Question 2 Shorts

# QNO3: Explain the differences among multicore systems, MICs, and GPGPUs

In Multicore CPU a single processor chip contains multiple cores for speeding up and parallel processing of general-purpose tasks. Whereas Many Integrated Cores (MICs) refers to the processors with a large number of simpler cores optimized for parallel workloads and this technique is used in super computers. Lastly, GPGPU stands for "general-purpose computing on graphics processing units". It refers to the use of a graphics processing unit (GPU) for tasks that are usually handled by a central processing unit (CPU).

# QNO2: Explain the concept of performance balance.

Performance balance means making sure the CPU, memory, and I/O systems are compatible and optimized to work efficiently together. If one part, like memory, is slow, it can bottleneck the system, even if the CPU is very fast and the throughput will fall drastically.

# QNO6: Define MIPS and FLOPS

**MIPS (Million Instructions Per Second):**

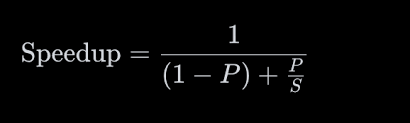
MIPS measures how many millions of instructions a CPU can execute per second. It is used to evaluate the speed of processors in handling general-purpose tasks, especially integer operations.

**FLOPS (Floating Point Operations Per Second):**

FLOPS measures how many floating-point calculations a system can perform per second. It is critical for tasks requiring high precision and mathematical computations, such as scientific simulations and machine learning.

# QNO4: Ahmdal’s Law

Amdahl's Law is a formula used to find the maximum speedup of a system when part of the task is improved. It states that even with faster parts, the total speedup is limited by the part of the task that cannot be improved. This is especially important in parallel computing, where only some parts of a program can run simultaneously. Formula:



# QNO5: Little’s Law

**Definition:** Little's Law states that the average number of items in a system (L) is equal to the arrival rate of items (λ) multiplied by the average time an item spends in the system (W). The formula is:

L = λW

It helps in analyzing and optimizing system performance, particularly in queuing systems like networks, databases, and manufacturing lines.