PARALLEL AND DISTRIBUTED COMPUTING

INTRODUCTION

The simultaneous growth in the availability of big data and in the number of simultaneous users on the Internet places particular pressure on the need to carry out computing tasks "in parallel," or simultaneously. Parallel and distributed computing occur across many different topic areas in computer science, including algorithms, computer architecture, networks, operating systems, and software engineering. During the early 21st century there was explosive growth in multiprocessor design and other strategies for complex applications to run faster. Parallel and distributed computing build on fundamental systems concepts. Such as concurrency, mutual exclusion, consistency in state/memory manipulation message-passing, and shared-memory models.

PARALLEL COMPUTING:
The simultaneous execution of the same task on multiple processors in order to obtain faster results is called parallel computing.

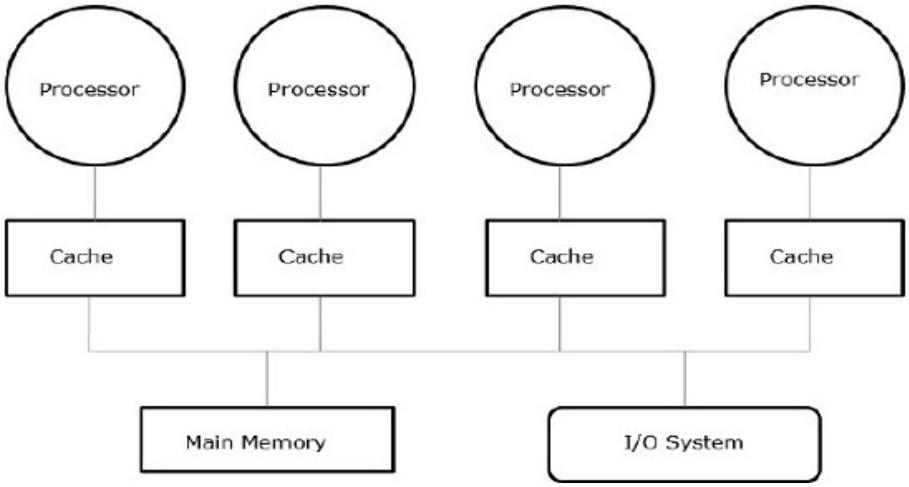
- HPC: High Performance/Productivity Computing
- Technical Computing
- Cluster computing

"In parallel computing, multiple processors perform multiple tasks assigned to them simultaneously. Memory in parallel systems can either be shared or distributed. Parallel computing provides concurrency and saves time and money."

DISTRIBUTED COMPUTING:
Uses or coordinates physically separate computing resources.

- Grid computing
- Cloud Computing

"In distributed computing, we have multiple autonomous computers which seem to the user as a single system. In distributed systems, there is no shared memory and computers communicate with each other through message passing. In distributed computing, a single task is divided among different computers."



DIFFERENCE BETWEEN PARALLEL COMPUTING AND DISTRIBUTED COMPUTING

Sr No.	PARALLEL COMPUTING	DISTRIBUTED COMPUTING
1	Many operations are performed simultaneously	System components are located at different locations
2	A single computer is required	Uses Multiple Computers
3	Multiple processors perform multiple operations	Multiple computers perform multiple operations
4	It may have shared or distributed memory	It has only distributed memory
5	Processors communicate with each other through the bus	Computers communicate with each other through message passing
6	Improves the system performance	Improves system scalability, fault tolerance, and resource-sharing capabilities

ADVANTAGES OF PARALLEL COMPUTING

The advantages of Parallel Computing over Serial Computing are as follows:

1. It saves time and money as many resources working together will reduce the time and cut potential costs.
2. It can be impractical to solve larger problems on Serial Computing.
3. It can take advantage of non-local resources when the local resources are finite.
4. Serial Computing 'wastes' the potential computing power, thus Parallel Computing makes better work of hardware.

TYPES OF PARALLELISM

1. **BIT-LEVEL PARALLELISM:**
   It is a form of parallel computing which is based on the increasing processor's size. It reduces the number of instructions that the system must execute in order to perform a task on large-sized data.
   Example: Consider a scenario where an 8-bit processor must compute the sum of two 16-bit integers. It must first sum up the 8 lower-order bits, then add the 8 higher-order bits, thus requiring two instructions to perform the operation. A 16-bit processor can perform the operation with just one instruction.

2. **INSTRUCTION-LEVEL PARALLELISM:**

   A processor can only address less than one instruction for each clock cycle phase. These instructions can be re-ordered and grouped which are later on executed concurrently without affecting the result of the program. This is called instruction-level parallelism.

3. **TASK PARALLELISM:**

   Task parallelism employs the decomposition of a task into subtasks and then allocating each of the subtasks for execution. The processors perform the execution of sub-tasks concurrently.

WHY PARALLEL COMPUTING?

- The whole real world runs in dynamic nature i.e. many things happen at a certain time but at different places concurrently. This data is extensively huge to manage.
- Real-world data needs more dynamic simulation and modeling, and for achieving the same, parallel computing is the key.
- Parallel computing provides concurrency and saves time and money.
- Complex, large datasets, and their management can be organized only and only using a parallel computing approach.
- Ensures the effective utilization of the resources. The hardware is guaranteed to be used effectively whereas in the serial computation only some part of the hardware was used and the rest was rendered idle.
- Also, it is impractical to implement real-time systems using serial computing.

APPLICATIONS OF PARALLEL COMPUTING:

- Databases and Data mining.
- Real-time simulation of systems.
- Science and Engineering.
- Advanced graphics, augmented reality, and virtual reality.

LIMITATIONS OF PARALLEL COMPUTING:

- It addresses such as communication and synchronization between multiple sub-tasks and processes which is difficult to achieve.
- The algorithms must be managed in such a way that they can be handled in a parallel mechanism.
- The algorithms or program must have low coupling and high cohesion. But it’s difficult to create such programs.
- More technically skilled and expert programmers can code a parallelism-based program well.

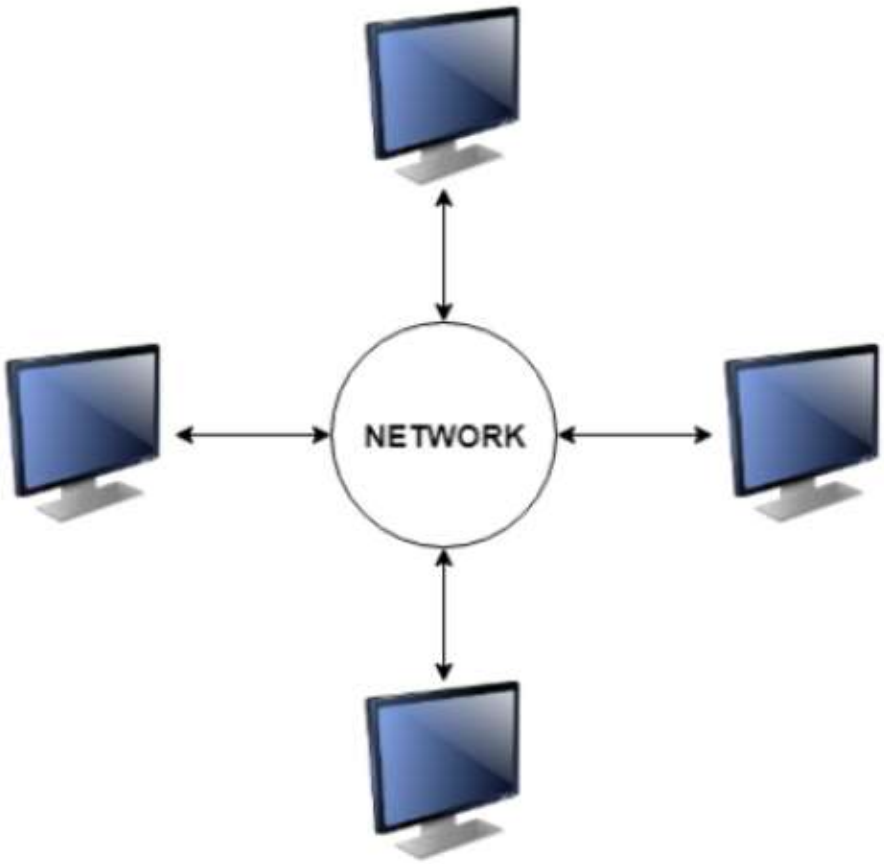
FUTURE OF PARALLEL COMPUTING:

The computational graph has undergone a great transition from serial computing to parallel computing. Tech giant such as Intel has already taken a step towards parallel computing by employing multicore processors.

Parallel computation will revolutionize the way computers work in the future, for the better good. With the world connecting to each other even more than before, Parallel Computing does a better role in helping us stay that way. With faster networks, distributed systems, and multi-processor computers, it becomes even more necessary.

WHAT IS DISTRIBUTED SYSTEM?

A distributed system contains multiple nodes that are physically separate but linked together using the network. All the nodes in this system communicate with each other and handle processes in tandem. Each of these nodes contains a small part of the distributed operating system software.
A diagram to better explain the distributed system is –



DISTRIBUTED OPERATING SYSTEM

TYPES OF DISTRIBUTED SYSTEMS

The nodes in the distributed systems can be arranged in the form of client/server systems or peer to peer systems. Details about these are as follows −

CLIENT/SERVER SYSTEMS:

In client-server systems, the client requests a resource and the server provides that resource. A server may serve multiple clients at the same time while a client is in contact with only one server. Both the client and server usually communicate via a computer network and so they are a part of distributed systems.

PEER-TO-PEER SYSTEMS:

The peer-to-peer systems contain nodes that are equal participants in data sharing. All the tasks are equally divided between all the nodes. The nodes interact with each other as required to share resources. This is done with the help of a network.

ADVANTAGES OF DISTRIBUTED SYSTEMS

Some advantages of Distributed Systems are as follows:
All the nodes in the distributed system are connected to each other. So nodes can easily share data with other nodes.
More nodes can easily be added to the distributed system i.e. it can be scaled as required.
Failure of one node does not lead to the failure of the entire distributed system. Other nodes can still communicate with each other.
Resources like printers can be shared with multiple nodes rather than being restricted to just one.

DISADVANTAGES OF DISTRIBUTED SYSTEMS

Some disadvantages of Distributed Systems are as follows:

- It is difficult to provide adequate security in distributed systems because the nodes as well as the connections need to be secured.
- Some messages and data can be lost in the network while moving from one node to another.
- The database connected to the distributed systems is quite complicated and difficult to handle as compared to a single-user system.
- Overloading may occur in the network if all the nodes of the distributed system try to send data at once.

DISTRIBUTED COMPUTING MODELS

There are certain technologies working behind the cloud computing platforms making cloud computing flexible, reliable, and usable. These technologies are listed below:

- Virtualization
- Service-Oriented Architecture (SOA)
- Grid Computing
- Utility Computing