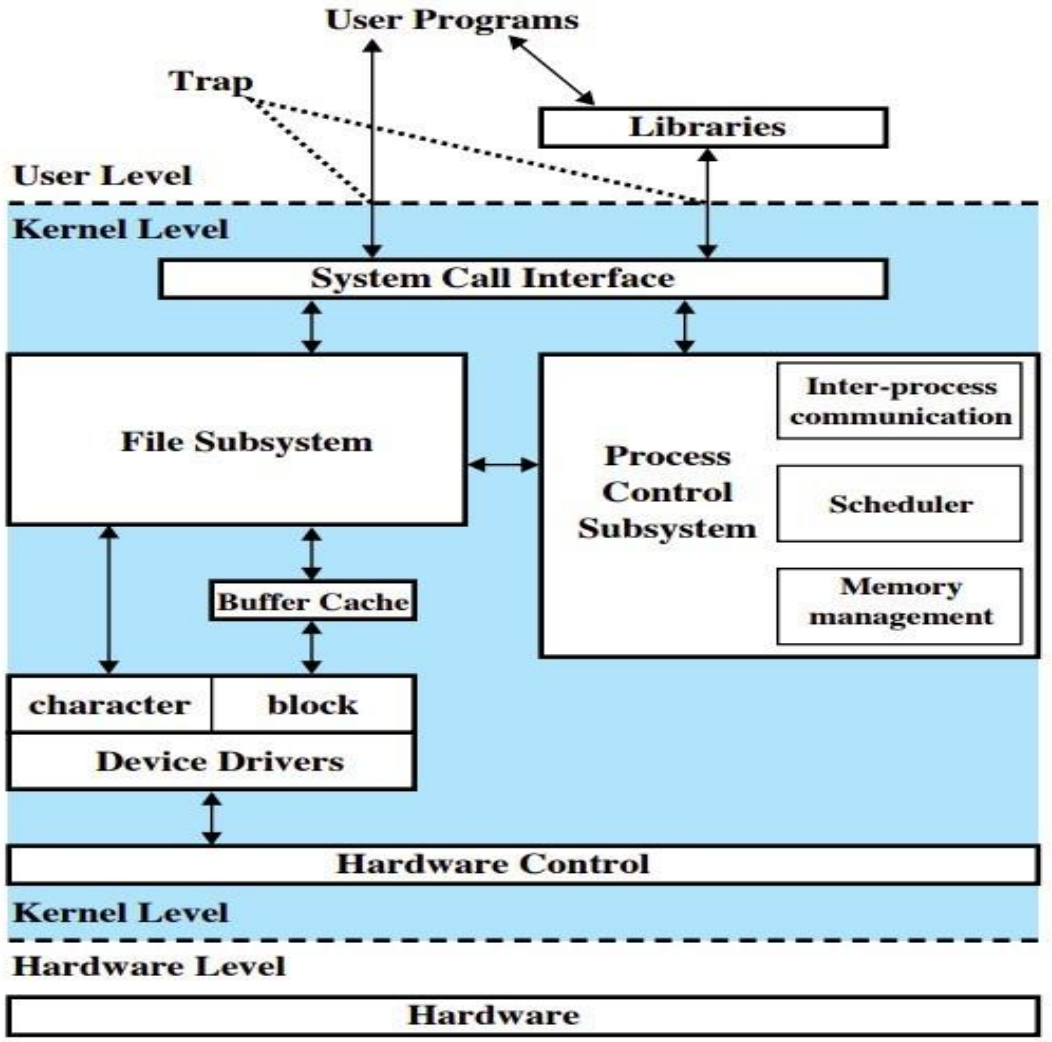
Computer Fundamentals And Operating System Concepts

➢ UNIX Operating System:

- UNIX: UNICS – **Uniplexed Information & Computing Services/System.**
- UNIX was developed at **AT&T Bell Labs** in US, in the decade of 1970's by **Ken Thompson, Denies Ritchie** and team.
- It was first run on a machine **DEC-PDP-7** (Digital Equipment Corporation
- Programmable Data Processing-7).
- UNIX is the first **multi-user, multi-programming** & **multi-tasking**
- operating system.
- UNIX was specially designed for developers by developers
- System architecture design of UNIX is followed by all modern OS's like Windows, Linux, MAC OS X, Android etc..., and hence UNIX is referred as **mother of all modern operating systems.**

Operating System Concepts



Operating System Concepts

- Kernel acts as an interface between programs and hardware.
- Operating System has subsystems like **System Call Interface Block, File Subsystem Block, Process Control Subsystem Block (which contains IPC, Memory Management & CPU Scheduling), Device Driver, Hardware Control/Hardware Abstraction Layer.**
- **There are two major subsystems:**

1. Process Control Subsystem
2. File Subsystem

- - In UNIX, whatever is that can be stored is considered as a file and whatever is active is reffered as a process.
- **- File has space & Process has life.**

Operating System Concepts

- From UNIX point of view all devices are considered as a file
- In UNIX, devices are catagorises into two catagories:

1. Character Devices: Devices from which data gets transfered character by character --> character special device file

e.g. keyboard, mouse, printer, monitor etc...

2. Block Devices: Devices from which data gets transfered block by block --> block special device file

e.g. all storage devices.

- Device Driver: It is a program/set of programs enable one or more hardware devices to communicate with the computer's operating system.

- Hardware Control Layer/Block does communication with control logic block i.e. controller of a hardware.

- **System Calls:** are the functions defined in a C, C++ & Assembly languages, which provides interface of services made available by the kernel for the user (programmer user).

-If programmers want to use kernel services in their programs, it can be called directly through system calls or indirectly through set of libary functions provided by that programming language.

- There are 6 catagories of system calls:

1. **Process Control System Calls:** e.g. fork(), _exit(), wait() etc...
2. **File Operations System Calls:** e.g. open(), read(), write(), close() etc...
3. **Device Control System Calls:** e.g. open(), read(), write(), ioctl() etc...

4.Accounting Information System Calls: e.g. getpid(), getppid(), stat() etc...

5.Protection & Security System Calls: e.g. chmod(), chown() etc...

6.Inter Process Communication System Calls: e.g. pipe(), signal(), msgget() etc...

- In UNIX 64 system calls are there.

- In Linux more than 300 system calls are there

- In Windows more than 3000 system calls are there

-When system call gets called the CPU switched from user defined code to system defined code, and hence system calls are also called as **software interrupts/trap.**

➢ Dual Mode Operation:

- System runs in two modes:

1. System Mode
2. User Mode

1. System Mode:

-When the CPU executes system defined code instructions, system runs in a system mode.

-System mode is also reffered as kernel mode/monitor mode/supervisor mode/priviledged mode.

2. User Mode:

- When the CPU executes user defined code instructions, system runs in a user mode.
- User mode is also reffered as non-priviledged mode.
- Throughout execution, the CPU keeps switch between kernel mode and user mode

➢ Dual Mode Operation:

- Throughout an execution of any program, the CPU keeps switcesh in between kernel mode and user mode and hence system runs in two modes, it is reffered as **dual mode operation.**
- To differentiate between user mode and kernel mode one bit is there onto the CPU which is maintained by an OS, called as **mode bit**, by which the CPU identifies wheather currently executing instruction is of either system defined code instruction/s or user defined code instruction/s.

In Kernel mode value of **mode bit = 0**, whereas

In User mode **mode bit = 1.**

➢ Process Management

- When we say an OS does process management it means an OS is responsible for **process creation**, **to provide environment for an execution of a process**, **resource allocation, scheduling, resources management, inter process comminication, process coordination**, and **terminate the process.**

Q. What is a Program?

- **User view:** Program is a finite set of instructions written in any programming language given to the machine to do specific task.
- **System view:** Program is an **executable file** in HDD which divided logically into sections like **exe header, bss section, data section, rodata section, code section, symbol table.**

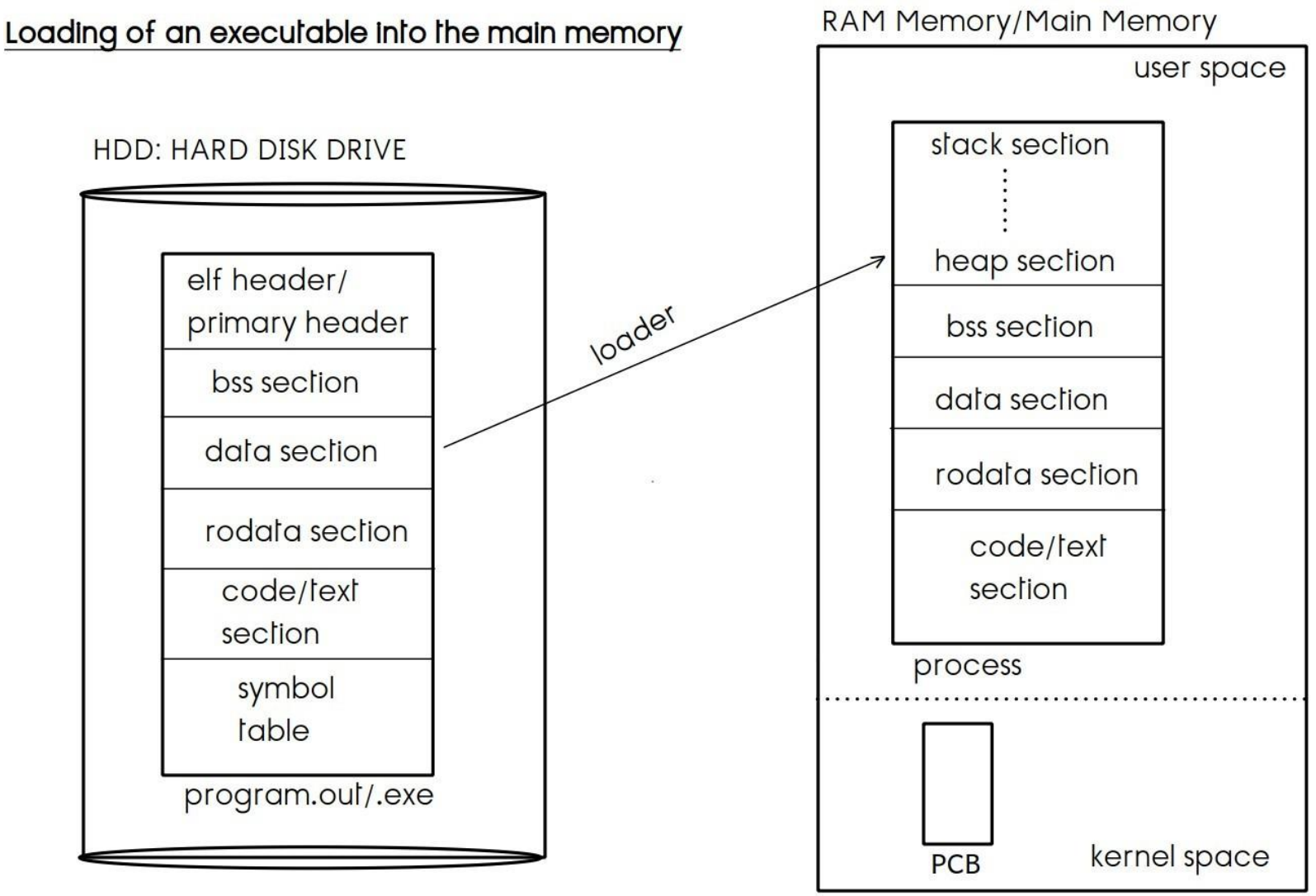
Q. What is process

User view:

- Program in execution is called as a process.
- Running program is called as a process.
- When a program gets loaded into the main memory it is reffered as a process.
- Running instance of a program is reffered as a process.

System view:

- Process is a program loaded into the main memory which has got PCB into the main memory inside kernel space and program itself into the main memory inside user space has got **bss section, rodata section, code section**, and two new sections gets added for the process:
- **stack section:** contains function activation records of called functions.
- **heap section:** dynamically allocated memory


Loading of an executable into the main memory
HDD: HARD DISK DRIVE
elf header/
primary header
bss section
data section
rodata section
code/text
section
symbol
table
program.out/.exe
loader
RAM Memory/Main Memory
user space
stack section
heap section
bss section
data section
rodata section
code/text
section
process
PCB
kernel space

Operating System Concepts

- As a kernel, core program of an a OS runs continuosly into the main memory, **part of the main memory which is occupied by the kernel reffered as kernel space and whichever part is left is reffered as an user space**, so main memory is divided logically into two parts: **kernel space & user space.**
- User programs gets loaded into the user space only.
- When we execute a program or upon submission of a process very first one structure gets created into the main memory inside kernel space by an OS in which all the information which is required to control an execution of that process can be kept, this structure is reffered as a **PCB: Process Control Block**, is also called as a **Process Descriptor.**
- Per process one PCB gets created and PCB remains inside the main memory throughout an execution of a program, upon exit PCB gets destroyed from the main memory.
- PCB mainly contains: PID, PPID, PC, CPU sched information, memory management information, information about resources allocated for that process, execution context etc...

➢ Process States:

- Throughout execution, process goes through different states out of which at a time it can be only in a one state.

States of the process:

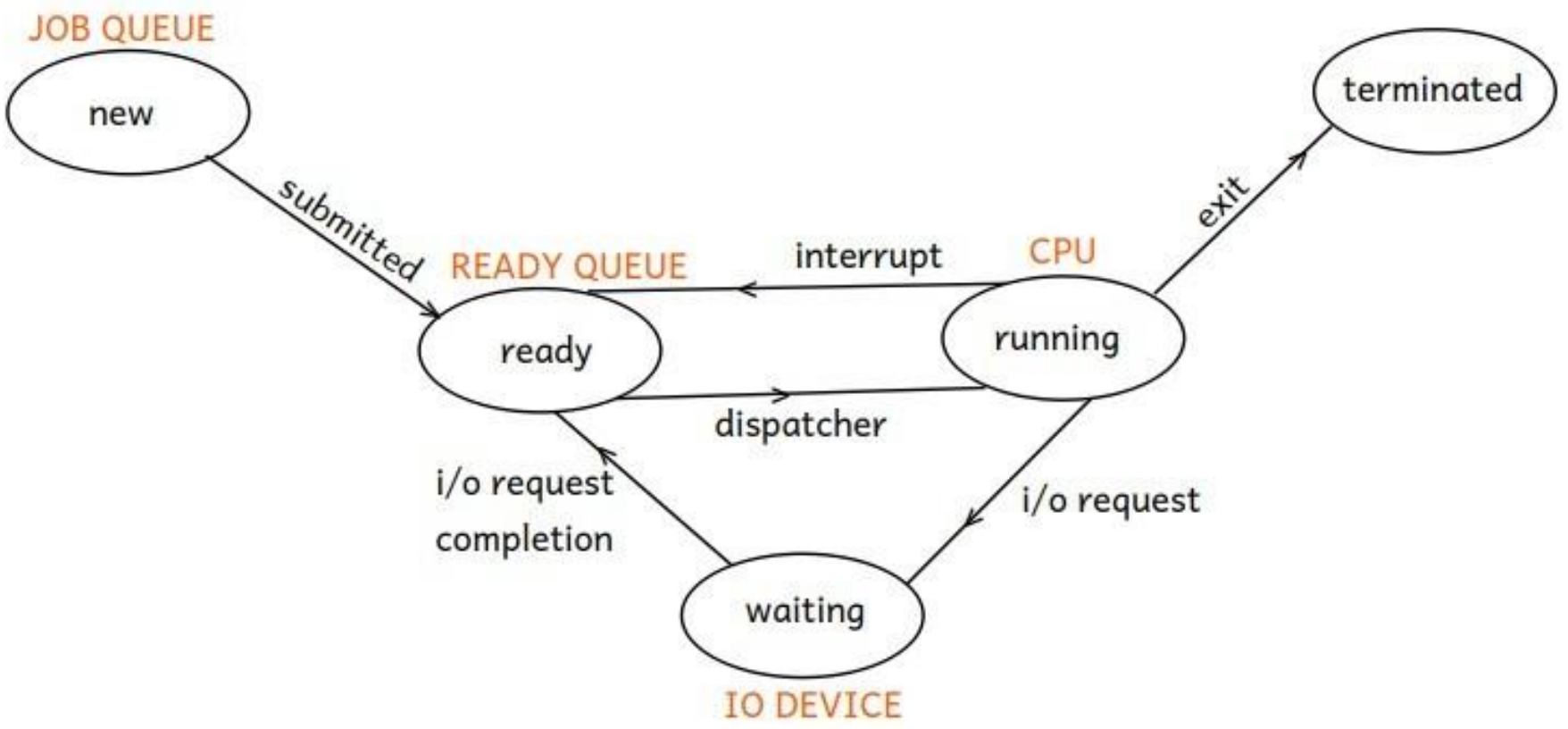
1.New state: upon submission or when a PCB for a process gets created into the main memory process is in a new state.

2.Ready state: after submission, if process is in the main memory and waiting for the CPU time, it is in a ready state.

3.Running state: if currently the CPU is executing any process then state of that process is considered as a running state.

4.Waiting state: if a process is requesting for any i/o device then state of that process is considered as a waiting state.

5.Terminated state: upon exit, process goes into terminated state and its PCB gets destroyed from the main memory.



PROCESS STATE DIAGRAM

Operating System Concepts

1. multi-programming: system in which more than one processes can be submitted/an execution of more than one programs can be started at a time.

- degree of multi-programming: no. of programs that can be submitted into the system at a time.

2. multi-tasking: system in which, the CPU can execute more than one programs simultaneously/concurrently, the speed at which it executes multiple programs simultaneosuly it seems that it executes multiple programs at a time.

At a time the CPU can execute only one program

- time-sharing: system in which CPU time gets shared among all running programs.

3.multi-threading: system in which it seems that the CPU can execute more than one threads which are of either same process or are of different processes simultaneously/concurrently.

4.multi-processor: system can run on a machine in which more than one CPU's are connected in a closed circuit.

5. multi-user: system in which multiple users can loggedin at a time.

➢ Process States:

- To keep track on all running programs, an OS maintains few **data structures** reffered as **kernel data structures:**

1. **Job queue:** it contains list of PCB's of all submitted processes.
2. **Ready queue:** it contains list of PCB's of processes which are in the main memory and waiting for the CPU time.
3. **Waiting queue:** it contains list of PCB's of processes which are requesting for that particular device.

4. **Job Scheduler/Long Term Scheduler:** it is a system program which selects/schedules jobs/processes from job queue to load them onto the ready queue.
5. **CPU Scheduler/Short Term Scheduler:** it is a system program which selects/schedules job/process from ready queue to load it onto the CPU.

- **Dispatcher:** it is a system program which loads a process onto the CPU which is scheduled by the CPU scheduler, and the time required for the dispatcher to stops an execution of one process and to starts an execution of another process is reffered as **dispatcher latency.**

➢ Context Switch:

- As during context-switch, the CPU gets switched from an execution context of one process onto an execution context of another process, and hence it is reffered as "**context-switch**".
- **context-switch = state-save + state-restore**
- **state-save** of suspended process can be done i.e. an execution context of suspended process gets saved into its PCB.
- **state-restore** of a process which is scheduled by the CPU scheduler can be done by the dispacther, dispather copies an execution context of process scheduled by the cpu scheduler from its PCB and restore it onto the CPU registers.
- When a high priority process arrived into the ready queue, low priority process gets suspended by means of sending an interrupt, and control of the CPU gets allocated to the high priority process, and its execution gets completed first, then low priority process can be resumed back, i.e. the CPU starts executing suspended process from the point at which it was suspended and onwards.

- CPU Scheduler gets called in the following four cases:

Case-1: Running -> Terminated

Case-2: Running -> Waiting

Case-3: Running -> Ready

Case-4: Waiting -> Ready

- There are two types of CPU scheduling:

1. Non-preemptive: under non-preemptive cpu scheduling, process releases the control of the CPU by its own i.e. voluntarily.

e.g. in above case 1 & case 2

2. Preemptive: under preemptive cpu scheduling, control of the CPU taken away forcefully from the process.

e.g. in above case 3 & 4.

- **Following algorithms used for CPU Scheduling:**

1. **FCFS (First Come First Served) CPU Scheduling**
2. **SJF (Shortest Job First) CPU Scheduling**
3. **Round Robin CPU Scheduling**
4. **Priority CPU Scheduling**

- Multiple algorithms are there for CPU scheduling, so there is need to decide which algorithm is best suited at specific situation and which algorithm is an efficient one, to decide this there are certain criterias called as **scheduling criterias**:

cpu utilization, throughput, waiting time, response time and turn-around-time.

➢ CPU Scheduling Criterias:

1.CPU Utilization: one need to select such an algorithm in which utilization of the CPU must be as **maximum** as a possible.

2. Throughput: total work done per unit time.

One need to select such an algorithm in which throughput must be as **maximum** possible.

3.Waiting Time: it is the toal amount of time spent by the process into the ready queue for waiting to get control of the CPU from its time of submission.

One need to select such an algorithm in which waiting time must be as **minimum** as possible.

4.Response Time: it is a time required for the process to get first response from the CPU from its time of submission.

One need to select such an algorithm in which response time must be as **minimum** as possible.

5. Turn-Around-Time: it is the total amount of time required for the process to complete its execution from its time of submission.

One need to select such an algorithm in which turn-around-time must be as **minimum** as possible.

- **Execution Time:** it is the total amount of time spent by the process onto the CPU to complete its execution.

OR CPU Burst Time: total no. of CPU cycles required for the process to complete its execution.

- **Turn-Around-Time = Waiting Time + Execution Time.**
- Turn-around-time is the sum of periods spent by the process into ready queue for waiting and onto the CPU for execution from its time of submission.

1. FCFS (First Come First Served) CPU Scheduling

- In this algorithm, process which is arrived first into the ready queue gets the control of the CPU first i.e. control of the CPU gets allocated for processes as per their order of an arrival into the ready queue.
- This algorithm is simple to impelement and can be implemented by using fifo queue.
- It is a non-preemptive scheduling algorithm.

- **Convoy effect:** in fcfs, due to an arrival of longer process before shorter processes, shorter processes has to wait for longer duration and due to which average waiting time gets increases, which results into an increase in an average turn-around-time and hence overall system performance gets down.

2. SJF(Shortest Job First) CPU Scheduling:

-In this algorithm, process which is having minimum CPU burst time gets the control of the CPU first, and whenever tie is there it can be resolved by using fcfs.

- SJF algorithm ensures minimum waiting time.

-Under non-preemptive SJF, algorithm fails if the submission time of processes are not same, and hence it can be implemented as preemptive as well.

- Non-preemptive SJF is also called as **SNTF(Shortest-Next-Time-First).**

- Preemptive SJF is also called as **SRTF(Shortest-Remaining-Time-First).**

Starvation: in this algorithm, as shorter processes has got higher priority, process which is having larger CPU burst time may gets blocked i.e. control of the CPU will never gets allocated for it, such situation is called as **starvation/indefinite blocking.**