen.
LCDs are less energy efficient than LEDs and are thicker in size.	LEDs are more energy-efficient and are much thinner in size as compared to LCDs.
Its resolution is low.	Its resolution is high.
Its contrast ratio is high.	Its contrast ratio is low.

Direct current can reduce the span life of LCDs.	Direct current does not have any effect on LEDs.
LCDs display area is large.	LEDs display area is small.
The switching time of LCD is slow.	The switching time of LED is fast.

PRINTERS

● Printer is an output device, which is used to print information on paper.

There are two types of printers −

✓ **Impact Printers**
✓ **Non-Impact Printers**

✓ **Impact Printers**
Impact printers print the characters by striking them on the ribbon, which is then pressed on the paper.

Characteristics of Impact Printers are the following −

1. Very low consumable costs
2. Very noisy
3. Useful for bulk printing due to low cost
4. There is physical contact with the paper to produce an image

These printers are of two types −

1. Character printers
2. Line printers
3. Character Printers

1. Character printers
● Character printers are the printers which print one character at a time.

These are further divided into two types:

✓ **Dot Matrix Printer(DMP)**
✓ **Daisy Wheel**

✓ **Dot Matrix Printer**
● In the market, one of the most popular printers is Dot Matrix Printer.
● These printers are popular because of their ease of printing and economical price.

● Each character printed is in the form of pattern of dots and head consists of a Matrix of Pins of size (5*7, 7*9, 9*7 or 9*9) which comes out to form a character which is why it is called Dot Matrix Printer.

Advantages
● Inexpensive
● Widely Used
● Other language characters can be printed

Disadvantages
● Slow Speed
● Poor Quality

✓ **Daisy Wheel**
● Head is lying on a wheel and pins corresponding to characters are like petals of Daisy (flower) which is why it is called Daisy Wheel Printer.
● These printers are generally used for word-processing in offices that require a few letters to be sent here and there with very nice quality.

Advantages
● More reliable than DMP
● Better quality
● Fonts of character can be easily changed

Disadvantages
● Slower than DMP
● Noisy
● More expensive than DMP

2. **Line Printers**
● Line printers are the printers which print one line at a time.

- These are of two types −

  ✓ **Drum Printer**
  ✓ **Chain Printer**

  ✓ **Drum Printer**
- This printer is like a drum in shape hence it is called drum printer.
- The surface of the drum is divided into a number of tracks.
- Total tracks are equal to the size of the paper, i.e. for a paper width of 132 characters, drum will have 132 tracks.
- A character set is embossed on the track.
- Different character sets available in the market are 48 character set, 64 and 96 characters set.
- One rotation of drum prints one line. Drum printers are fast in speed and can print 300 to 2000 lines per minute.

Advantages
- Very high speed

Disadvantages
- Very expensive
- Characters fonts cannot be changed

  ✓ **Chain Printer**
- In this printer, a chain of character sets is used; hence it is called Chain Printer.
- A standard character set may have 48, 64, or 96 characters.

Advantages
- Character fonts can easily be changed.
- Different languages can be used with the same printer.

Disadvantages
- Noisy

  ✓ **Non-impact Printers**
- Non-impact printers print the characters without using the ribbon. These printers print a complete page at a time, thus they are also called as Page Printers.

These printers are of two types −

✓ **Laser Printers**
✓ **Inkjet Printers**

Characteristics of Non-impact Printers
- Faster than impact printers
- They are not noisy
- High quality
- Supports many fonts and different character size

✓ **Laser Printers**
- These are non-impact page printers.
- They use laser lights to produce the dots needed to form the characters to be printed on a page.

Advantages
- Very high speed
- Very high quality output
- Good graphics quality
- Supports many fonts and different character size

Disadvantages
- Expensive
- Cannot be used to produce multiple copies of a document in a single printing

✓ **Inkjet Printers**
- Inkjet printers are non-impact character printers based on a relatively new technology.
- They print characters by spraying small drops of ink onto paper.
- Inkjet printers produce high quality output with presentable features.

- They make less noise because no hammering is done and these have many styles of printing modes available.
- Color printing is also possible. Some models of Inkjet printers can produce multiple copies of printing also.

Advantages
- High quality printing
- More reliable

Disadvantages
- Expensive as the cost per page is high
- Slow as compared to laser printer

I/O INTERFACING

- The method that is used to transfer information between internal storage and external I/O devices is known as I/O interface.
- The CPU is interfaced using special communication links by the peripherals connected to any computer system.
- These communication links are used to resolve the differences between CPU and peripheral.
- There exists special hardware components between CPU and peripherals to supervise and synchronize all the input and output transfers that are called interface units.

Mode of Transfer:
- The binary information that is received from an external device is usually stored in the memory unit.
- The information that is transferred from the CPU to the external device is originated from the memory unit.
- CPU merely processes the information but the source and target is always the memory unit.
- Data transfer between CPU and the I/O devices may be done in different modes.

- Data transfer to and from the peripherals may be done in any of the three possible ways

✓ **Programmed I/O.**
✓ **Interrupt- initiated I/O.**
✓ **Direct memory access (DMA).**

Now let's discuss each mode one by one.

✓ **Programmed I/O:**
- It is due to the result of the I/O instructions that are written in the computer program.
- Each data item transfer is initiated by an instruction in the program.
- Usually the transfer is from a CPU register and memory.
- In this case it requires constant monitoring by the CPU of the peripheral devices.

Example of Programmed I/O:
- In this case, the I/O device does not have direct access to the memory unit.
- A transfer from I/O device to memory requires the execution of several instructions by the CPU, including an input instruction to transfer the data from device to the CPU and store instruction to transfer the data from CPU to memory.
- In programmed I/O, the CPU stays in the program loop until the I/O unit indicates that it is ready for data transfer.
- This is a time consuming process since it needlessly keeps the CPU busy.
- This situation can be avoided by using an interrupt facility.

✓ **Interrupt- initiated I/O:**
- Since in the above case we saw the CPU is kept busy unnecessarily.
- This situation can very well be avoided by using an interrupt driven method for data transfer.
- By using interrupt facility and special commands to inform the interface to issue an interrupt request signal whenever data is available from any device.
- In the meantime the CPU can proceed for any other program execution.
- The interface meanwhile keeps monitoring the device.
- Whenever it is determined that the device is ready for data transfer it initiates an interrupt request signal to the computer.
- Upon detection of an external interrupt signal the CPU stops momentarily the task that it was already performing, branches to the service program to process the I/O transfer, and then return to the task it was originally performing.

Note:
- Both the methods programmed I/O and Interrupt-driven I/O require the active intervention of the processor to transfer data between memory and the I/O module, and any data transfer must transverse a path through the processor.
- Thus both these forms of I/O suffer from two inherent drawbacks.
- The I/O transfer rate is limited by the speed with which the processor can test and service a device.
- The processor is tied up in managing an I/O transfer; a number of instructions must be executed for each I/O transfer.

✓ **Direct Memory Access:**
- The data transfer between a fast storage media such as magnetic disk and memory unit is limited by the speed of the CPU.
- Thus we can allow the peripherals directly communicate with each other using the memory buses, removing the intervention of the CPU.
- This type of data transfer technique is known as DMA or direct memory access.
- During DMA the CPU is idle and it has no control over the memory buses.

● The DMA controller takes over the buses to manage the transfer directly between the I/O devices and the memory unit.
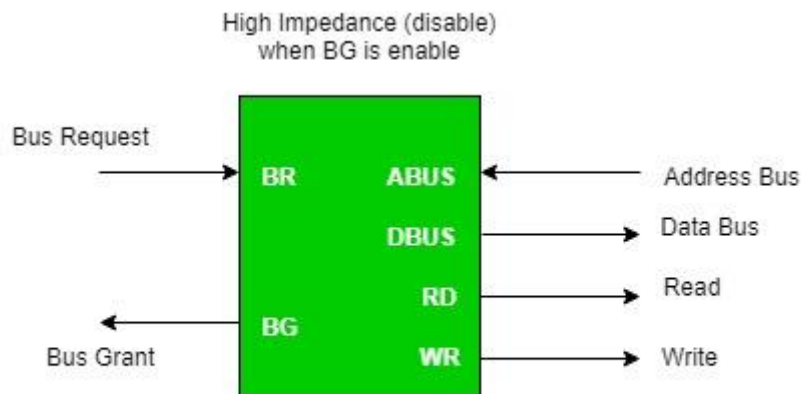

Figure - CPU Bus Signals for DMA Transfer

Bus Request:
  ● It is used by the DMA controller to request the CPU to relinquish the control of the buses.

Bus Grant:
  ● It is activated by the CPU to inform the external DMA controller that the buses are in high impedance state and the requesting DMA can take control of the buses.
  ● Once the DMA has taken the control of the buses it transfers the data. This transfer can take place in many ways.

Types of DMA transfer using DMA controller:

Burst Transfer:
  ● DMA returns the bus after complete data transfer. A register is used as a byte count, being decremented for each byte transfer, and upon the byte count reaching zero, the DMAC will release the bus.
  ● When the DMAC operates in burst mode, the CPU is halted for the duration of the data transfer.

Steps involved are:

✓ Bus grant request time.
✓ Transfer the entire block of data at transfer rate of device because the device is usually slow than the speed at which the data can be transferred to CPU.
✓ Release the control of the bus back to CPU

So, total time taken to transfer the N bytes
= Bus grant request time + (N) * (memory transfer rate) + Bus release control time.

Where,
X µsec =data transfer time or preparation time (words/block)
Y µsec =memory cycle time or cycle time or transfer time (words/block)
% CPU idle (Blocked) =(Y/X+Y)*100

% CPU Busy=(X/X+Y)*100

Cyclic Stealing:
- An alternative method in which DMA controller transfers one word at a time after which it must return the control of the buses to the CPU.
- The CPU delays its operation only for one memory cycle to allow the direct memory I/O transfer to "steal" one memory cycle.

Steps Involved are:

✔ Buffer the byte into the buffer
✔ Inform the CPU that the device has 1 byte to transfer (i.e. bus grant request)
✔ Transfer the byte (at system bus speed)
✔ Release the control of the bus back to CPU.
- Before moving on transfer next byte of data, device performs step 1 again so that bus isn't tied up and the transfer won't depend upon the transfer rate of device.
- **So, for 1 byte of transfer of data, time taken by using cycle stealing mode (T) = time required for bus grant + 1 bus cycle to transfer data + time required to release the bus,**

It will be

N x T

- In cycle stealing mode we always follow pipelining concept that when one byte is getting transferred then Device is parallel preparing the next byte.
- "The fraction of CPU time to the data transfer time" if asked then cycle stealing mode is used.

Where,
X µsec =data transfer time or preparation time (words/block)
Y µsec =memory cycle time or cycle time or transfer time (words/block)
% CPU idle (Blocked) =(Y/X)*100
% CPU busy=(X/Y)*100

Interleaved mode:
- In this technique, the DMA controller takes over the system bus when the microprocessor is not using it.
- An alternate half cycle i.e. half cycle DMA + half cycle processor.

POLLING
- In polling is not a hardware mechanism, it's a protocol in which CPU steadily checks whether the device needs attention.
- Wherever device tells process unit that it desires hardware processing, in polling process unit keeps asking the I/O device whether or not it desires CPU processing.
- The CPU ceaselessly check every and each device hooked up thereto for sleuthing whether or not any device desires hardware attention.
- Each device features a command-ready bit that indicates the standing of that device, i.e., whether or not it's some command to be dead by hardware or not.

- If command bit is ready one, then it's some command to be dead else if the bit is zero, then it's no commands.
- Let's see that the difference between interrupt and polling:

S.NO	Interrupt	Polling
1.	In interrupt, the device notices the CPU that it requires its attention.	Whereas, in polling, CPU steadily checks whether the device needs attention.
2.	An interrupt is not a protocol, it's a hardware mechanism.	Whereas it isn't a hardware mechanism, it's a protocol.
3.	In interrupt, the device is serviced by interrupt handler.	While in polling, the device is serviced by CPU.
4.	Interrupt can take place at any time.	Whereas CPU steadily ballots the device at regular or proper interval.
5.	In interrupt, interrupt request line is used as indication for indicating that device requires servicing.	While in polling, Command ready bit is used as indication for indicating that device requires servicing.
6.	In interrupts, processor is simply disturbed once any device interrupts it.	On the opposite hand, in polling, processor waste countless processor cycles by repeatedly checking the command-ready little bit of each device.

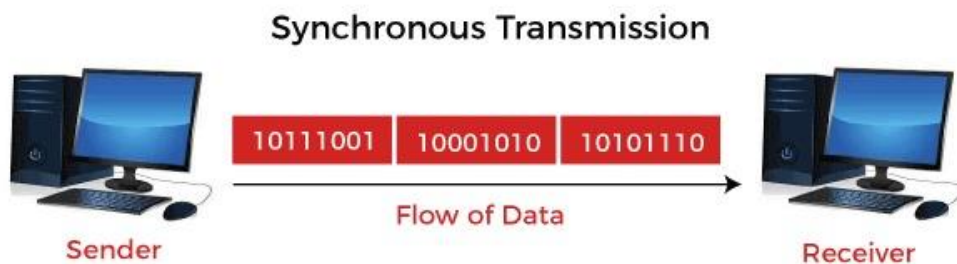## <u>SERIAL COMMUNICATION</u>

- Serial communication is the process of sequentially transferring the information/bits on the same channel.
- Due to this, the cost of wire will be reduced, but it slows the transmission speed.
- Generally, communication can be described as the process of interchanging information between individuals in the form of audio, video, verbal words, and written documents.
- The serial protocol is run on every device that can be our mobile, personal computers, and many more with the help of following some protocols.
- The protocol is a type of reliable and secure form of communication that contains a set of rules addressed with the help of a source host and a destination host.
- In serial communication, binary pulses are used to show the data.
- Binary contains the two numbers 0 and 1.
- 0 is used to show the LOW or 0 Volts, and 1 is used to show the HIGH or 5 Volts.

- The serial communication can either be asynchronous or synchronous.

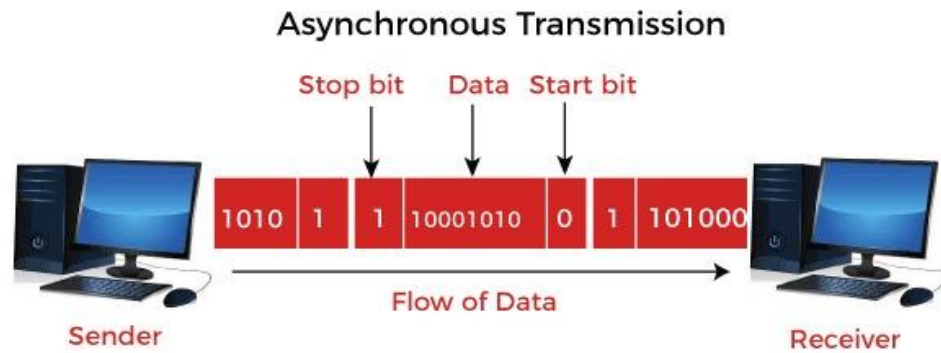## Serial Communication
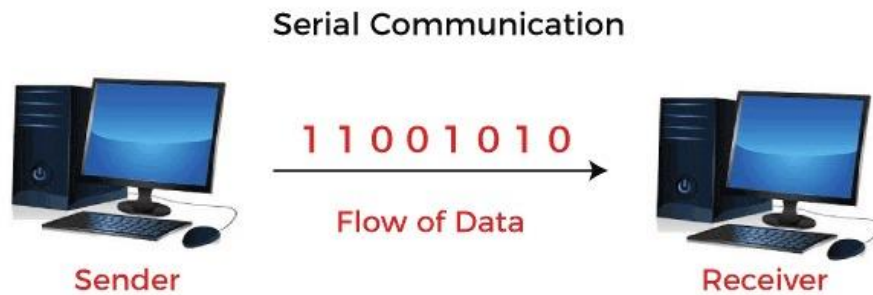
Synchronous Communication

- In synchronous communication, the frames or data will be constructed with the help of combining the groups of bits.
- That frames will be continuously sent in time with a master clock.
- It uses a synchronized clock frequency to operate the data of sender or receiver.
- In synchronous communication, there is no need to use the gaps, start bits and stop bits.
- The time taken by the sender and receiver is synced that's why the frequency of timing error will be less, and the data will move faster.
- On the basis of the timing being synced correctly between the sender and receiver devices, the data accuracy is totally dependent.
- The synchronous serial transmission is more expensive as compared to asynchronous serial transmission.

Synchronous Transmission

Asynchronous Communication

- In asynchronous communication, the groups of bits will be treated as an independent unit, and these data bits will be sent at any point in time.
- In order to make synchronization between sender and receiver, the stop bits and start bits are used between the data bytes.
- These bits are useful to ensure that the data is correctly sent.
- The time taken by data bits of sender and receiver is not constant, and the time between transmissions will be provided by the gaps.
- In asynchronous communication, we don't require synchronization between the sender and receiver devices, which is the main advantage of asynchronous communication.
- This method is also cost-effective.
- In this method, there can be a case when data transmission is slow, but it is not compulsory, and it is the main disadvantage of the asynchronous method.

Asynchronous Transmission

Stop bit Data Start bit

| 1010 | 1 | 1 | 10001010 | 0 | 1 | 101000 |

Flow of Data

Sender   Receiver

UART

- UART represents Universal Asynchronous Receiver Transmitter.
- It is dedicated to hardware related to serial communication.
- UART is one of the most generally used serial communication techniques.
- UART is being used in several applications like GPS Receivers, Bluetooth Modules, GSM and GPRS Modems, Wireless Communication Systems, RFID-based applications, etc.

Functions of the UART
There are various functions of UART which are as follows −

- It can change parallel data into serial data for outbound communications
- It can change serial data into parallel data for inbound communications
- It can add a parity checking bit on outbound transmissions and tests the parity bit for inbound transmissions
- It can handle interrupt requests and device administration, which can need the computer and the device to organize the speed of operation.

UART Communication
- In this communication, there are two types of UARTs such as transmitting UART and receiving UART, and the communication between these two can be done explicitly by each other.
- The data transmission of a UART can be completed by using a data bus in the form of parallel by other devices such as a microcontroller, memory, CPU, etc.
- After receiving the parallel data from the bus, it forms a data packet by inserting three bits like start, stops, and parity.
- It reads the data packet bit by bit and converts the received data into the parallel form to remove the three bits of the data packet.

→ Packet ←

| 1 Start Bit | Data Bits (5 to 9) | 0 to 1 (Parity Bits) | 1 to 2 (Stop Bits) |

→ Data Frame ←
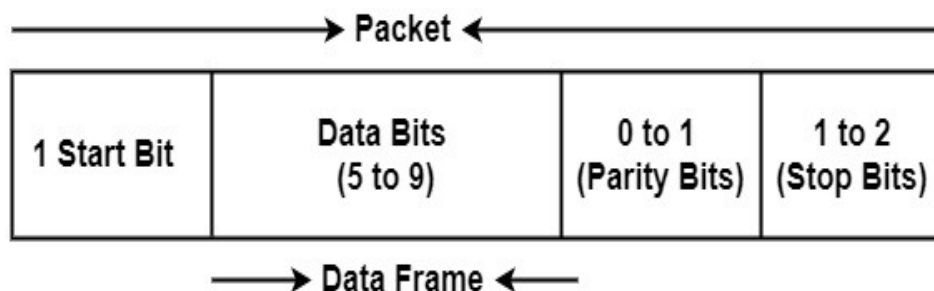
Start Bit
- Start-bit is also referred to as a synchronization bit that is located before the actual data.
- Usually, an inactive data transmission line is reserved at a high-voltage level.
- To start the data transmission, the UART transmission burden the data line from a high voltage level (1) to a low voltage level (0).

Stop Bit
- The Stop Bit is located at the ending of the data packet.
- Generally, this bit is 2-bits lengthy but commonly on bit only used.
- It can stop the broadcast; the UART maintains the data-line on high voltage.

Parity Bit
- Parity bit allows the receiver to provide whether the collected record is right or not.
- It is a low-level fault checking system & parity bit is accessible in two ranges including Even Parity and Odd Parity.

Data Bits or Data Frame
- The data bits involve the real data being a transfer from the sender to the receiver.
- The data frame length can be between 5 & 8.
- If the parity bit is not used when the data frame length could be 9-bit long.
- Usually, the LSB of the data to be sent first then it is very beneficial for transmitting.

RS232

- RS-232 is a standard communication protocol for connecting computers and their peripheral devices to enable serial data exchange.
- n simple terms, RS232 represents the voltage for the path used for data exchange between the devices.
- It determines the common voltage and signal level, common pin wire configuration and minimum, amount of control signals.
- RS232 represents the signals connecting between DTE and DCE.
- Therefore, DTE represents Data Terminal Equipment and an example for DTE is a computer.
- DCE represents Data Communication Equipment or Data Circuit Terminating Equipment and an example for DCE is a modem.
- RS232 was introduced in the 1960s and was initially referred to as EIA Recommended Standard 232.
- RS232 is one of the first serial communication standards with provided simple connectivity and compatibility across multiple manufacturers.
- The DTEs in RS32 are electromechanical typewriters and DCEs are modems.

- The RS-232 interface works in a mixture with UART (universal asynchronous receiver/transmitter).
- It is a part of an integrated circuit integrated within the processor or controller.
- It creates bytes and sequentially sends the single bits in a frame.
- A frame is a defined structure, carrying a meaningful sequence of bits or bytes of data.
- It has a start bit followed by 8 data bits, a parity bit, and a stop bit.
- Once data is modified into bits separate line drivers are used to changing the logic level of UART to RS-232 logic.
- Finally, the signals are shared along the interface cable at the particular voltage level of RS-232.

- The data is transmitted serially on RS232.
- Each bit is transmitted one after the other.
- This mode of transmission needed that the receiver is aware of when the actual data bits are appearing to synchronize itself with arriving data.
- Therefore, logic 0 is sent as a start bit.
- The start bit in the frame signals the receiver that a new character is arriving.
- Once the receiver acknowledges the next five to eight bits are sent which defines the character.
- This is followed by the parity bit used for error detection.
- The parity bit can determine an even or an odd number of ones in the set of bits.
- For error detection, it can add a more bit to the data word.
- The transmitter evaluates the value of the bit based on the data sent and the receiver also implements the same computation.
- It tests the parity value to the computed value.
- The stop bit supports the receiver to recognize the end of the message.
- The start bit continually has space value and the stop bit has mark value.
- This generates a framing error condition in the receiving UART.
- The device then attempts to resynchronize on more incoming bits.
- At the other end again the line driver interface changes it into UART compatible logic levels.
- At the destination, a second UART re-assembles the bits into bytes.
- This is how RS232 creates the data exchange compatible and reliable.

USB

- USB represents Universal Serial Bus.
- It is an industry-standard developed in the mid-1990s that represents the cables, connectors, and communication protocols used in a bus for connection, communication, and power supply among computer and electronic devices.
- USB operates at 12 Mbps with particular consideration for low-cost peripherals.
- It provides up to 127 devices with both synchronous and asynchronous data transfers.
- A USB cable has a rectangular "TYPE A" plug at the computer end and a square "TYPE B" plug at the peripheral end.

- A Universal Serial Bus (USB) is a common interface that allows the connection between devices and host controllers including a personal computer (PC).

- It connects peripheral devices including digital cameras, mice, keyboards, printers, scanners, media devices, external hard drives, and flash drives.
- Hot swapping is one of the major advantages of USB, which enables a device to be eliminated from the system without the requirement of rebooting the system.
- Because previous ports required that a PC be reboot while inserting or deleting a new device with them.
- By rebooting the system, the device will be reconfigured and can avoid electrostatic discharge.

Advantages of USB

There are various advantages of USB which are as follows −

✔ **Single Interface for multiple devices−** The versatile feature of USB eliminates the complexity of several connector types and hardware needed for each peripheral.

✔ **Compact Size** − USB sockets are small in dimensions as compared to RS232 or parallel ports. There is no external power required. The USB interface was established from the first day to double duty as a DC power supply. Several host devices through its USB port can supply 5V DC delivering 500mA (USB 1.0 and 2.0) to 900 mA (USB 3.0) to the peripheral.

✔ **Speed** − USB offers multiple speed modes which create it more effective and swift compared to RS232 and parallel ports. It provides speed ranges from 1.5Mbit/s to 5Gbit/s. With the introduction of USB 3.1 in 2013, the speed has been raised to 10Gbit/s.

✔ **Reliability** − The USB protocol can hook errors during data transfer and inform the transmitter to retransmit the information. The generic USB driver and unique driver software provide error-free data communication.

Disadvantages of USB

There are various disadvantages of USB which are as follows −

✔ **Peer to Peer Communication** − According to the USB standard, the connection takes place between the host and the peripheral. Two hosts cannot connect directly with each other. The same is the case for a peripheral. Similarly, interfaces like FireWire provides peripheral to peripheral communication. For overcoming this drawback, the USB introduced the term OTG (On the Go). The OTG device usually functions as a peripheral, but it can also function as a host with some small capability when needed.

✔ **Distance** − As per the USB standards, the connecting cable can be seeing 5 meters, further which, USB hubs require to be used for enlarging the connectivity.

✔ **Broadcasting** − Universal Serial Bus does not support the broadcasting nature, only single messages can be connected between the host and peripheral.

HIGH PERFORMANCE COMPUTING

- The term High-performance computing is occasionally used as a synonym for supercomputing.
- Although technically a supercomputer is a system that performs at or near currently highest operational rate for computers.
- Some supercomputers work at more than a petaflop (10^{12}) floating points opm per second.
- The most common HPC system all scientific engineers & academic institutions.
- Some Government agencies particularly military are also relying on APC for complex applications.

High-performance Computers:

- Processors, memory, disks, and OS are elements of high-performance computers of interest to small & medium size businesses today are really clusters of computers.
- Each individual computer in a commonly configured small cluster has between one and four processors and today's processors typically are from 2 to 4 crores, HPC people often referred to individual computers in a cluster as nodes.
- A cluster of interest to a small business could have as few as 4 nodes on 16 crores.
- Common cluster size in many businesses is between 16 & 64 crores or from 64 to 256 crores.
- The main reason to use this is that in its individual node can work together to solve a problem larger than any one computer can easily solve.
- These nodes are so connected that they can communicate with each other in order to produce some meaningful work.
- There are two popular HPC's software i. e, Linux, and windows.
- Most of installations are in Linux because of its supercomputer but one can use it with his / her requirements.

Importance of High performance Computing:
- It is used for scientific discoveries, game-changing innovations, and to improve quality of life.
- It is a foundation for scientific & industrial advancements.
- It is used in technologies like IoT, AI, 3D imaging evolves & amount of data that is used by organization is increasing exponentially to increase ability of a computer, we use High-performance computer.
- HPC is used to solve complex modeling problems in a spectrum of disciplines. It includes AI, Nuclear Physics, Climate Modelling, etc.
- HPC is applied to business uses as well as data warehouses & transaction processing.

Need of High performance Computing:
- It will complete a time-consuming operation in less time.
- It will complete an operation under a light deadline and perform a high numbers of operations per second.
- It is fast computing, we can compute in parallel over lot of computation elements CPU, GPU, etc.
- It set up very fast network to connect between elements.

Need of ever increasing Performance:
- Climate modeling
- Drug discovery
- Data Analysis
- Protein folding
- Energy research

Main Components:

User → Computers cluster → Data storage

- In HPC architecture computer servers must be networked together in a cluster.
- Programs & algorithms are run simultaneously on server in cluster.
- Cluster is networked to data storage to capture output.
- Together this achieves a complete set of meaningful work.

PIPELINING

- To improve the performance of a CPU we have two options:
  1) Improve the hardware by introducing faster circuits.
  2) Arrange the hardware such that more than one operation can be performed at the same time.
- Since there is a limit on the speed of hardware and the cost of faster circuits is quite high, we have to adopt the 2nd option.

- Pipelining is a process of arrangement of hardware elements of the CPU such that its overall performance is increased.
- Simultaneous execution of more than one instruction takes place in a pipelined processor.
- Let us see a real-life example that works on the concept of pipelined operation.
- Consider a water bottle packaging plant.
- Let there be 3 stages that a bottle should pass through, Inserting the bottle(I), Filling water in the bottle(F), and Sealing the bottle(S).
- Let us consider these stages as stage 1, stage 2 and stage 3 respectively.
- Let each stage take 1 minute to complete its operation.
- Now, in a non-pipelined operation, a bottle is first inserted in the plant, after 1 minute it is moved to stage 2 where water is filled.
- Now, in stage 1 nothing is happening.
- Similarly, when the bottle moves to stage 3, both stage 1 and stage 2 are idle.
- But in pipelined operation, when the bottle is in stage 2, another bottle can be loaded at stage 1.
- Similarly, when the bottle is in stage 3, there can be one bottle each in stage 1 and stage 2.
- So, after each minute, we get a new bottle at the end of stage 3.
- Hence, the average time taken to manufacture 1 bottle is:

> **Without pipelining = 9/3 minutes = 3m**
> **I F S | | | | | |**
> **| | | I F S | | |**
> **| | | | | | I F S (9 minutes)**
>
> **With pipelining = 5/3 minutes = 1.67m**
>
> **I F S | |**
> **| I F S |**
> **| | I F S (5 minutes)**

- Thus, pipelined operation increases the efficiency of a system.

Design of a basic pipeline

- In a pipelined processor, a pipeline has two ends, the input end and the output end.

- Between these ends, there are multiple stages/segments such that the output of one stage is connected to the input of the next stage and each stage performs a specific operation.
- Interface registers are used to hold the intermediate output between two stages.
- These interface registers are also called latch or buffer.
- All the stages in the pipeline along with the interface registers are controlled by a common clock.
- Execution in a pipelined processor Execution sequence of instructions in a pipelined processor can be visualized using a space-time diagram.
- For example, consider a processor having 4 stages and let there be 2 instructions to be executed.
- We can visualize the execution sequence through the following space-time diagrams:

Non-overlapped execution:

Stage / Cycle	1	2	3	4	5	6	7	8
S1	I1				I2			
S2		I1				I2		
S3			I1				I2	
S4				I1				I2

Total time = 8 Cycle

Overlapped execution:

Stage / Cycle	1	2	3	4	5
S1	I1	I2			
S2		I1	I2		
S3			I1	I2	
S4				I1	I2

Total time = 5 Cycle
Pipeline Stages
- RISC processor has 5 stage instruction pipeline to execute all the instructions in the RISC instruction set.
- Following are the 5 stages of the RISC pipeline with their respective operations:

✔ **Stage 1 (Instruction Fetch)**
  - In this stage the CPU reads instructions from the address in the memory whose value is present in the program counter.
✔ **Stage 2 (Instruction Decode)**
  - In this stage, instruction is decoded and the register file is accessed to get the values from the registers used in the instruction.
✔ **Stage 3 (Instruction Execute)**
  - In this stage, ALU operations are performed.
✔ **Stage 4 (Memory Access)**
  - In this stage, memory operands are read and written from/to the memory that is present in the instruction.
✔ **Stage 5 (Write Back)**

- In this stage, computed/fetched value is written back to the register present in the instructions.

Performance of a pipelined processor

- Consider a 'k' segment pipeline with clock cycle time as 'Tp'.
- Let there be 'n' tasks to be completed in the pipelined processor.
- Now, the first instruction is going to take 'k' cycles to come out of the pipeline but the other 'n – 1' instructions will take only '1' cycle each, i.e., a total of 'n – 1' cycles.
- So, time taken to execute 'n' instructions in a pipelined processor:

$$ET_{pipeline} = k + n - 1 \text{ cycles}$$
$$= (k + n - 1) \text{ Tp}$$

- In the same case, for a non-pipelined processor, the execution time of 'n' instructions will be:

$$ET_{non-pipeline} = n * k * Tp$$

- So, speedup (S) of the pipelined processor over the non-pipelined processor, when 'n' tasks are executed on the same processor is:

S = Performance of Non-pipelined processor / Performance of pipelined processor

- As the performance of a processor is inversely proportional to the execution time, we have,

$$S = ET_{non-pipeline} / ET_{pipeline}$$
$$=> S = [n * k * Tp] / [(k + n - 1) * Tp]$$
$$S = [n * k] / [k + n - 1]$$

- When the number of tasks 'n' is significantly larger than k, that is, $n \gg k$

$$S = n * k / n$$
$$S = k$$

where 'k' are the number of stages in the pipeline.

- Also, Efficiency = Given speed up / Max speed up = S / S_{max}
- We know that $S_{max} = k$
- So, Efficiency = S / k
- Throughput = Number of instructions / Total time to complete the instructions
- So, Throughput = $n / (k + n - 1) * Tp$

Note: The cycles per instruction (CPI) value of an ideal pipelined processor is 1.

PARALLEL PROCESSING

- Parallel processing can be described as a class of techniques which enables the system to achieve simultaneous data-processing tasks to increase the computational speed of a computer system.
- A parallel processing system can carry out simultaneous data-processing to achieve faster execution time.
- For instance, while an instruction is being processed in the ALU component of the CPU, the next instruction can be read from memory.

- The primary purpose of parallel processing is to enhance the computer processing capability and increase its throughput, i.e. the amount of processing that can be accomplished during a given interval of time.
- A parallel processing system can be achieved by having a multiplicity of functional units that perform identical or different operations simultaneously.
- The data can be distributed among various multiple functional units.
- The following diagram shows one possible way of separating the execution unit into eight functional units operating in parallel.

- The operation performed in each functional unit is indicated in each block if the diagram:
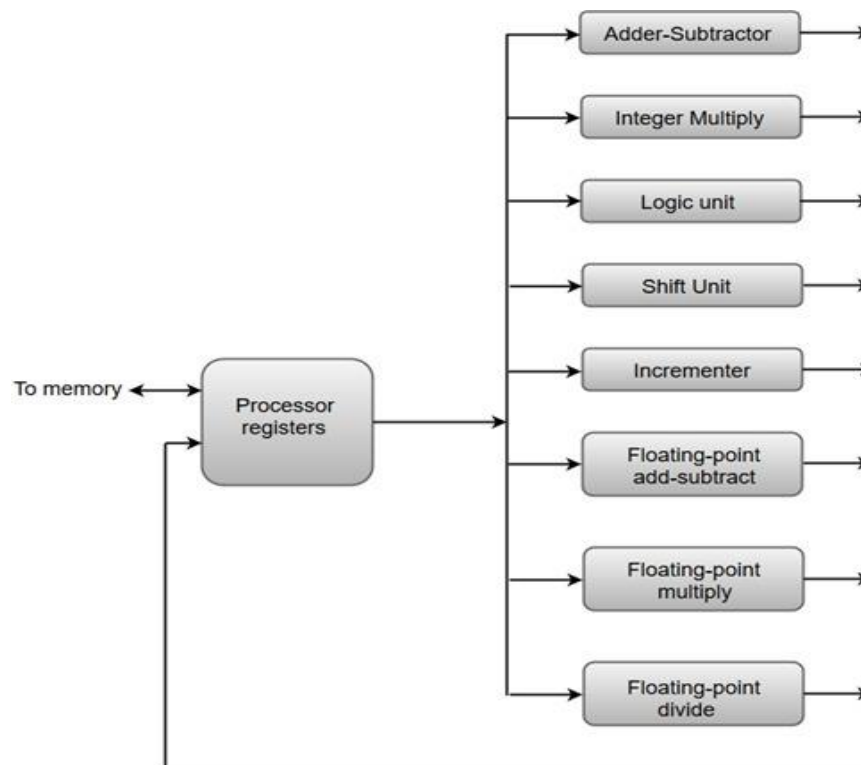
- The adder and integer multiplier performs the arithmetic operation with integer numbers.
- The floating-point operations are separated into three circuits operating in parallel.
- The logic, shift, and increment operations can be performed concurrently on different data.
- All units are independent of each other, so one number can be shifted while another number is being incremented.
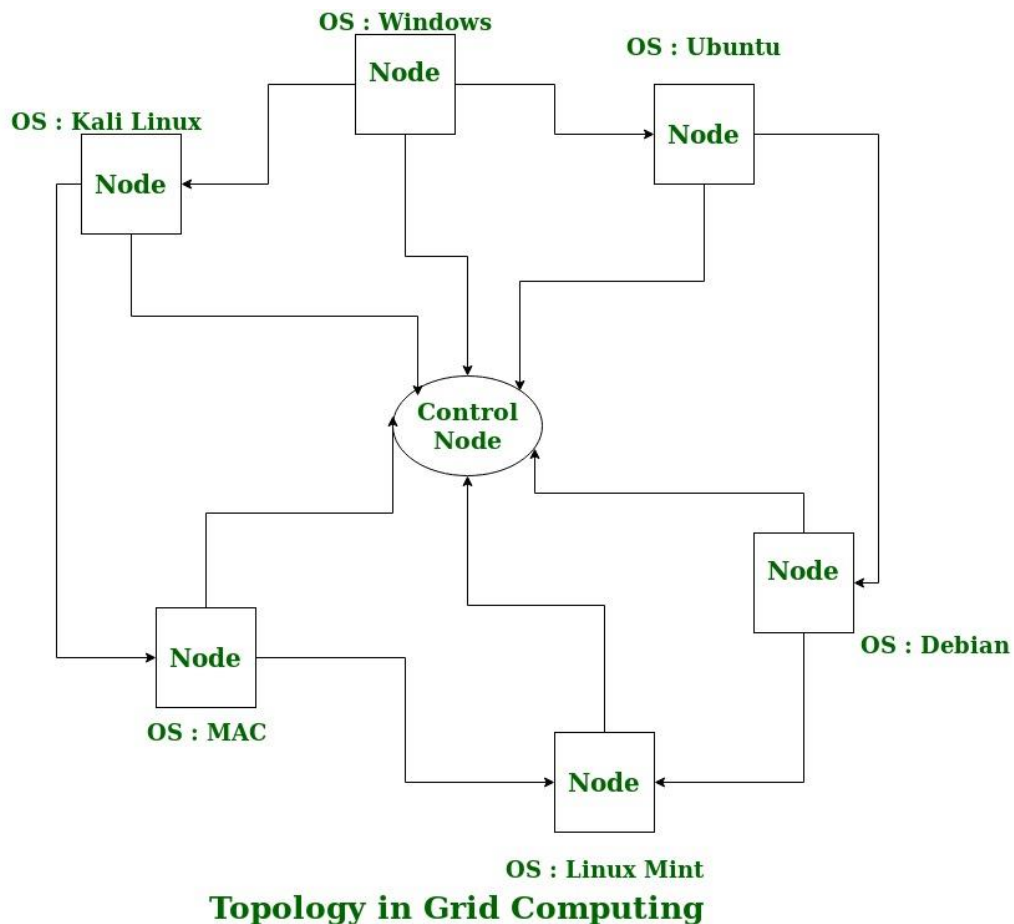
GRID COMPUTING

- Grid Computing can be defined as a network of computers working together to perform a task that would rather be difficult for a single machine.
- All machines on that network work under the same protocol to act as a virtual supercomputer.

- The task that they work on may include analyzing huge datasets or simulating situations that require high computing power.
- Computers on the network contribute resources like processing power and storage capacity to the network.
- Grid Computing is a subset of distributed computing, where a virtual supercomputer comprises machines on a network connected by some bus, mostly Ethernet or sometimes the Internet.
- It can also be seen as a form of Parallel Computing where instead of many CPU cores on a single machine, it contains multiple cores spread across various locations.
- The concept of grid computing isn't new, but it is not yet perfected as there are no standard rules and protocols established and accepted by people.

Working:

- A Grid computing network mainly consists of these three types of machines

    1. Control Node:
- A computer, usually a server or a group of servers which administrates the whole network and keeps the account of the resources in the network pool.
    2. Provider:
- The computer contributes its resources to the network resource pool.
    3. User:
- The computer that uses the resources on the network.

- When a computer makes a request for resources to the control node, the control node gives the user access to the resources available on the network.
- When it is not in use it should ideally contribute its resources to the network.
- Hence a normal computer on the node can swing in between being a user or a provider based on its needs.
- The nodes may consist of machines with similar platforms using the same OS called homogeneous networks, else machines with different platforms running on various different OSs called heterogeneous networks.
- This is the distinguishing part of grid computing from other distributed computing architectures.

- For controlling the network and its resources a software/networking protocol is used generally known as Middleware.
- This is responsible for administrating the network and the control nodes are merely its executors.
- As a grid computing system should use only unused resources of a computer, it is the job of the control node that any provider is not overloaded with tasks.
- Another job of the middleware is to authorize any process that is being executed on the network.
- In a grid computing system, a provider gives permission to the user to run anything on its computer, hence it is a huge security threat for the network.
- Hence a middleware should ensure that there is no unwanted task being executed on the network.
- The meaning of the term Grid Computing has changed over the years, according to "The Grid: Blueprint for a new computing infrastructure" by Ian Foster and Carl Kesselman published in 1999, the idea was to consume computing power like electricity is consumed from a power grid.

- This idea is similar to the current concept of cloud computing, whereas now grid computing is viewed as a distributed collaborative network.
- Currently, grid computing is being used in various institutions to solve a lot of mathematical, analytical, and physics problems.


Topology in Grid Computing

Advantages of Grid Computing:
- It is not centralized, as there are no servers required, except the control node which is just used for controlling and not for processing.
- Multiple heterogeneous machines i.e. machines with different Operating Systems can use a single grid computing network.
- Tasks can be performed parallelly across various physical locations and the users don't have to pay for them (with money).

Disadvantages of Grid Computing:
- The software of the grid is still in the involution stage.
- A super fast interconnects between computers resources is the need of hour.
- Licensing across many servers may make it prohibitive for some applications.
- Many groups are reluctant with sharing resources.

CLUSTER COMPUTING

- Cluster computing is a collection of tightly or loosely connected computers that work together so that they act as a single entity.
- The connected computers execute operations all together thus creating the idea of a single system.
- The clusters are generally connected through fast local area networks (LANs)
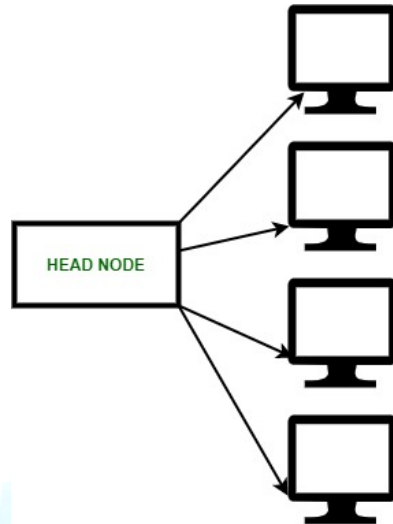
Importance of Cluster Computing

- Cluster computing gives a relatively inexpensive, unconventional to the large server or mainframe computer solutions.
- It resolves the demand for content criticality and process services in a faster way.
- Many organizations and IT companies are implementing cluster computing to augment their scalability, availability, processing speed and resource management at economic prices.
- It ensures that computational power is always available.

- It provides a single general strategy for the implementation and application of parallel high-performance systems independent of certain hardware vendors and their product decisions.
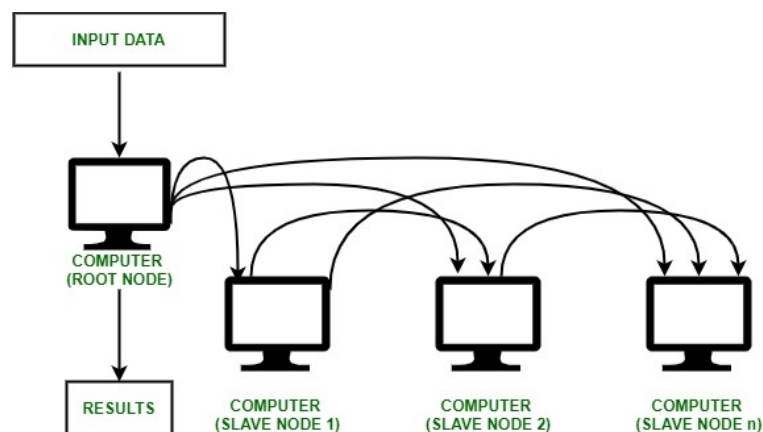
A SIMPLE CLUSTER COMPUTING LAYOUT

Types of Cluster computing:

1. High performance (HP) clusters:
- HP clusters use computer clusters and supercomputers to solve advance computational problems.
- They are used to performing functions that need nodes to communicate as they perform their jobs.
- They are designed to take benefit of the parallel processing power of several nodes.

2. Load-balancing clusters:
- Incoming requests are distributed for resources among several nodes running similar programs or having similar content.
- This prevents any single node from receiving a disproportionate amount of task.
- This type of distribution is generally used in a web-hosting environment.

3. High Availability (HA) Clusters:
- HA clusters are designed to maintain redundant nodes that can act as backup systems in case any failure occurs.
- Consistent computing services like business activities, complicated databases, customer services like e-websites and network file distribution are provided.
- They are designed to give uninterrupted data availability to the customers.

Classification of Cluster:

1. Open Cluster:
- IPs are needed by every node and those are accessed only through the internet or web. This type of cluster causes enhanced security concerns.

2. Close Cluster:
- The nodes are hidden behind the gateway node, and they provide increased protection. They need fewer IP addresses and are good for computational tasks.

Cluster Computing Architecture:

- It is designed with an array of interconnected individual computers and the computer systems operating collectively as a single standalone system.
- It is a group of workstations or computers working together as a single, integrated computing resource connected via high speed interconnects.
- A node – Either a single or a multiprocessor network having memory, input and output functions and an operating system.
- Two or more nodes are connected on a single line or every node might be connected individually through a LAN connection.

CLUSTER COMPUTING ARCHITECTURE
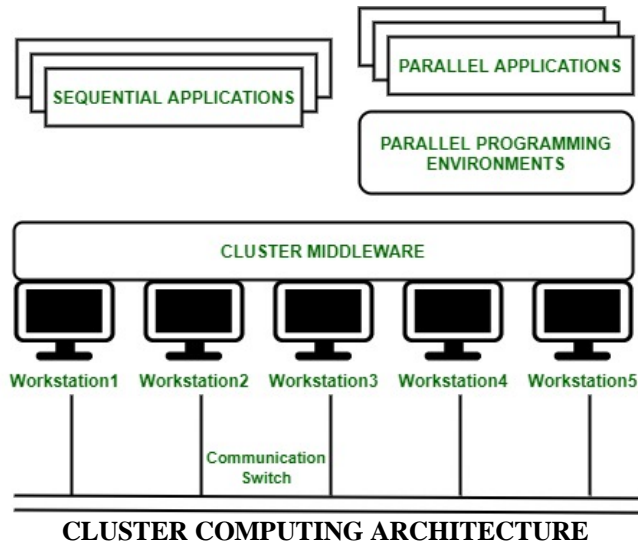
Components of a Cluster Computer:

- Cluster Nodes
- Cluster Operating System
- The switch or node interconnect
- Network switching hardware

CLUSTER COMPONENTS
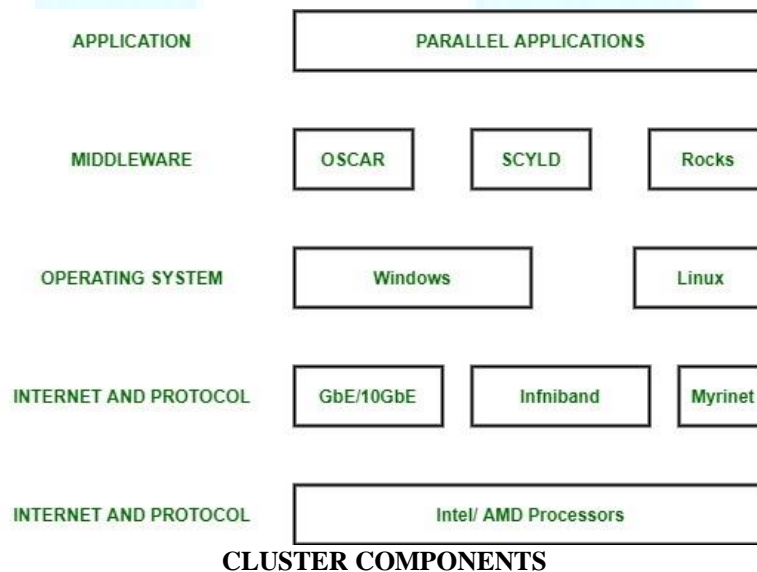
Advantages of Cluster Computing:
1. High Performance:
- The systems offer better and enhanced performance than that of mainframe computer networks.

2. Easy to manage:
- Cluster Computing is manageable and easy to implement.

3. Scalable:
- Resources can be added to the clusters accordingly.

4. Expandability:
- Computer clusters can be expanded easily by adding additional computers to the network.
- Cluster computing is capable of combining several additional resources or the networks to the existing computer system.

5. Availability:
- The other nodes will be active when one node gets failed and will function as a proxy for the failed node.
- This makes sure for enhanced availability.

6. Flexibility:
- It can be upgraded to the superior specification or additional nodes can be added.

Disadvantages of Cluster Computing:

1. High cost:
- It is not so much cost-effective due to its high hardware and its design.

2. Problem in finding fault:
- It is difficult to find which component has a fault.

3. More space is needed:
- Infrastructure may increase as more servers are needed to manage and monitor.

Applications of Cluster Computing:
- Various complex computational problems can be solved.
- It can be used in the applications of aerodynamics, astrophysics and in data mining.
- Weather forecasting.
- Image Rendering.
- Various e-commerce applications.
- Earthquake Simulation.
- Petroleum reservoir simulation.