**OPERATING SYSTEMS**

**UNIT-I**

**Chapter – 1**

**INTRODUCTION**

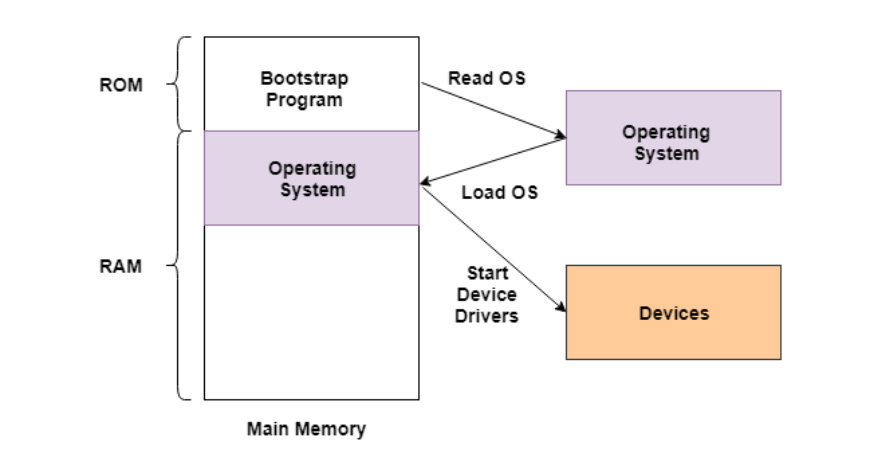
Computer System Organization, Computer System Structure.

**Computer System Organization**

**In this we have study the organization of the Computer System is as follows:**

1. **Computer Startup:**

* For a computer to start running—for instance, when it is powered up orrebooted—it needs to have an initial program to run.
* This initial program,or bootstrap program, tends to be simple.
* **A bootstrap program is the first code that is executed when the computer system is started. The entire operating system depends on the bootstrap program to work correctly as it loads the operating system.**
* A figure that demonstrates the use of the bootstrap program is as follows −



* In the above image, the bootstrap program is a part of ROM which is the non-volatile memory. The operating system is loaded into the RAM by the bootstrap program after the start of the computer system. Then the operating system starts the device drivers.

## Bootstrapping Process

* The bootstrapping process does not require any outside input to start. Any software can be loaded as required by the operating system rather than loading all the software automatically.
* The bootstrapping process is performed as a chain i.e., at each stage, it is the responsibility of the simpler and smaller program to load and execute the much more complicated and larger program. This means that the computer system improves in increments by itself.
* The booting procedure starts with the hardware procedures and then continues onto the software procedures that are stored in the main memory. The bootstrapping process involves self-tests, loading BIOS, configuration settings, hypervisor, operating system etc.

## Benefits of Bootstrapping

* Without bootstrapping, the computer user would have to download all the software components, including the ones not frequently required. With bootstrapping, only those software components need to be downloaded that are legitimately required and all extraneous components are not required. This process frees up a lot of space in the memory and consequently saves a lot of time.

1. **Computer System Structure:**

****

**Fig:** Abstract view of the components of a computer system.

* Computer system can be divided into four components:

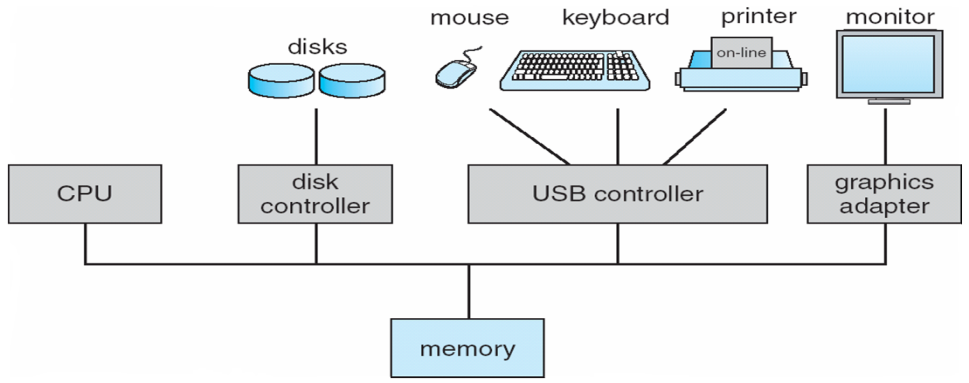
**Hardware:** Provides basic computing resources CPU, memory, I/O devices.

**Operating system:** Controls and coordinates use of hardware among various applications and users.

**Application programs:** Define the ways in which the system resources are used to solve the computing problems of the users (Word processors, compilers, web browsers, database systems, video games).

**Users:** People, machines, other computers.

1. **Computer System Operation:**



**Fig : A modern computer system.**

* A modern general-purpose computer system consists of one or more CPUsand a number of device controllers connected through a common bus thatprovides access to shared memory in above figure.
* Each device controller is incharge of a specific type of device (for example, disk drives, audio devices,or video displays).
* The CPU and the device controllers can execute in parallel,competing for memory cycles.
* To ensure orderly access to the shared memory,a memory controller synchronizes access to the memory.
* I/O devices and the CPU can execute concurrently.
* Each device controller has a local buffer.
* CPU moves data from/to main memory to/from local buffers.
* I/O is from the device to local buffer of controller.
* Device controller informs CPU that it has finished its operation by causing an interrupt.

1. **Interrupt Timeline:**

* **Interrupts are signals sent to the CPU by external devices, normally I/O devices. They tell the CPU to stop its current activities and execute the appropriate part of the operating system.**

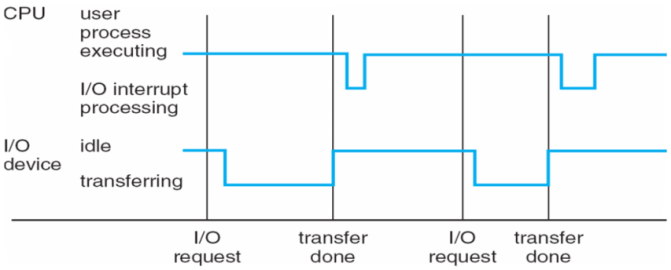


Fig: Interrupt timeline for a single process doing output.

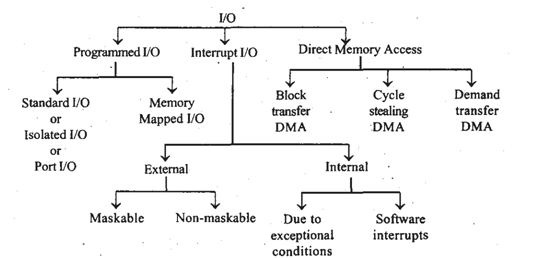
* When the CPU is interrupted, it stops what it is doing and immediatelytransfers execution to a fixed location. The fixed location usually containsthe starting address where the service routine for the interrupt is located.The interrupt service routine executes; on completion, the CPU resumes theinterrupted computation. A timeline of this operation is shown in Figure.

There are three types of interrupts:

* **Hardware Interrupts** are generated by hardware devices to signal that they need some attention from the OS. They may have just received some data (e.g., keystrokes on the keyboard or data on the ethernet card); or they have just completed a task which the operating system previous requested, such as transferring data between the hard drive and memory.
* **Software Interrupts** are generated by programs when they want to request a [system call](http://faculty.salina.k-state.edu/tim/ossg/glossary.html#term-system-call) to be performed by the operating system.
* **Traps** are generated by the CPU itself to indicate that some error or condition occurred for which assistance from the operating system is needed.
* Interrupts are important because they give the user better control over the computer. Without interrupts, a user may have to wait for a given application to have a higher priority over the CPU to be ran. This ensures that the CPU will deal with the process immediately.
* An operating system is interrupt driven.

1. **I/O Structure:**

* **I/O Structure consists of Programmed I/O, Interrupt driven I/O, DMS, CPU, Memory, External devices, these are all connected with the help of Peripheral I/O Buses and General I/O Buses.**
* Different types of I/O Present inside the system are shown below −

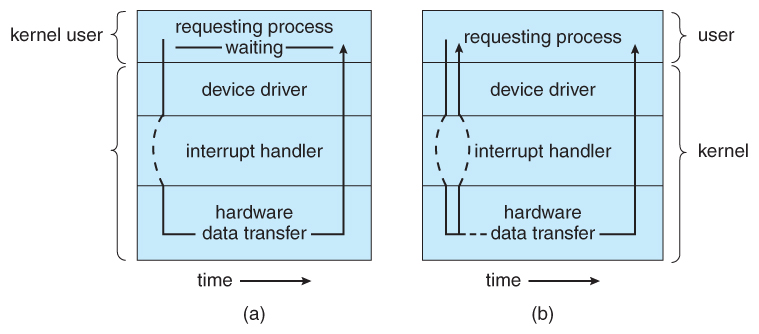


**Programmed I/O:**

* In the programmed I/O when we write the input then the device should be ready to take the data otherwise the program should wait for some time so that the device or buffer will be free then it can take the input.
* Once the input is taken then it will be checked whether the output device or output buffer is free then it will be printed. This process is continued every time in transferring of the data.

**I/O Interrupts:**

* To initiate any I / O operation, the CPU first loads the registers to the device controller. Then the device controller checks the contents of the registers to determine what operation to perform.
* There are two possibilities if I / O operations want to be executed. These are as follows −
* **Synchronous I / O** − The control is returned to the user process after the I/O process is completed. (Or**) In this scheme CPU execution waits while I/O proceeds.**
* **Asynchronous I/O** − The control is returned to the user process without waiting for the I/O process to finish. Here, I/O process and the user process run simultaneously. (or) **I/O proceeds concurrently with CPU execution.**



**Fig: Two I/O methods: (a) synchronous and (b) asynchronous.**

**DMA Structure:**

* **Direct Memory Access (DMA) is a method of handling I / O. Here the device controller directly communicates with memory without CPU involvement.**
* After setting the resources of I/O devices like buffers, pointers, and counters, the device controller transfers blocks of data directly to storage without CPU intervention.
* **DMA is generally used for high speed I / O devices.**

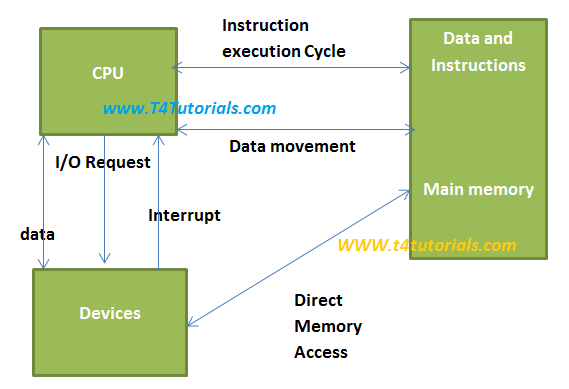


Fig:DMA

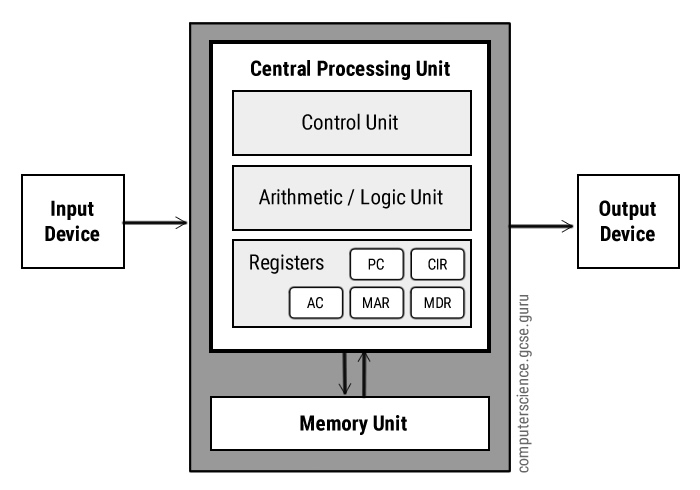
* **This is a method that allows an input/output (I/O) device to send or receive data directly to or from the main memory, bypassing the CPU to speed up memory operations.**
* **The process is managed by a chip known as a DMA controller (DMAC).**

1. **Storage Structure**



Fig: Storage-Device Hierarchy

* In generally the Storage systems organized hierarchy isSpeed, Costand Volatility.
* The wide variety of storage systems can be organized in a hierarchy according to speed and cost. The higher levels are expensive, but they are fast. As we move down the hierarchy, the cost per bit generally decreases,whereas the access time generally increases.
* **Various levels of memory are:**
* **1. Registers :**

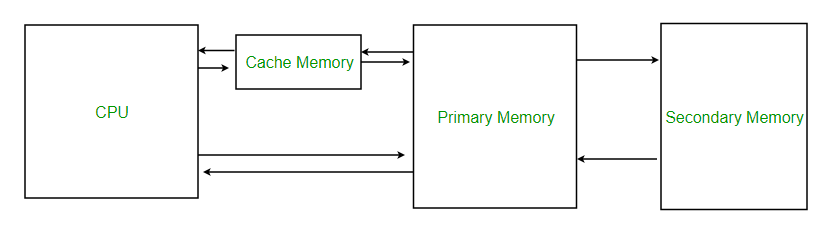


**Fig: Registers in CPU**

* **Register memory is the smallest and fastest memory in a computer. It is not a part of the main memory and is located in the CPU in the form of registers, which are the smallest data holding elements. A register temporarily holds frequently used data, instructions, and memory address that are to be used by CPU.**
* Registers are a type of computer memory used to quickly accept, store, and transfer data and instructions that are being used immediately by the CPU. The registers used by the CPU are often termed as Processor registers.
* A processor register may hold an instruction, a storage address, or any data (such as bit sequence or individual characters).
* The computer needs processor registers for manipulating data and a register for holding a memory address. The register holding the memory location is used to calculate the address of the next instruction after the execution of the current instruction is completed.
* Following is the list of some of the most common registers used in a basic computer:

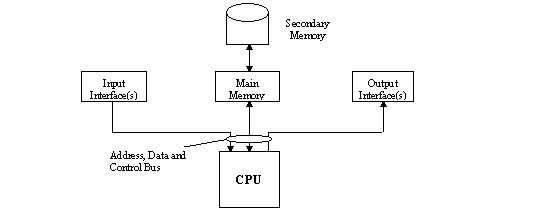
|  |  |  |  |
| --- | --- | --- | --- |
| **Register** | **Symbol** | **Number of bits** | **Function** |
| Data register | DR | 16 | Holds memory operand |
| Address register | AR | 12 | Holds address for the memory |
| Accumulator | AC | 16 | Processor register |
| Instruction register | IR | 16 | Holds instruction code |
| Program counter | PC | 12 | Holds address of the instruction |
| Temporary register | TR | 16 | Holds temporary data |
| Input register | INPR | 8 | Carries input character |
| Output register | OUTR | 8 | Carries output character |

* **2. Cache memory**:
* **It is a small-sized type of volatile computer memory that provides high-speed data access to a processor and stores frequently used computer programs, applications and data.**



**Fig: Cache Memory**

* **3. Main memory:**



**Fig : Main Memory**

* **Only large storage media that the CPU can access directly.**
* **This is Random access memory.**
* **This is also called as volatile memory.**
* **4. Secondary storage** :
* **Extension of main memory and that provides large nonvolatile storage capacity.**
* **5. Solid-state disks**:
* **It is Nonvolatile.**
* **Faster than hard disks.**
* It is used in various technologies and Becoming more popular.
* **6. Hard disks** :
* **Hard disks for personal computers can store terabytes of information.**
* Rigid metal or glass platters covered with magnetic recording material.
* Disk surface is logically divided into tracks, which are subdivided into sectors.
* The disk controller determines the logical interaction between the device and the computer.
* A [storage](https://www.webopedia.com/TERM/M/mass_storage.html) [medium](https://www.webopedia.com/TERM/M/media.html) from which [data](https://www.webopedia.com/TERM/D/data.html) is [read](https://www.webopedia.com/TERM/R/read.html) and to which it is written by lasers.
* **7. Optical disks**:
* Can [store](https://www.webopedia.com/TERM/S/store.html) much more data -- up to 6 [gigabytes](https://www.webopedia.com/TERM/G/gigabyte.html) (6 billion [bytes](https://www.webopedia.com/TERM/B/byte.html)) -- than most portable magnetic media, such as floppies.
* There are three basic types of optical disks:

[**CD-ROM**](https://www.webopedia.com/TERM/C/CD_ROM.html)**:**Like audio CDs, CD-ROMs come with data already encoded onto them. The data is permanent and can be read any number of times, but CD-ROMs cannot be modified.

[**WORM**](https://www.webopedia.com/TERM/W/worm.html)**:** Stands for write-once, [read](https://www.webopedia.com/TERM/R/read.html)-many. With a WORM [disk drive](https://www.webopedia.com/TERM/D/disk_drive.html), you can [write](https://www.webopedia.com/TERM/W/write.html)data onto a WORM disk, but only once. After that, the WORM disk behaves just like a CD-ROM.

**Erasable:** Optical disks that can be erased and loaded with new data, just like magnetic disks. These are often referred to as [EO](https://www.webopedia.com/TERM/E/erasable_optical_disk.html) (erasable optical) disks.

* These three technologies are not [compatible](https://www.webopedia.com/TERM/C/compatible.html) with one another; each requires a different type of disk drive and [disk](https://www.webopedia.com/TERM/D/disk.html). Evenwithin one category, there are many competing [formats](https://www.webopedia.com/TERM/F/format.html), although CD-ROMs are relatively standardized.
* **8. A magnetic disk:**
* This is a storage device that uses a magnetization process to write, rewrite and access data.
* It is covered with a magnetic coating and stores data in the form of tracks, spots and sectors.
* Hard disks, zip disks and floppy disks are common examples of magnetic disks.

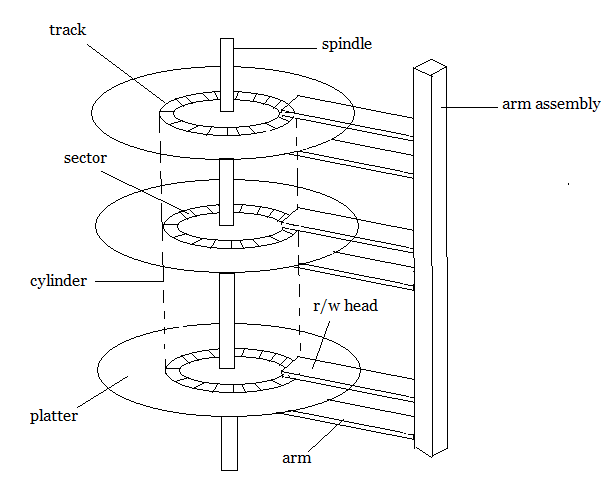


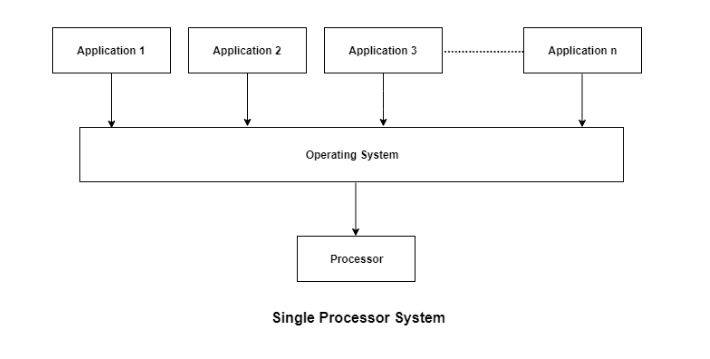
Fig:Magnetic Disk

* **9. Magnetic Tape:**
* A magnetic tape, in computer terminology, is a storage medium that allows for data archiving, collection, and backup.

**Computer-System Architecture**

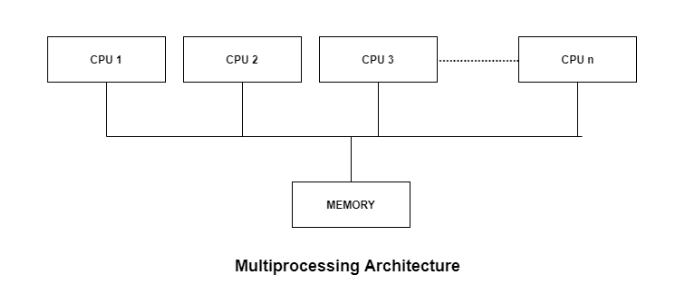
* A computer system can be organized in a number of different ways, which wecan categorize roughly according to the number of general-purpose processorsused.They are

1. **Single Processor Systems (or) Loosely Coupled Systems:**



* Most systems use single processor systems.
* **They perform only one process at a given time, and it carries out the next process in the queue only after the current process is completed.**
* OS monitors the status of them and also sends them next executable instruction.
* It relieves CPU of disk scheduling and other tasks.
* It is suitable for general purpose computers, as **it cannot run multiple processes in parallel.**

1. **Multi Processor Systems (or) Tightly Coupled Systems:**



* Also known as parallel or tightly coupled systems.
* Most computer systems are single processor systems i.e they only have one processor. However, multiprocessor or parallel systems are increasing in importance nowadays.
* **These systems have multiple processors working in parallel that share the computer clock, memory, bus, peripheral devices etc.**

**The Advantages of Multi processor systems are** :

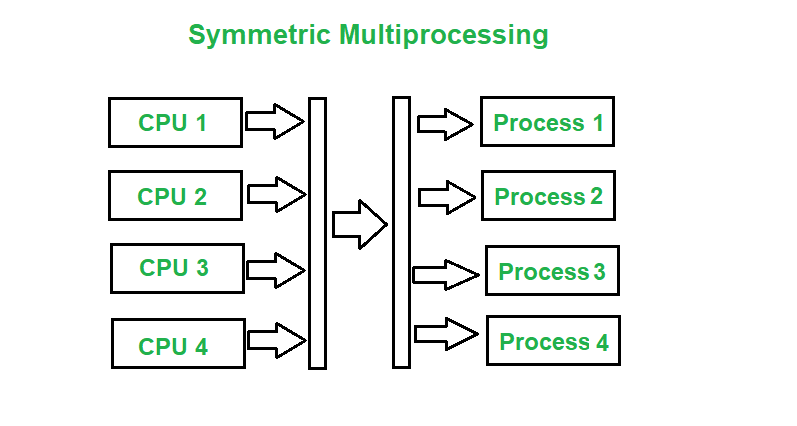
**Increased throughput:**As there are a number of processors, more work can be done in less time. These multiple processors run parallel to each other increasing the performance of the system.

**Reliability and failure-free:**Failure of any processor will not affect the functionality of the system, as there are a number of processors. We can expect failure free service from multi-processor system.

**Economy of scale:**Multi-Processor Systems cost less than a number of individual single processor system. In the case of multi-processor system expenditure for system cabinet, memory power supply, accessories are saved as these systems share resources like power supply, memory and also space.

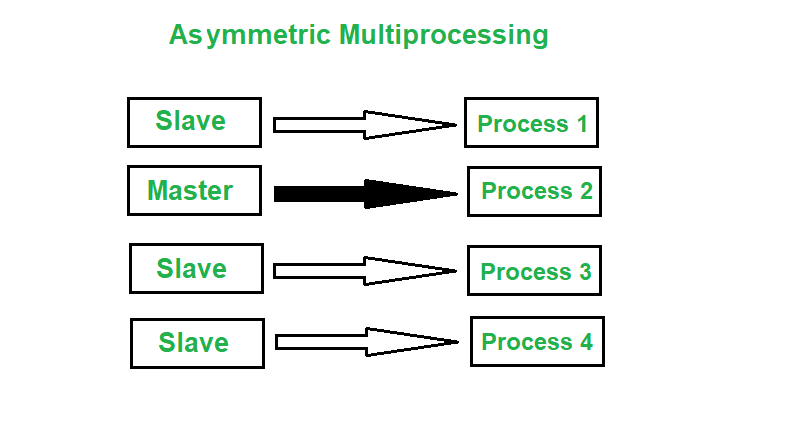
* There are mainly two types of multiprocessors i.e. symmetric and asymmetric multiprocessors. Details about them are as follows:

**(c) SymmetricMultiprocessors:**



* **In these types of systems, each processor contains a similar copy of the operating system and they all communicate with each other. All the processors are in a peer to peer relationship i.e. no master - slave relationship exists between them.**
* An example of the symmetric multiprocessing system is the Encore version of Unix for the Multimax Computer.

**(d)Asymmetric Multiprocessors:**



* **In asymmetric systems, each processor is given a predefined task. There is a master processor that gives instruction to all the other processors.**
* Asymmetric multiprocessor system contains a master slave relationship.
* Asymmetric multiprocessor was the only type of multiprocessor available before symmetric multiprocessors were created. Now also, this is the cheaper option.
* **Difference Between Asymmetric and Symmetric Multiprocessing:**

| ASYMMETRIC MULTIPROCESSING | SYMMETRIC MULTIPROCESSING |
| --- | --- |
| In asymmetric multiprocessing, the processors are not treated equally. | In symmetric multiprocessing, all the processors are treated equally. |
| Tasks of the operating system are done by master processor. | Tasks of the operating system are done individual processor |
| No Communication between Processors as they are controlled by the master processor. | All processors communicate with another processor by a shared memory. |
| In asymmetric multiprocessing, process are master-slave. | In symmetric multiprocessing, the process is taken from the ready queue. |
| Asymmetric multiprocessing systems are cheaper. | Symmetric multiprocessing systems are costlier. |
| Asymmetric multiprocessing systems are easier to design | Symmetric multiprocessing systems are complex to design |

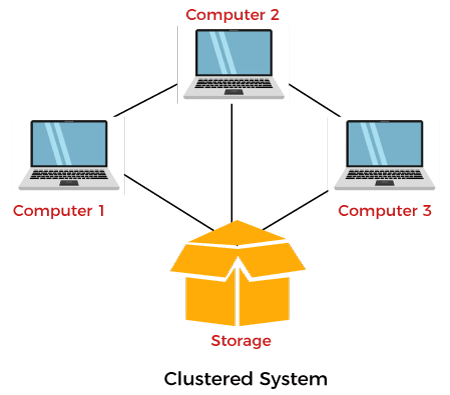
**(e) A dual-core design:**



Fig : A dual-core design with two cores placed on the same chip.

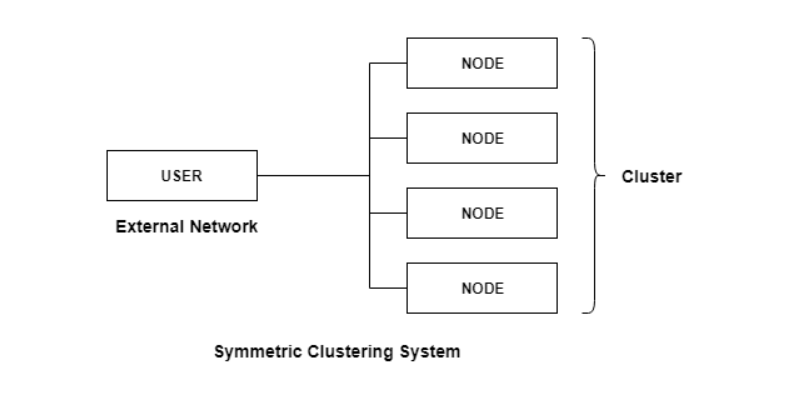
* A dual-core processor is a [CPU](https://techterms.com/definition/cpu) with two processors or "execution cores" in the same [integrated circuit](https://techterms.com/definition/integratedcircuit). Each processor has its own [cache](https://techterms.com/definition/cache) and controller, which enables it to function as efficiently as a single processor. However, because the two processors are linked together, they can perform operations up to twice as fast as a single processor can.
* We show a dual-core design with two cores on the samechip. In this design, each core has its own register set as well as its own localcache.
* Other designs might use a shared cache or a combination of local andshared caches.
* Aside from architectural considerations, such as cache,memory,and bus contention, these multicore CPUs appear to the operating system asN standard processors.
* This characteristic puts pressure on operating systemdesigners—and application programmers—to make use of those processingcores.
* **Note:**A core, or CPU core, is the "brain" of a [CPU](https://www.computerhope.com/jargon/c/cpu.htm). It receives [instructions](https://www.computerhope.com/jargon/c/compinst.htm), and performs calculations, or operations, to satisfy those instructions. A CPU can have multiple cores.
* **Note:** A CPU with a single core is called a uniprocessor. When a system has more than one core, it is called a multicore. A CPU with two cores is called a dual-core processor while a processor with four cores is called a quad-core processor. Moreover, high-performance computers can have six to eight cores.
* **Note**:**The main difference between multicore and multiprocessor is that the multicore refers to a single CPU with multiple execution units while the multiprocessor refers to a system that has two or more CPUs.**

**(f)Clustered Systems:**



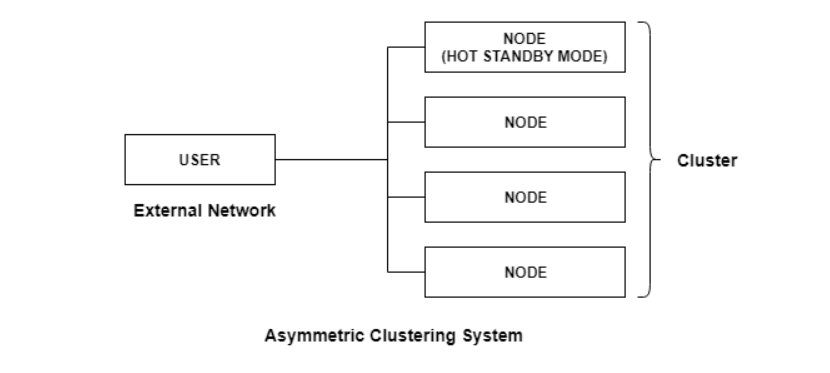
**Fig :** General structure of a clustered system.

* **Clustered computers share storage and are closely linked via a local-area network LANor a faster interconnect**, such as InfiniBand.
* Cluster operating systems are a combination of **software** and **hardware** clusters. Hardware clusters aid in the sharing of high-performance disks among all computer systems, while software clusters give a better environment for all systems to operate.
* Clustering can be structured asymmetrically or symmetrically.
* **In symmetric clustering** :
* **Multiple nodes help run all applications in this system, and it monitors all nodes simultaneously.**
* A diagram that demonstrates symmetric clustering system is:



**In asymmetricclustering**:

* **In the asymmetric cluster system, one node out of all nodes is in hot standby mode, while the remaining nodes run the essential applications.**
* A diagram that demonstrates symmetric clustering system is:



**How Asymmetric Clustering Works:**The following steps demonstrate the working of the asymmetric clustering system:

* There is a master node in asymmetric clustering that directs all the slaves’ nodes to perform the tasks required. The requests are delegated by the master node.
* A distributed cache is used in asymmetric clustering to improve the performance of the system.
* Resources such as memory, peripheral devices etc. are divided between the nodes of the asymmetric clustering system at boot time.
* **The advantages of Clustred Systems are :**
* Clustering is usually used to provide high-availability service—that is,service will continue even if one or more systems in the cluster fail.
* Clusters can also be used to provide high-performance computingenvironments.Such systems can supply significantly greater computationalpower than single-processor or even SMP systems because they can run anapplication concurrently on all computers in the cluster.