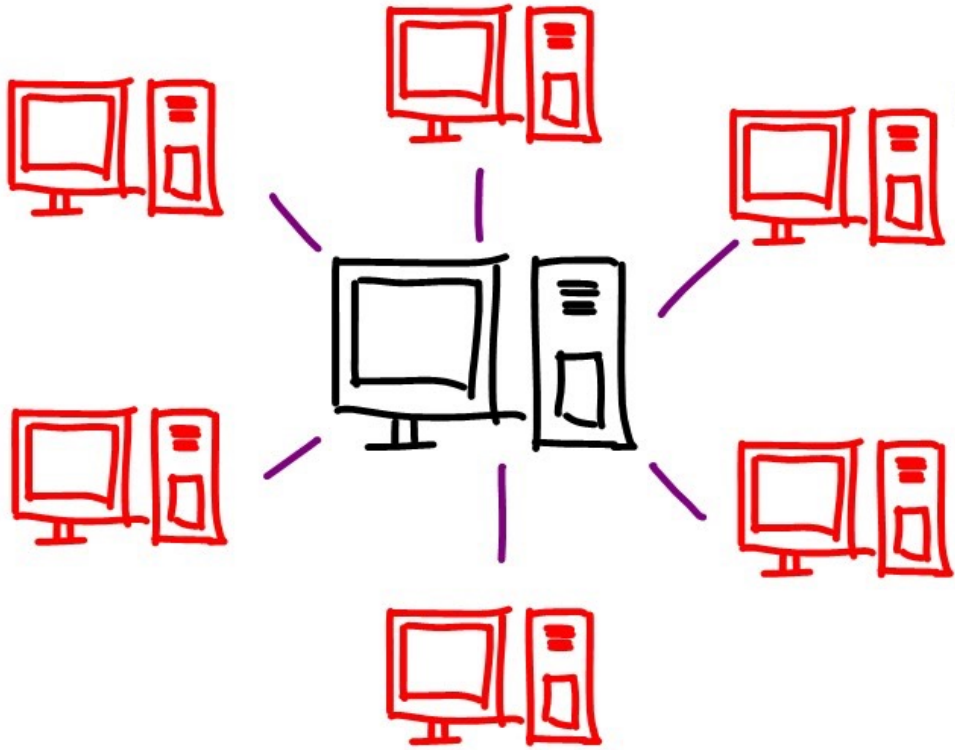
Introduction to Distributed and Parallel Computing
CS-401

Dr. Sanjay Saxena
Visiting Faculty, CSE, IIIT Vadodara
Assistant Professor, CSE, IIIT Bhubaneswar
Post doc – University of Pennsylvania, USA
PhD – Indian Institute of Technology(BHU), Varanasi

What is Distributed Systems/Computing?

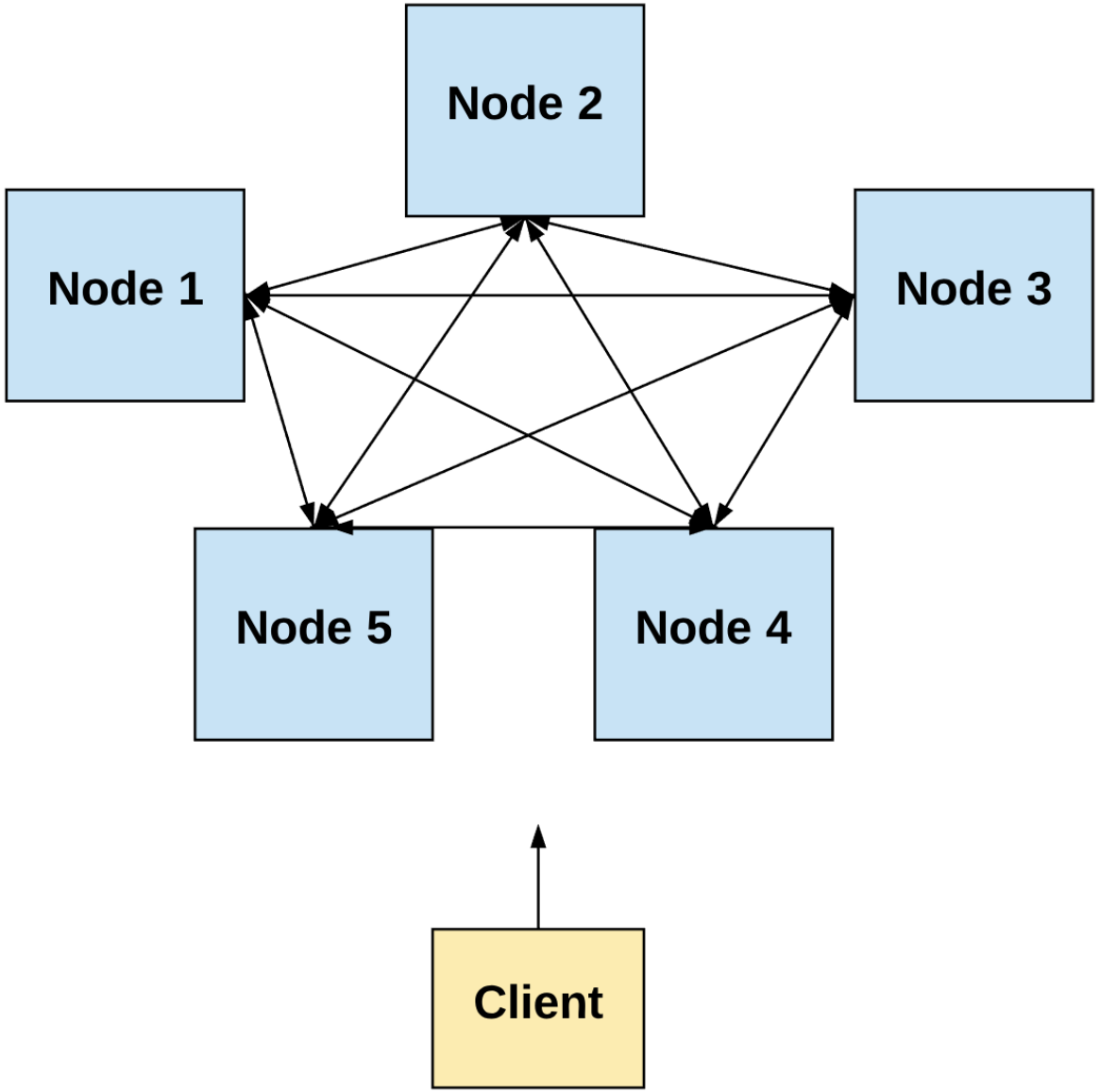
- Simply, Distributed System is a group of computers working together as to appear as a single computer to the end-user.
- Distributed computing refers to a system where processing and data storage is distributed across multiple devices or systems, rather than being handled by a single central device.

These machines have a shared state, operate concurrently and can fail independently without affecting the whole system's uptime.



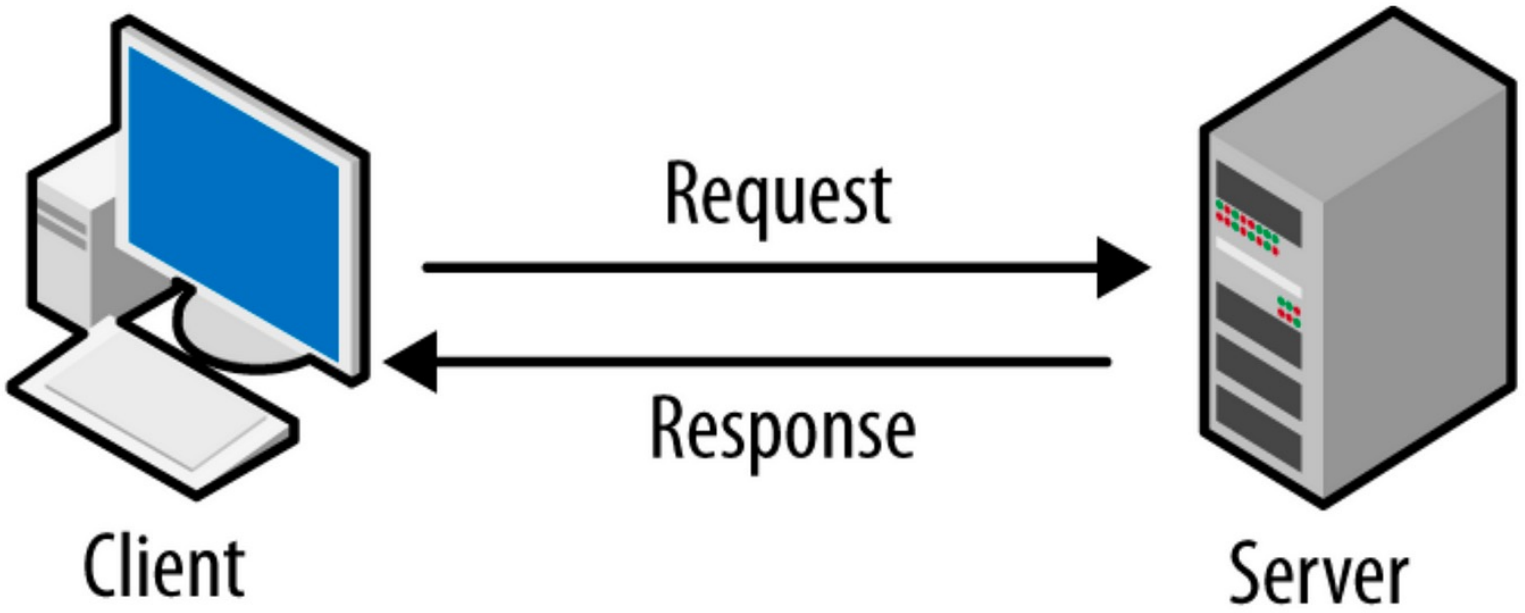
Distributed computing refers to the use of distributed systems to complete computing tasks.

Contd..



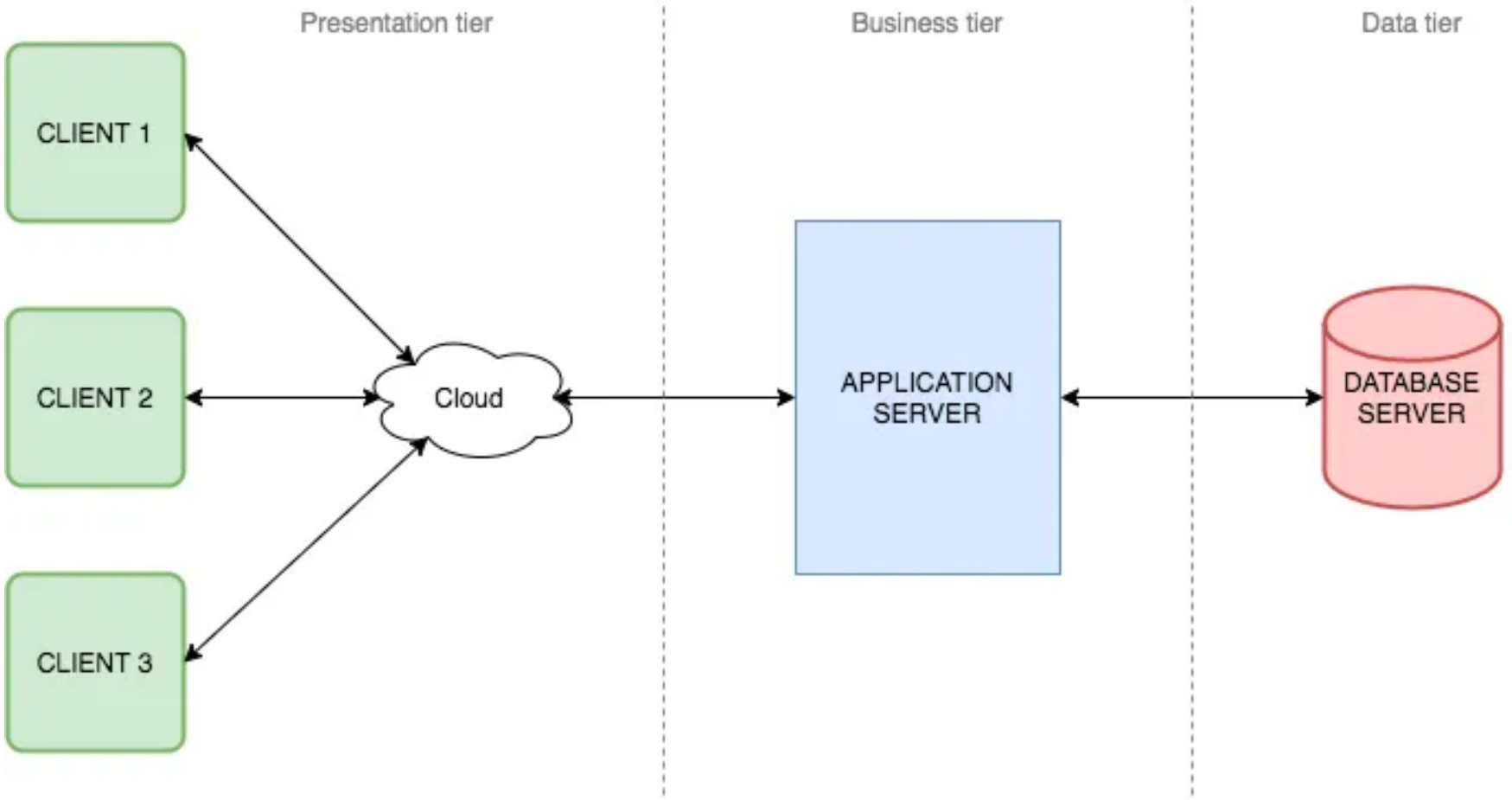
Types of distributed systems

Client-Server — In this client sends requests to the server then, the server responds to the client requests. Clients contact the server for data, then format it and display it to the end-user. The end-user can also make changes from the client-side and commit them back to the server to make it permanent. There are three main methods — sockets, remote procedure calls, pipes.



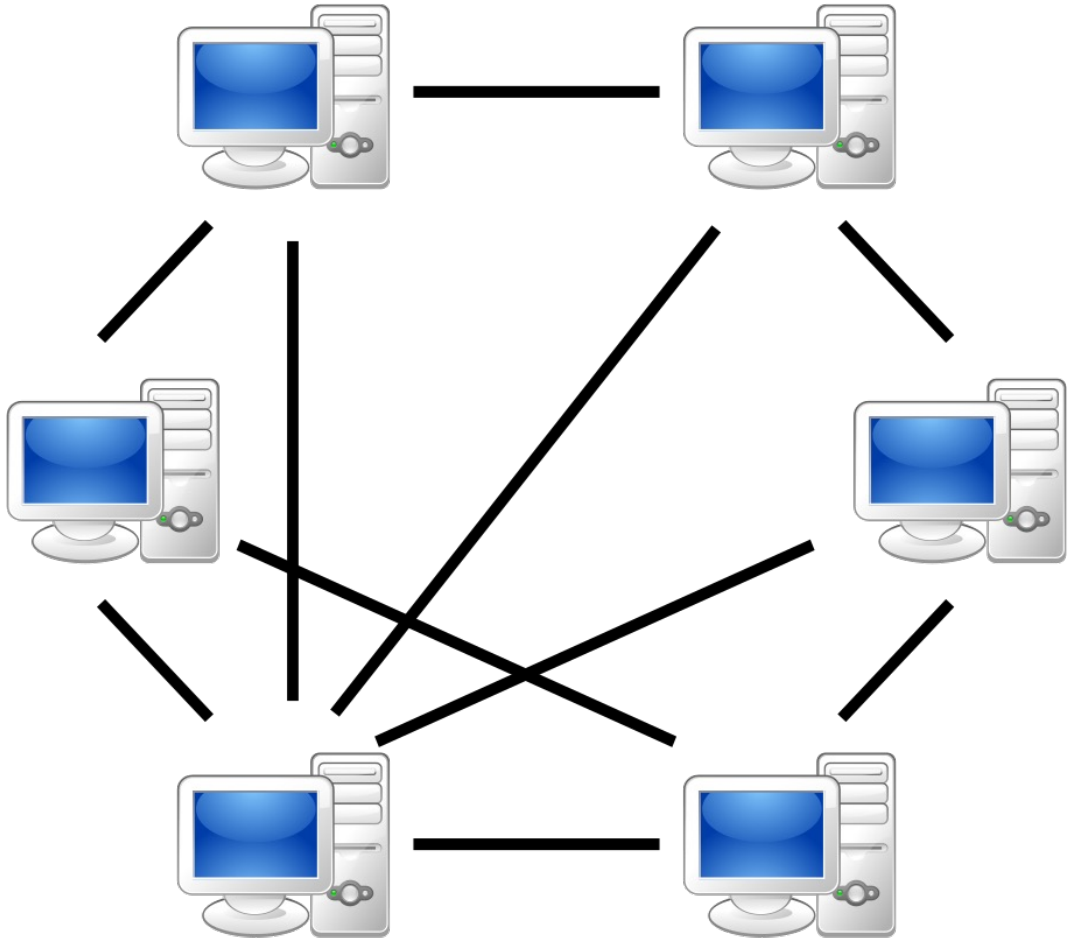
Contd..

Three-tier — It helps move the client intelligence to a middle-tier to use stateless clients. The client doesn't directly communicate with the database server. The client interacts with the server, which in turn interacts with the database. It simplifies application deployment. Most web applications are three-tier.



Peer-to-peer

There are no special machines that provide a service or manage the resources. All of the responsibilities are uniformly distributed among all the machines (peers). They serve both as a client and as a server. Torrent is a good example of a peer-to-peer network.



Different Components

• **Devices or Systems:** The devices or systems in a distributed system have their own processing capabilities and may also store and manage their own data.

• **Network:** The network connects the devices or systems in the distributed system, allowing them to communicate and exchange data.

• **Resource Management:** Distributed systems often have some type of resource management system in place to allocate and manage shared resources such as computing power, storage, and networking.

Characteristics

• **Multiple Devices or Systems:** Processing and data storage is distributed across multiple devices or systems.

• **Peer-to-Peer Architecture:** Devices or systems in a distributed system can act as both clients and servers, as they can both request and provide services to other devices or systems in the network.

• **Shared Resources:** Resources such as computing power, storage, and networking are shared among the devices or systems in the network.

• **Horizontal Scaling:** Scaling a distributed computing system typically involves adding more devices or systems to the network to increase processing and storage capacity. This can be done through hardware upgrades or by adding additional devices or systems to the network..

Applications Area of Distributed System

Finance and Commerce: Amazon, eBay, Online Banking, E-Commerce websites.

Information Society: Search Engines, Wikipedia, Social Networking, Cloud Computing.

Cloud Technologies: AWS, Salesforce, Microsoft Azure, SAP.

Entertainment: Online Gaming, Music, youtube.

Healthcare: Online patient records, Health Informatics.

Education: E-learning.

Transport and logistics: GPS, Google Maps.

Environment Management: Sensor technologies.

Advantages and Disadvantages

Scalability: Distributed systems are generally more scalable than centralized systems, as they can easily add new devices or systems to the network to increase processing and storage capacity.

Reliability: Distributed systems are often more reliable than centralized systems, as they can continue to operate even if one device or system fails.

Flexibility: Distributed systems are generally more flexible than centralized systems, as they can be configured and reconfigured more easily to meet changing computing needs.

There are a few limitations to Distributed Computing System

Complexity: Distributed systems can be more complex than centralized systems, as they involve multiple devices or systems that need to be coordinated and managed.

Security: It can be more challenging to secure a distributed system, as security measures must be implemented on each device or system to ensure the security of the entire system.

Performance: Distributed systems may not offer the same level of performance as centralized systems, as processing and data storage is distributed across multiple devices or systems.

Significant Applications

Distributed Computing Systems have a number of applications, including:

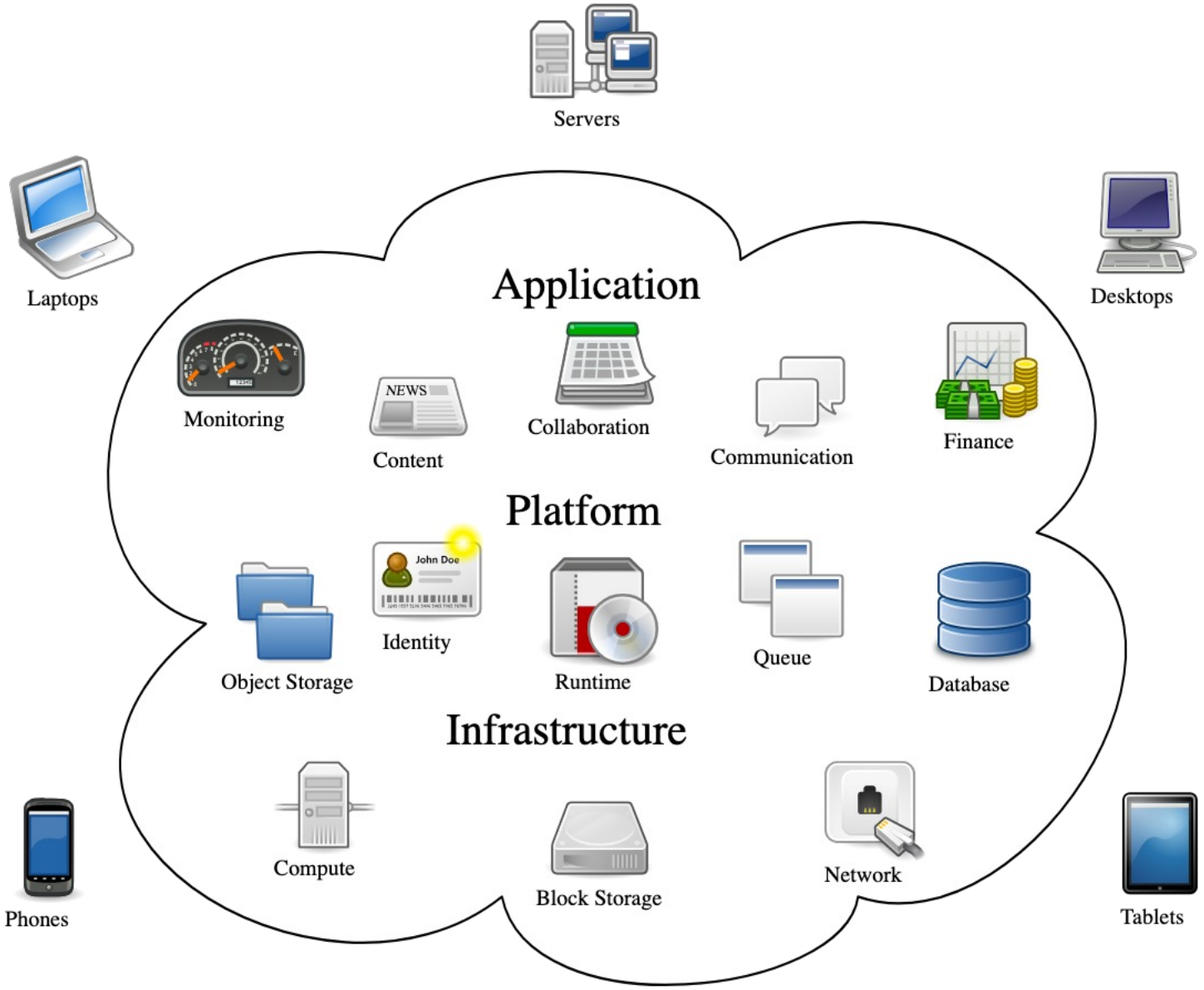
Cloud Computing: Cloud Computing systems are a type of distributed computing system that are used to deliver resources such as computing power, storage, and networking over the Internet.

Peer-to-Peer Networks: Peer-to-Peer Networks are a type of distributed computing system that is used to share resources such as files and computing power among users.

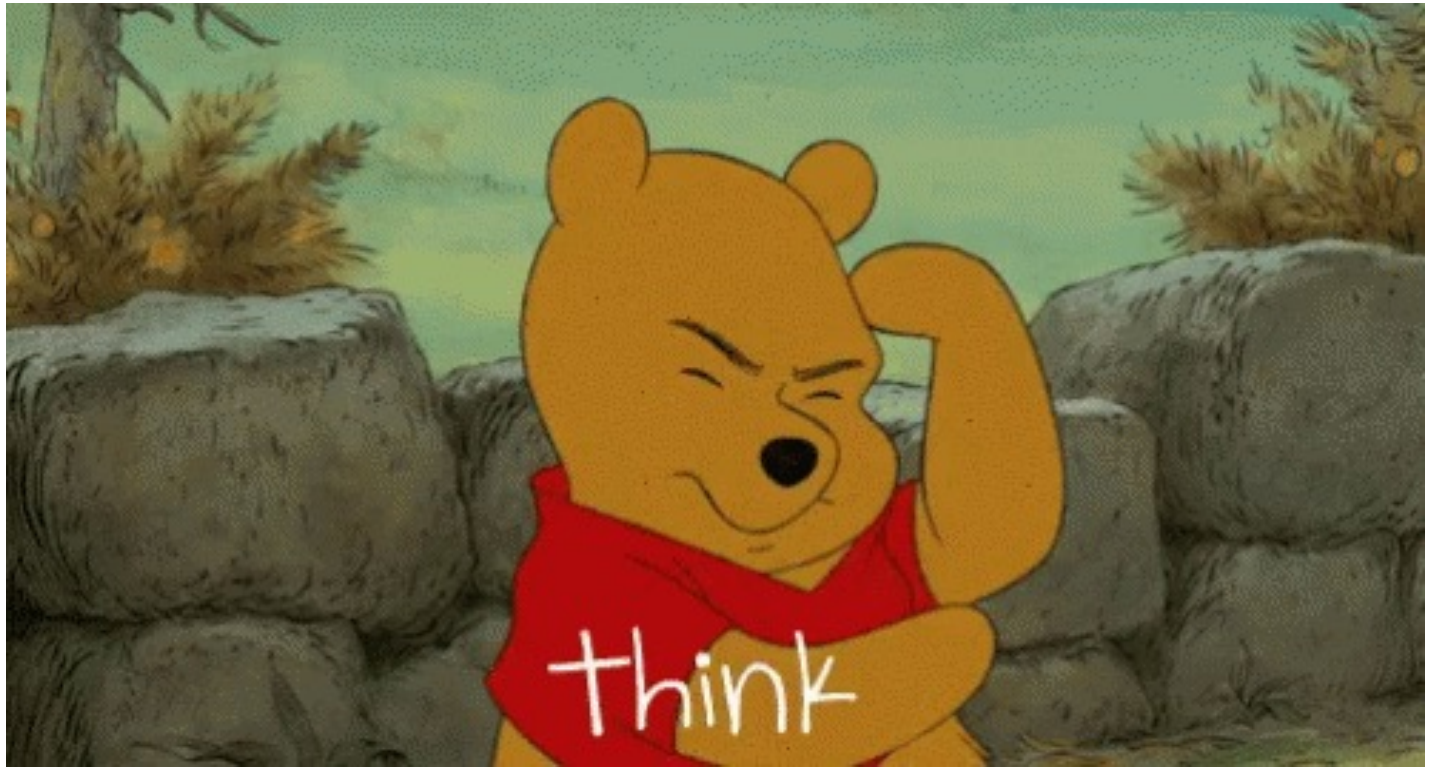
Distributed Architectures: Many modern computing systems, such as microservices architectures, use distributed architectures to distribute processing and data storage across multiple devices or systems.

Cloud Computing

Simply, Cloud computing is the delivery of computing services—including servers, storage, databases, networking, software, analytics, and intelligence—over the Internet (“the cloud”) to offer faster innovation, flexible resources, and economies of scale.



You may not realize it but we have been using cloud computing for over a decade!

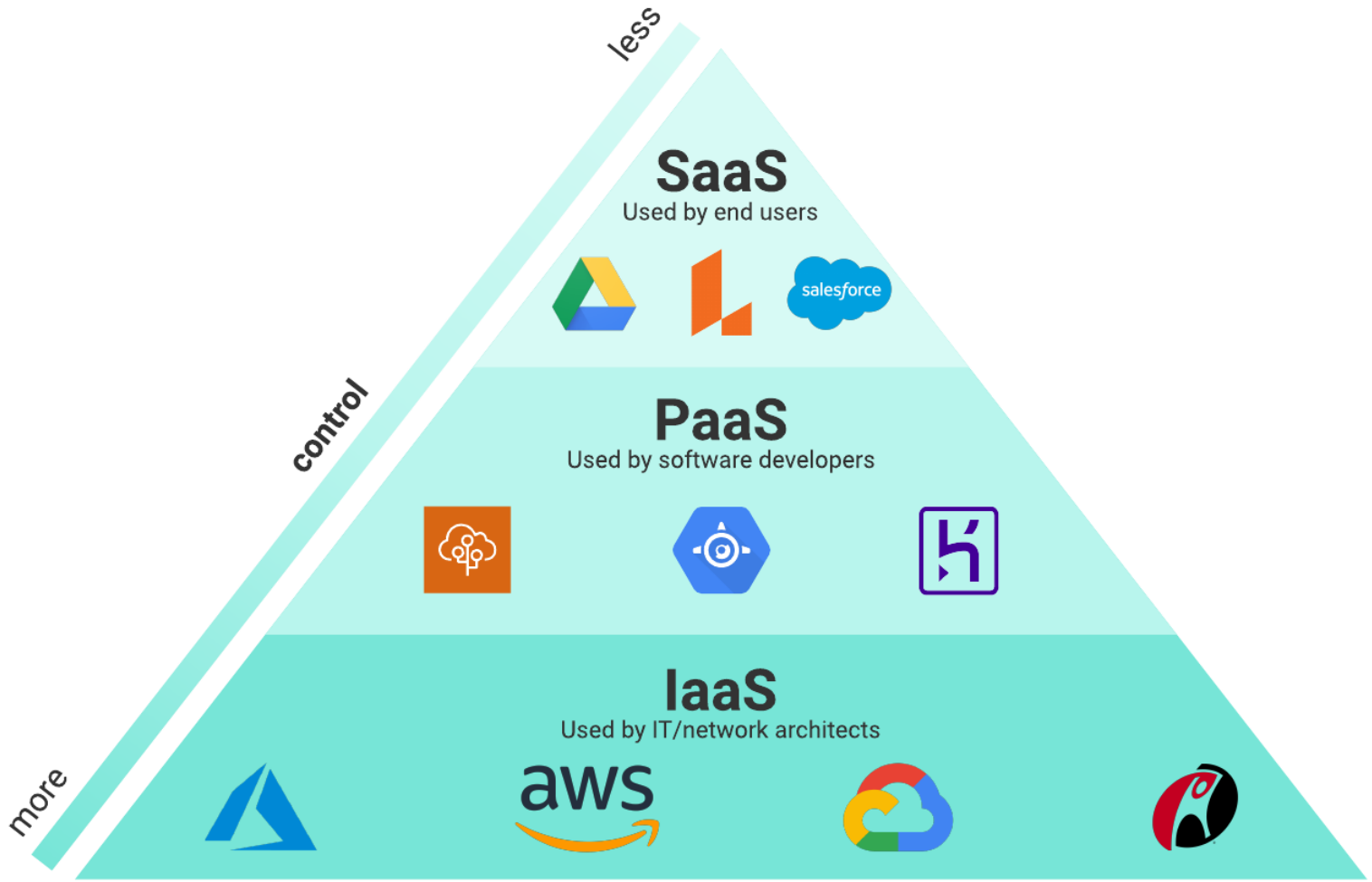


Ever sent an email with an attachment to yourself from your home computer, and then signed in to your email at work to access the attachment?

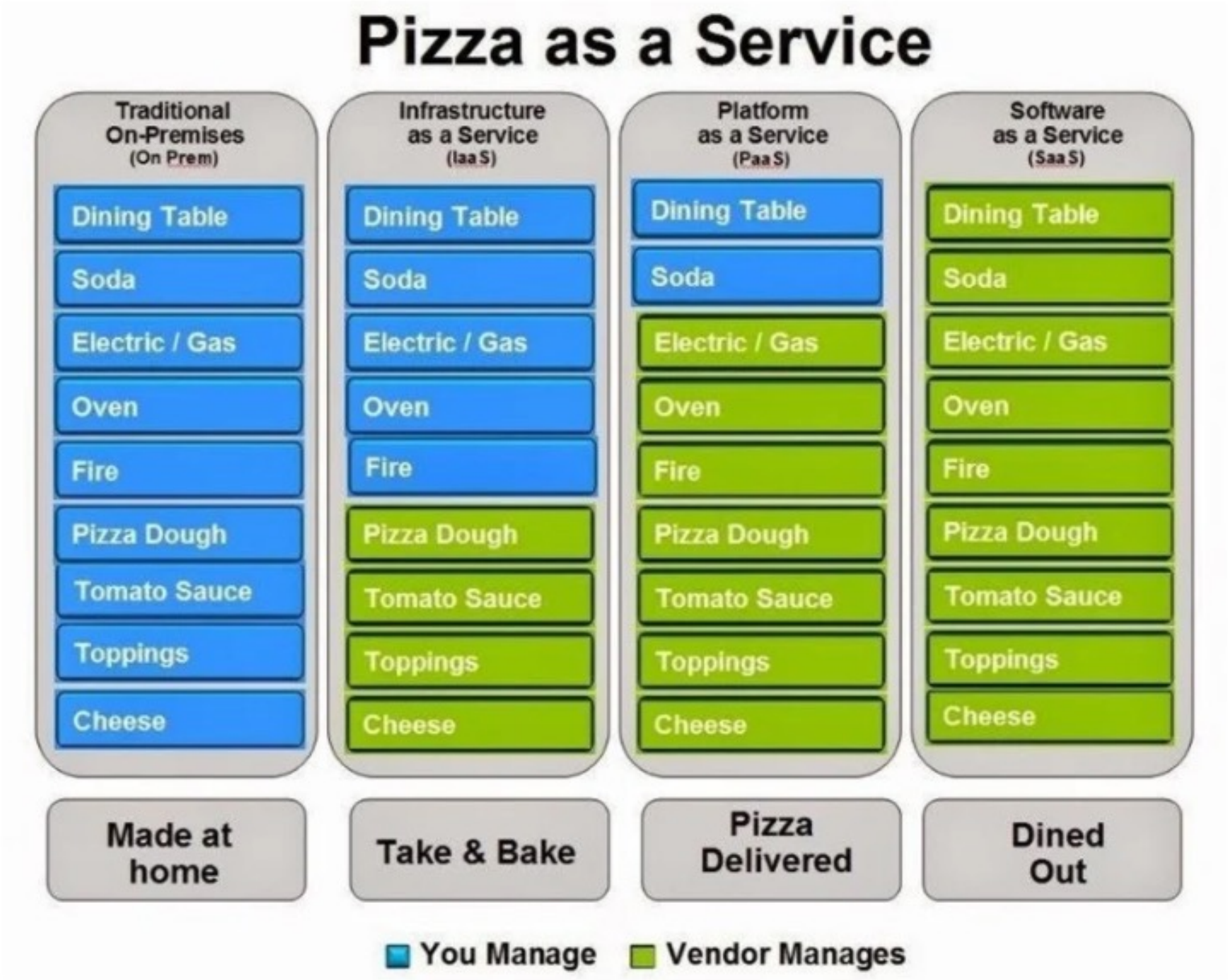
One of the earliest forms of cloud computing was the ability to save attachments in your email inbox and access them on another computer. That's basic cloud storage for you!

Cloud Service Models

Cloud solutions come in three primary service models: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS).



Real World Example



IaaS

IaaS gives users access to storage, networking, servers, and other computing resources via the cloud.

For example, many businesses use IaaS to support workload spikes during busy seasons (like holidays).

Popular IaaS providers include:

Amazon Web Services (AWS), Microsoft Azure, Google Compute Engine (GCE), the IaaS component of Google Cloud Platform (GCP)

PaaS

This service layer is primarily geared towards developers and operations professionals. Service providers rent out cloud-based platforms for users to develop and deliver applications.

Saas

Cloud application services are the most well-known of the cloud service models. The software is hosted, packaged, and delivered by a third party through the Internet (typically on a browser-based interface).

Benefits of Cloud Computing

Fewer Expenses
Moving to the Cloud will reduce the cost of IT systems management and maintenance. Instead of purchasing new systems and equipment for your business, you can take advantage of the resources from your service provider and reduce the cost of system upgrades, energy consumption, and labor.

Security
Most providers follow policies to ensure the security and protection of your data and infrastructure against threats. Also, it enables data backup and recovery which can be reflected throughout multiple services and can be accessed at any time.

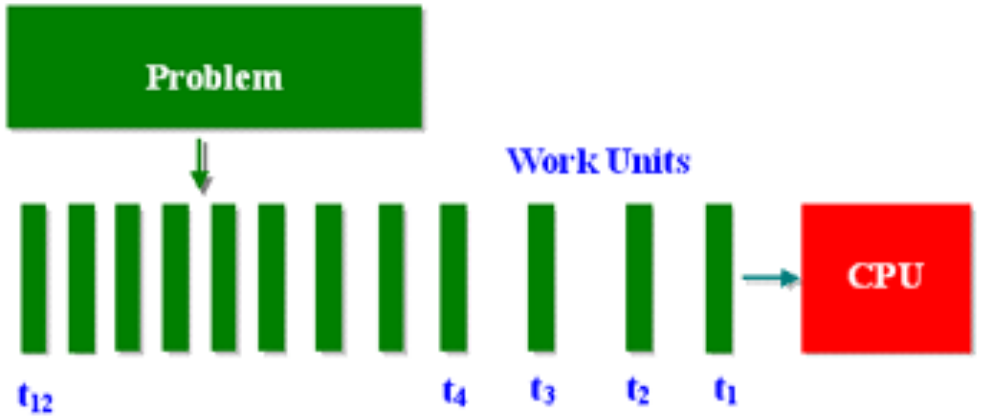
Flexibility
We all know how important it is for businesses to be flexible especially in these trying times. Cloud computing will allow your employees to work anywhere since they will get access to data, provided they have an internet connection.

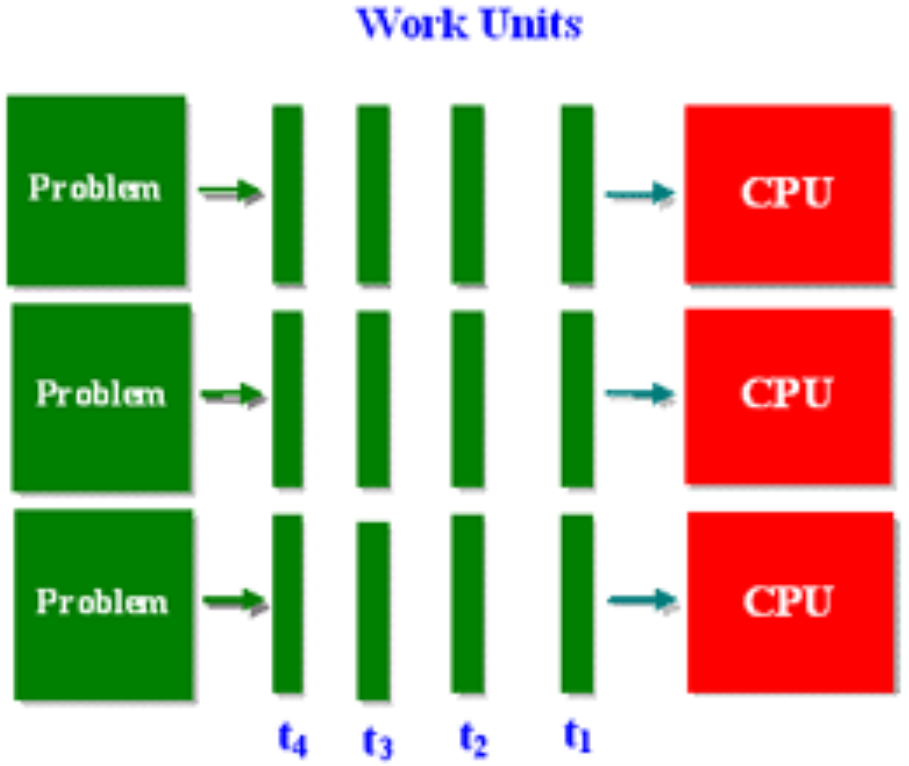
Collaboration
When your team needs to work on a certain project, and each are in different locations, they can utilize cloud computing to work and share files at the same time. It is secure and will allow your business to connect and communicate easily.

Automatic Updates
Cloud-based applications automatically update so you don’t need to call for support to manually install updates on your device every time. Updates could include the latest versions in addition to server upgrades and more, which allows your business to save time and money.

Parallel Computing

Parallel computing refers to the process of executing several processors an application or computation simultaneously.





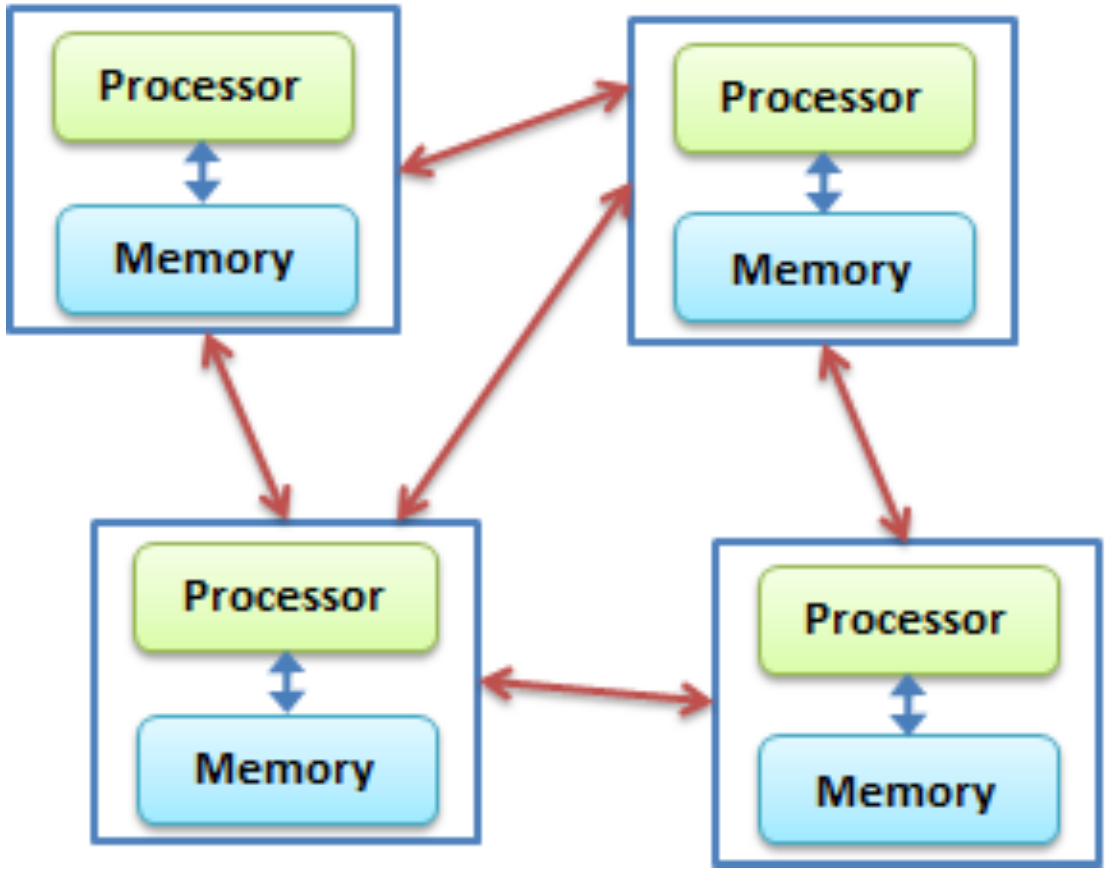
Advantages It saves time and money as many resources working together will reduce the time and cut potential costs.

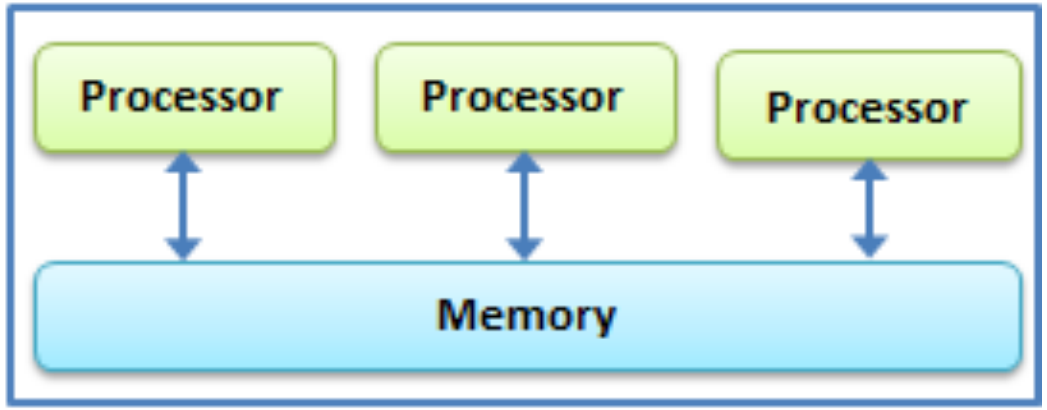
1.It can be impractical to solve larger problems on Serial Computing.

2.It can take advantage of non-local resources when the local resources are finite.

3.Serial Computing ‘wastes’ the potential computing power, thus Parallel Computing makes better work of the hardware.

Difference between Distributed computing and Parallel Computing





There are two types of computations: parallel computing and distributed computing. Parallel computing allows several processors to accomplish their tasks at the same time. In contrast, distributed computing splits a single task among numerous systems to achieve a common goal.

Features	Parallel Computing	Distributed Computing
Definition	It is a type of computation in which various processes runs simultaneously.	It is that type of computing in which the components are located on various networked systems that interact and coordinate their actions by passing messages to one another.
Communication	The processors communicate with one another via a bus.	The computer systems connect with one another via a network.
Functionality	Several processors execute various tasks simultaneously in parallel computing.	Several computers execute tasks simultaneously.
Number of Computers	It occurs in a single computer system.	It involves various computers.
Memory	The system may have distributed or shared memory.	Each computer system in distributed computing has its own memory.
Usage	It helps to improve the system performance	It allows for scalability, resource sharing, and the efficient completion of computation tasks.

Types of Parallelism

- **Bit-level parallelism –**
  It is the 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.

- **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.

- **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.

- **Data-level parallelism (DLP) –**
  Instructions from a single stream operate concurrently on several data – Limited by non-regular data manipulation patterns and by memory bandwidth

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 parallel computing's approach.
- Ensures the effective utilization of the resources. The hardware is guaranteed to be used effectively whereas in serial computation only some part of the hardware was used and the rest 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 programs 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.

Thanks & Cheers!!

Small aim is a crime; have great aim.
Bharat-Ratan A. P. J. Abdul Kalam