HPC:

- HPC is a technology which uses powerful computers or supercomputers to solve the complex computational problem at very high speed.

Need of HPC:

- Complete a time consuming problem in less time.
- Performs high no. of operations per sec.
- Compute parallel operations over lot of computation elements like CPU and GPU.

Parallel Programming:

- According to Moore's law the no. of transistors in an IC double about every two years.
- In recent era, physical limitations such as heat dissipation, power consumption and quantum effect slows down the performance of the system.
- In parallel programming, tasks are parallelized so that they can be run at same time by using multiple computers or multicore within a CPU.
- Parallel programming is critical for large scale proj. in which accuracy and speed are needed.

Need of Parallel Programming:

- With the rise of multicore processor, simply writing sequential programs no longer fully utilize the hardware capabilities of modern systems.
- Parallel programming allows us:
  - → Divide large problem into smaller sub-tasks.
  - → Execute multiple tasks simultaneously.
  - → Reduce execution time.
  - → leverage multicore, CPU & distributed architecture.

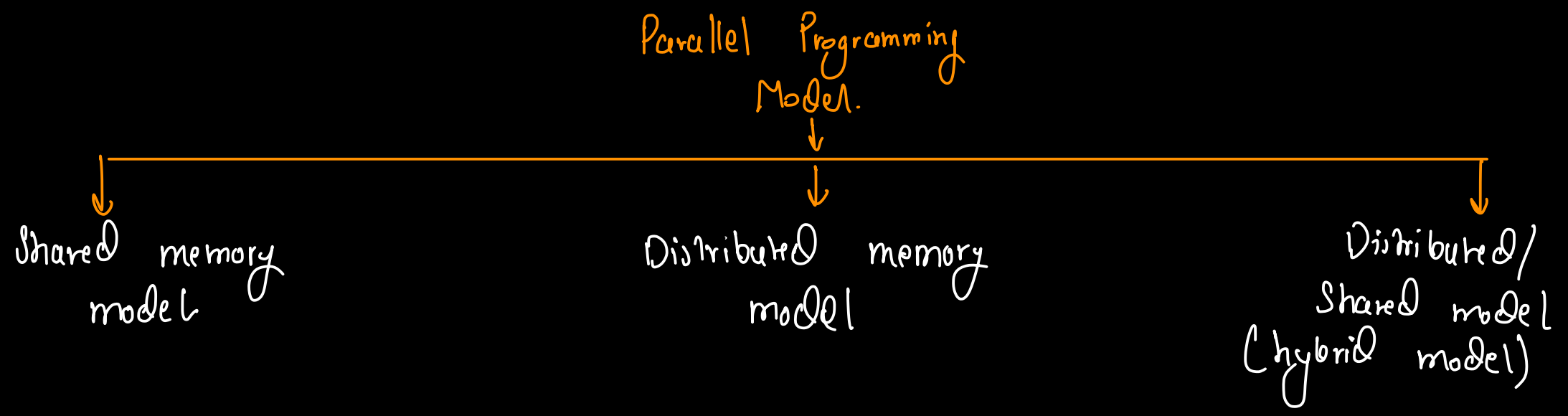
Applications:

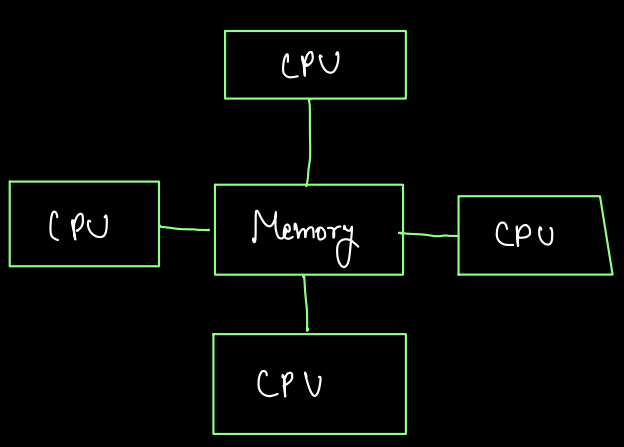
- Scientific computing & simulation
  - → weather forecasting
  - → Molecular dynamics

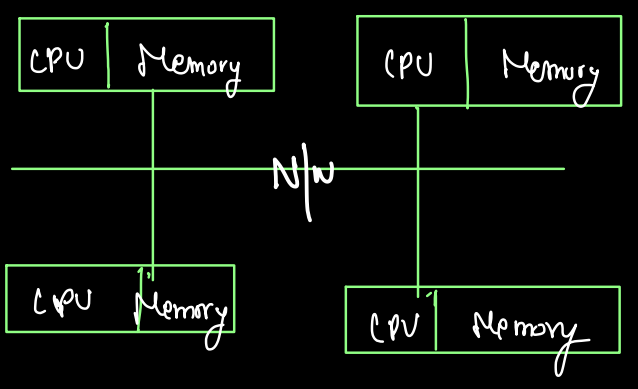
- Big Data & M.L.
  - → Data processing in Hel.
  - → Training deep learning models on GPU.
  - → Real time analysis on large dataset.
- Image and Signal Processing.
  - → Edge detection, filtering
  - → Medical image reconstruction.
- Games & Graphics.
  - → Physical Simulation
  - → 3D scenes.
- Search and optimization.
  - → Hel graph algorithm
  - → Genetic algorithm.

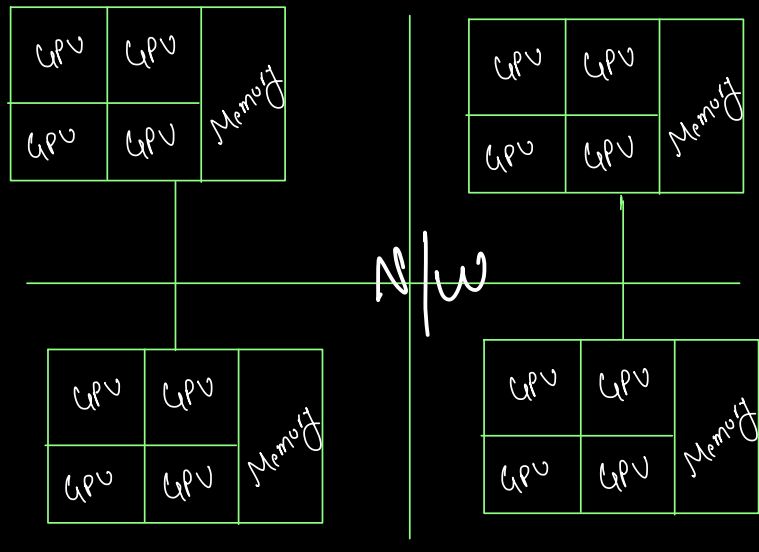
Benifits:

- Speed up
- Scalability
- Maximises resource utilization
- Responsiveness : Enable real time system to response fast and accurate.









- Pooling
- Shared network

Ilel Programming Software & its significance:

- Ilel programming software refers to tools, languages or framework that allow a program to execute multiple task or instructions simultaneously by using multiple processors, CPU core, threads or GPU.
- OpenMP (Open multi processing) → (App. prog. interface) API for shared memory.
- MPI (message passing interface): for distributed computing across multiple system.

CUDA (Compute unified Device architecture): programming model for GPU computing.

OpenCL (Open Computing languages): Framework for writing code that run on GPU & CPU both.

* Significance: Faster execution.
Efficient use of multicore processor.

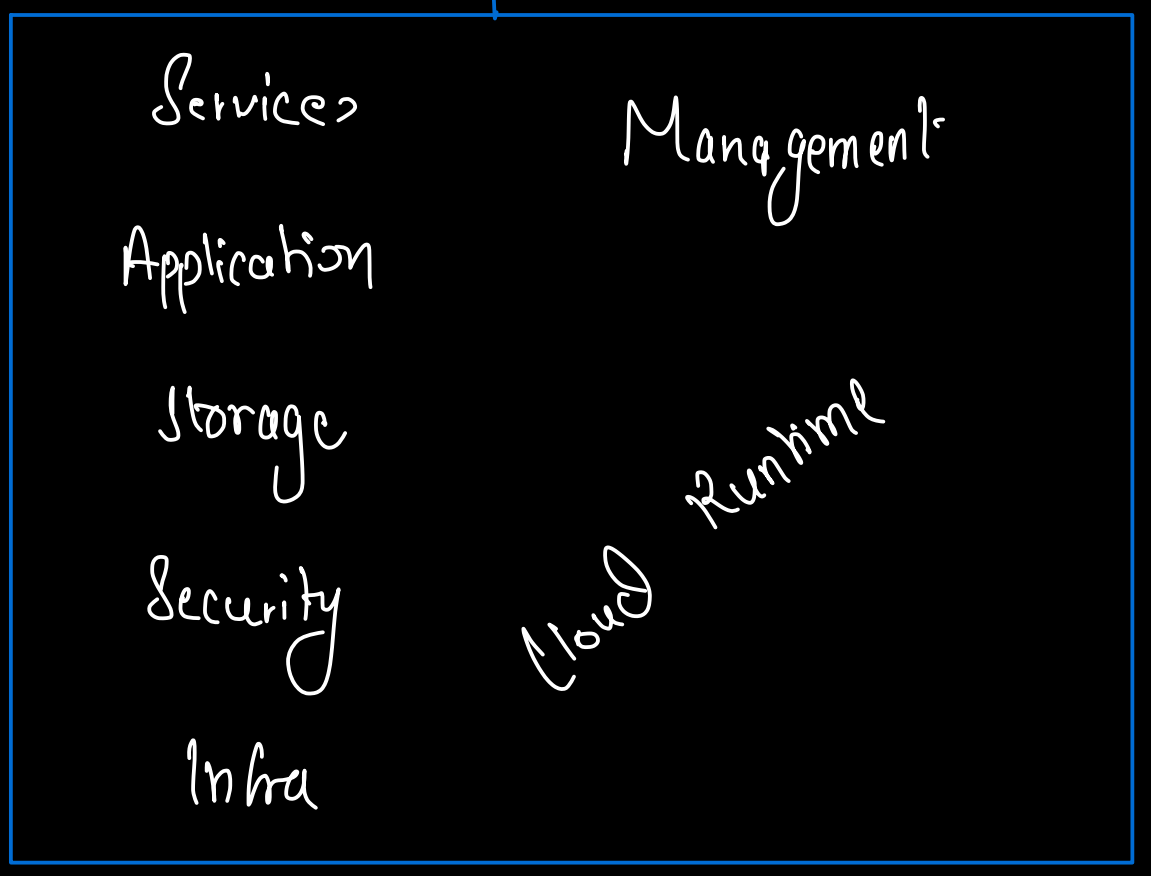
Cloud Computing:

- It means delivering computing like storage, Databases, networking, software over the network.
- Cost efficient, scalability, accessibility from anywhere & reduce hardware requirement.
- Cloud provides:
  - → AWS (amazon web services)
  - → Ms Azure
  - → Alibaba Cloud
  - → Google cloud platform
  - → IBM Cloud.
  - → Oracle cloud.

Key characteristics:

- On-demand self service
- Pay as you go pricing
- Resource pooling
- Broad Network Access
- Scalability.

Cloud Computing Architecture:



Services:

1. SaaS (Software as a Service)
   - ↳ Software over a Internet.
   - ↳ On Demand
   - ↳ end user / client
   - ↳ No need to install on PC.
   - ↳ Server / Resources are managed by vendor.
   - ↳ Platform independence.
2. PaaS (Platform as a Service):
   - ↳ 3rd party provider offers software & hardware tool needed to develop, test or run.
   - ↳ Developers use it.
   - ↳ No access of OS, Middleware, virtual machine.
   - ↳ Access to user interface is provided.

↳ No need to purchase expensive h/w or s/w.

Eg → Google App Engine.

IaaS:

- Provide access to IT tools like virtual computers, storage & network through internet.
- Operating System
- Virtual machine and storage
- IP addresses
- Provide Infra
- enhanced Scalability
- flexibility

FaaS: {Function as a Service}.

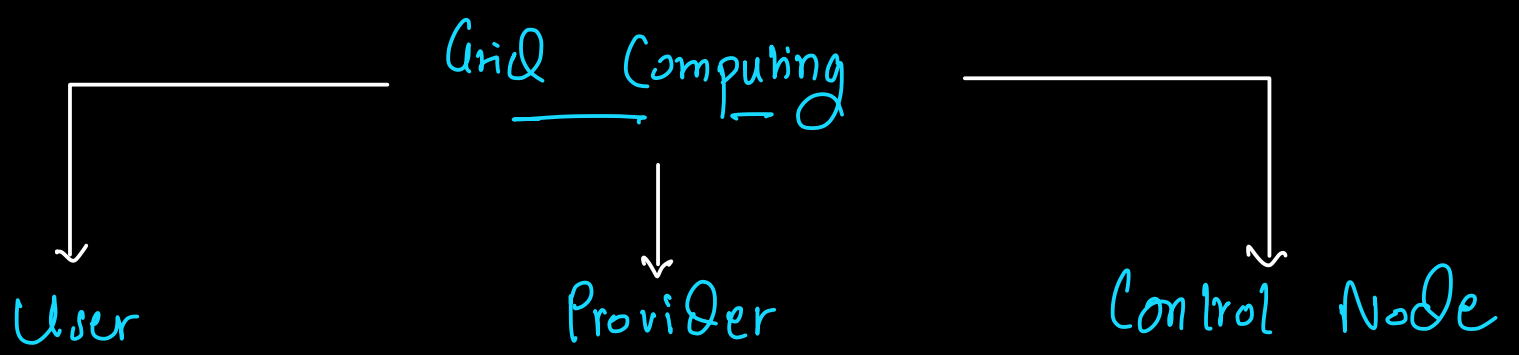
↳ Allow customers.

Grid Computing:

- It is a subset of Distributed computing.
- It is defined as a network of computers working together to perform a task that would be difficult for a machine.
- It combined computer resources from different locations to achieve a common goal.

Importance of Grid Computing:

1. Scalability
2. Resource utilization
3. Complex problem solving
4. Collaboration
5. Cost Saving.


Grid Computing
User
Provider
Control Node

Unit-3

Parallel Matrix Multiplication:

$$A = \left[\begin{array}{cc|cc} 2 & 3 & 4 & 5 \\ 9 & 8 & 7 & 6 \\ \hline 5 & 4 & 2 & 3 \\ 8 & 7 & 3 & 4 \end{array}\right]$$

P0 P1 P2 P3

$$B = \left[\begin{array}{cc|cc} 3 & 5 & 7 & 6 \\ 2 & 7 & 6 & 3 \\ \hline 7 & 5 & 3 & 2 \\ 4 & 3 & 2 & 5 \end{array}\right]$$

P0 P1 P2 P3

* Total Steps → $\sqrt{p} = \sqrt{4} = 2$

Step 1 →

$$\left[\begin{array}{cc|cc} 2 & 3 & 4 & 5 \\ 9 & 8 & 7 & 6 \\ \hline 5 & 4 & 2 & 3 \\ 6 & 7 & 3 & 4 \end{array}\right]$$

Q) A scientific simulator run on 30 nodes HPC Cluster using BLAS runtime

Each node: 80 Gflops

Workload = 24,0000 Gflops

Checkpoint = every 30 mins with 5 min overhead.

One node fails every 120 mm, causing rollback of 20 min

(i) Calculate the theoretical execution time with 100% efficiency

* Note: · GFLOPS → Giga floating - point operations per sec {unit of the computation}.

→ Time = $\frac{\text{Workload}}{\text{Throughput}}$ = $\frac{\text{Workload}}{\left(\frac{\text{performance}}{\text{per node}}\right) \times \text{no. of node}}$

= $\frac{24\,0000}{80 \times 30}$ = 100 sec

Q) A parallel BLAS routine performs 10 billion operation.

One node = 50 GFLOPS

20 nodes with efficiency of 90%

(i) Calculate ideal runtime

(ii) " actual/ adjusted runtime

(iii) Discuss the factor that causes less than perfect efficiency in BLAS based HPC workloads.

(i) 0.01 sec

(ii) $\frac{0.01}{9}$ = 0.011 sec

(i) Ideal Runtime = $\frac{\text{Workload}}{\text{no. of nodes} * \text{per}}$ = $\frac{10 \times 10^9}{50 \times 10^9 \times 20}$ = 0.01s

(ii) Actual Runtime = Ideal Runtime = $\frac{0.01}{0.9}$ = 0.011s

2

Q) Calculate the max. annual downtime allowed for:

(i) Mission Critical System (99.999% availability)

(ii) Business " " (99.9% availability)