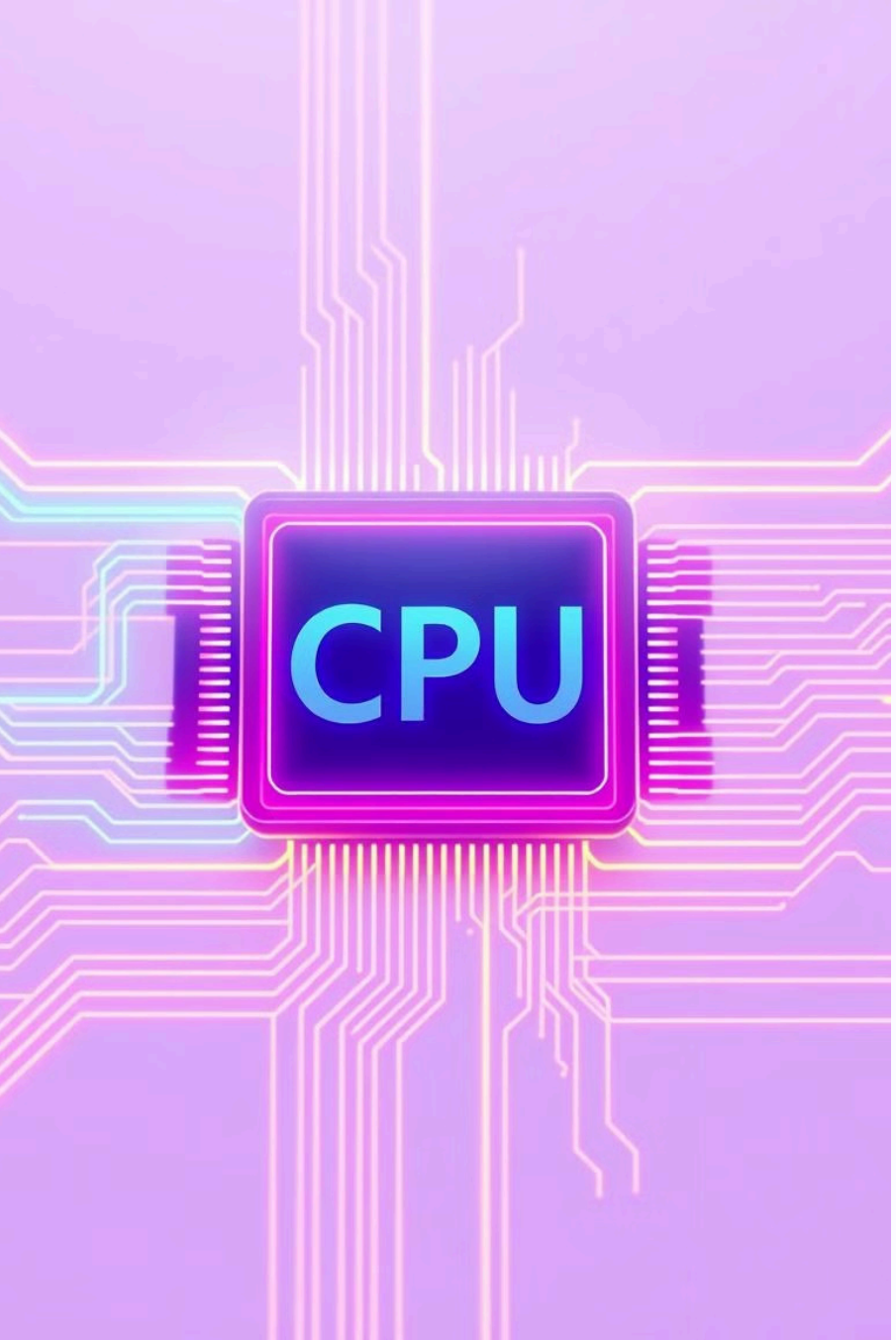


Introduction to CPU Scheduling

CPU scheduling is a fundamental concept in operating systems. It's the process of allocating and managing the use of the central processing unit (CPU) by various processes or threads.

K **by Khush karan Singh**

Definition of CPU Scheduling

CPU scheduling involves selecting which process or thread will be given access to the CPU at any given time. The scheduler decides the order in which processes will be executed and for how long each process will run.

Process Management

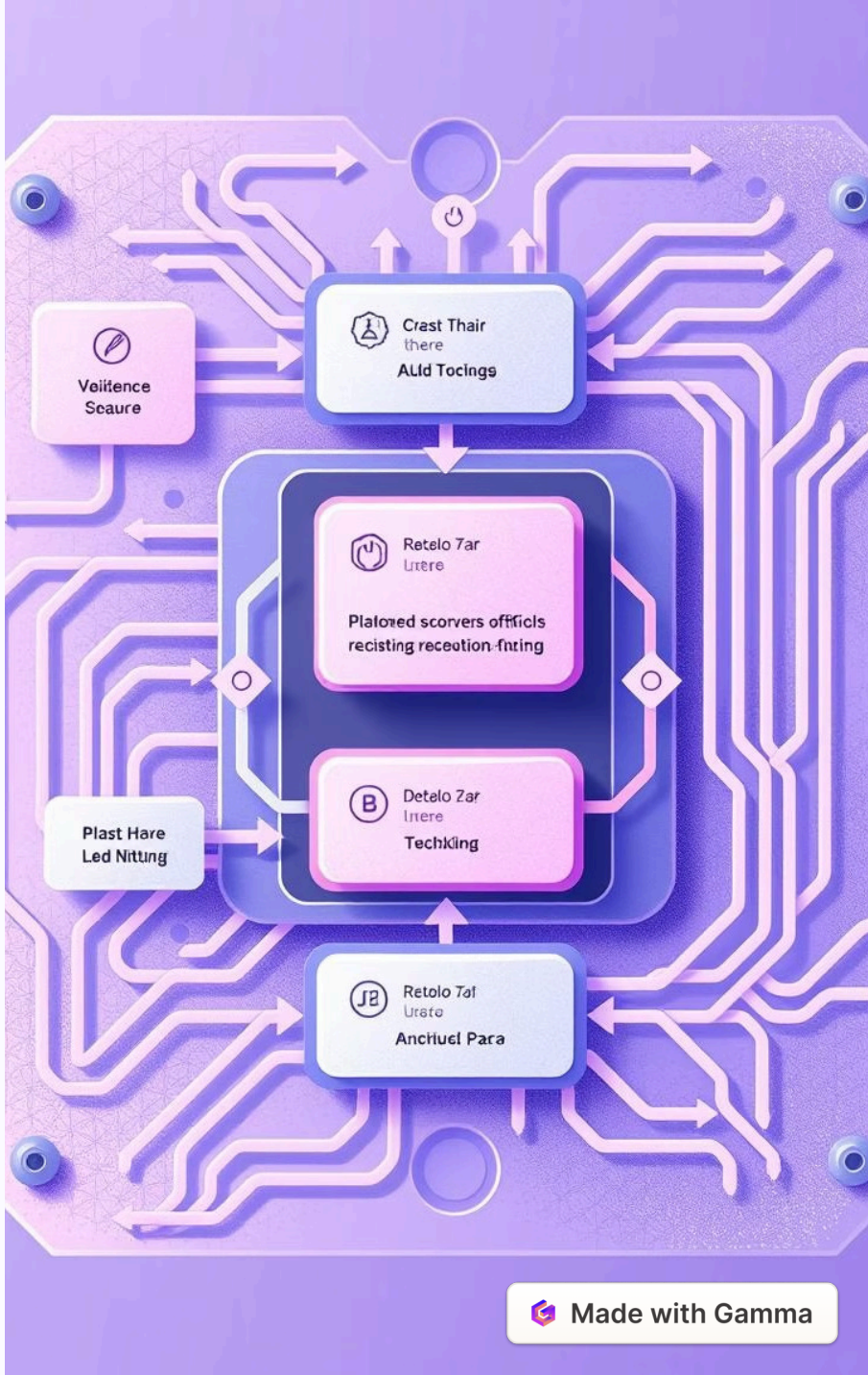
CPU scheduling is responsible for managing the execution of multiple processes in a multi-programming environment.

Resource Allocation

It determines how the CPU time is divided among competing processes, ensuring fairness and efficiency.

System Performance

CPU scheduling plays a critical role in optimizing system performance by minimizing response times, maximizing throughput, and balancing resource utilization.



Objectives of CPU Scheduling

CPU scheduling aims to achieve several important objectives, balancing the needs of various processes and the overall system.

1 Fairness

Ensuring that all processes get a fair share of the CPU time and are not starved of resources.

2 Efficiency

Maximizing the utilization of the CPU by minimizing idle time and maximizing the throughput of the system.

3 Response Time

Minimizing the time it takes for a process to respond to a user's request.

4 Throughput

Maximizing the number of processes that can be completed in a given time.

Types of CPU Scheduling Algorithms

There are various CPU scheduling algorithms, each with its own strengths and weaknesses, designed to address different system requirements and priorities.

Algorithm	Description
First-Come, First-Served (FCFS)	Processes are served in the order they arrive in the ready queue.
Shortest-Job-First (SJF)	Processes with the shortest execution times are served first.
Priority Scheduling	Processes with higher priority are served before those with lower priority.
Round-Robin (RR)	Processes are served in a cyclic manner, with each process given a fixed time slice.

First-Come, First-Served (FCFS) Scheduling

FCFS is a simple and straightforward algorithm. It follows the order of arrival, giving priority to the process that entered the ready queue first.

1

Process Arrival

Processes arrive in the system and join the ready queue.

2

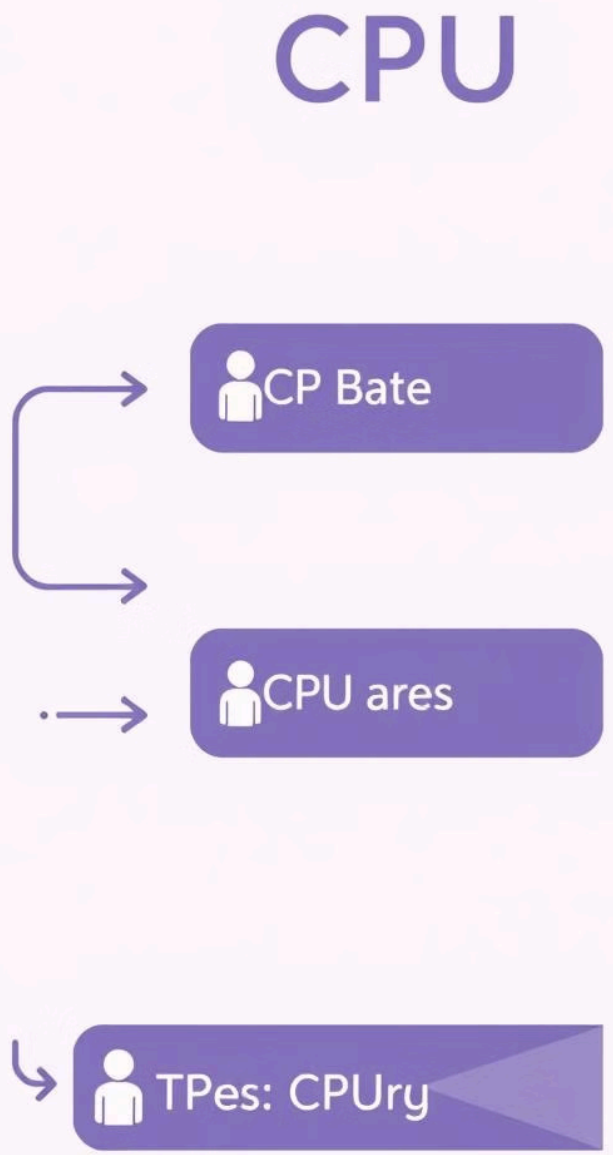
Queue Order

Processes are served in the order they appear in the queue.

3

CPU Allocation

The process at the front of the queue is allocated the CPU until it completes its execution.



Shortest-Job-First (SJF) Scheduling

SJF aims to minimize the average waiting time for processes. It prioritizes the process with the shortest estimated execution time.

Advantages

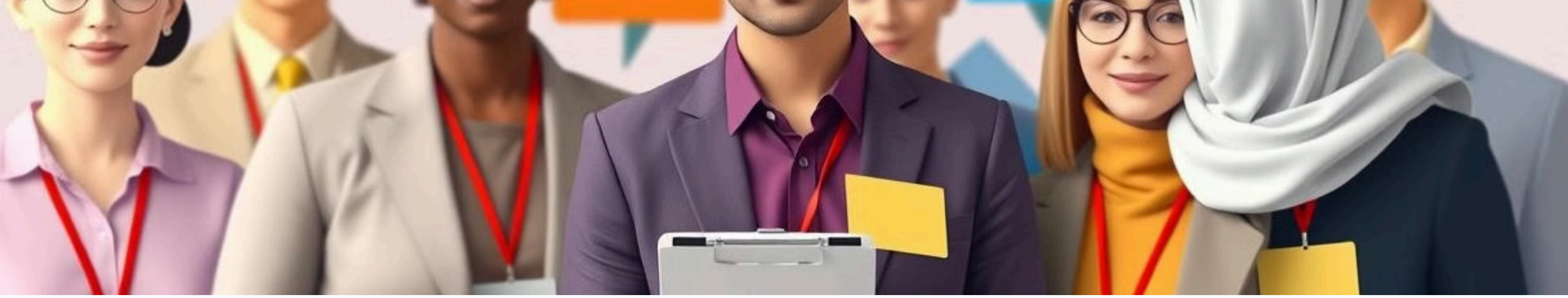
Reduces average waiting time.

Optimizes CPU utilization.

Disadvantages

Difficult to predict actual execution times.

Can lead to starvation of long processes.



Priority Scheduling

Priority scheduling assigns a priority to each process. Processes with higher priority are given preference over those with lower priority.

1

Priority Assignment

Each process is assigned a priority level.

2

Priority Queue

Processes are sorted in the ready queue based on their priorities.

3

CPU Allocation

The process with the highest priority is allocated the CPU.

Round-Robin (RR) Scheduling

RR is a time-sharing algorithm that gives each process a fixed time slice, also known as a quantum. It allows for fair sharing of CPU time among processes.

Time Quantum

Each process gets a predefined time slice to execute.

Cyclic Execution

Processes are served in a round-robin manner, cycling through the ready queue.

Context Switching

If a process doesn't complete within its time slice, it is preempted and moved to the back of the queue.