SNS COLLEGE OF ENGINEERING

Kurumbapalayam(Po), Coimbatore – 641

107 **An Autonomous Institution**

Accredited by NAAC-UGC with 'A' Grade

Approved by AICTE, Recognized by UGC & Affiliated to Anna University, Chennai

DEPARTMENT OF COMPUTER SCIENCE AND DESIGN

Course Code and Name :19TS501 AND CLOUD COMPUTING

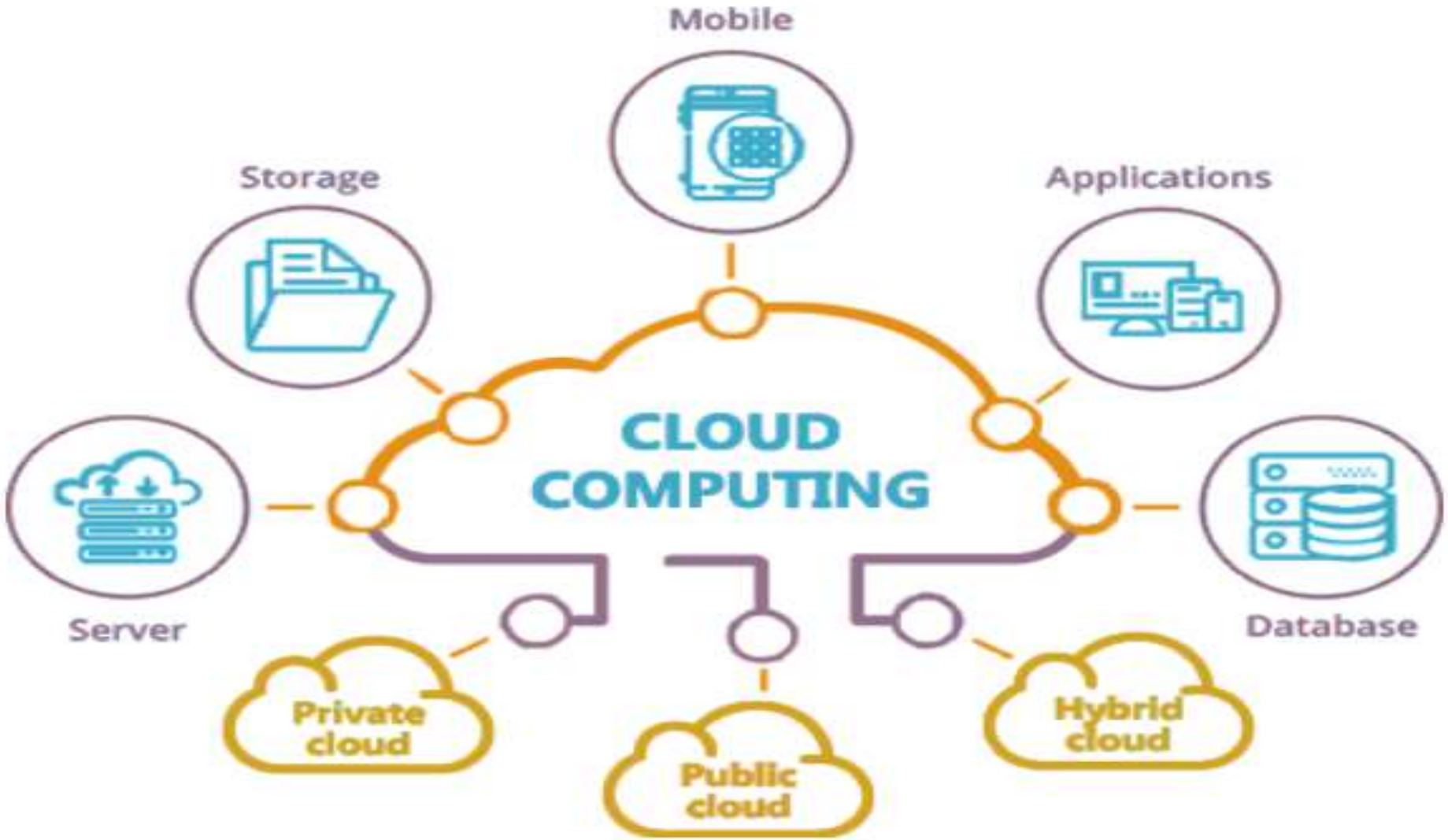
Unit 1: Cloud Computing

Topic : INTRODUCTION TO CLOUD COMPUTING

Introduction

What is Cloud Computing?

Cloud computing refers to the delivery of computing services over the internet. Instead of owning and maintaining physical servers and infrastructure, cloud computing allows users to access resources like servers, storage, databases, networking, and software applications over the internet from a cloud service provider.


Mobile
Storage
Applications
CLOUD
COMPUTING
Server
Database
Private
cloud
Public
cloud
Hybrid
cloud

KEY CONCEPTS:

1.On-Demand Self-Service:

Users can provision and manage resources, such as computing power and storage, without the need for human interaction with the service provider.

2. Broad Network Access:

Cloud services are accessible over the internet from various devices, such as laptops, smartphones, and tablets.

3.Resource Pooling:

Cloud providers use a multi-tenant model, where resources are shared among multiple users, allowing for efficient resource utilization.

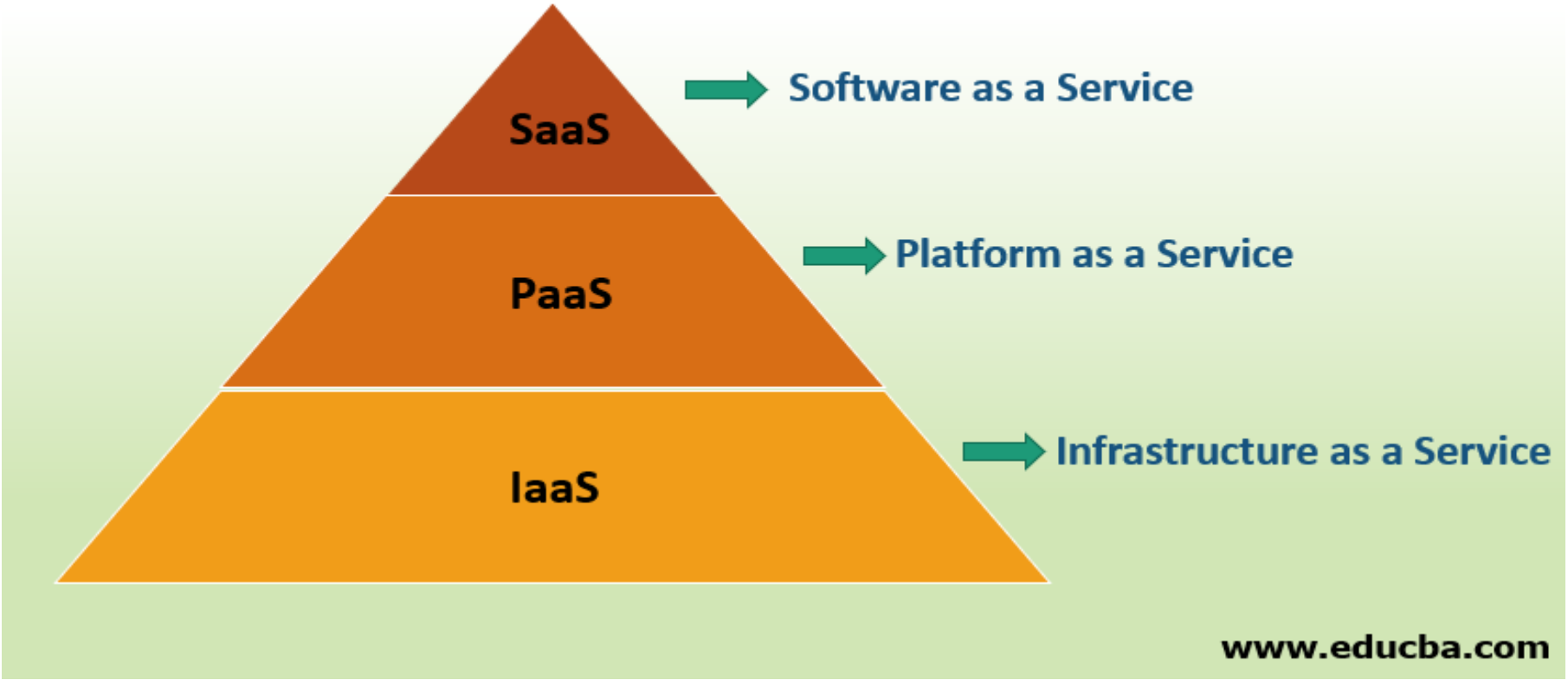
4.Rapid Elasticity:

Cloud resources can be scaled up or down based on demand, ensuring that you pay for only the resources you use.

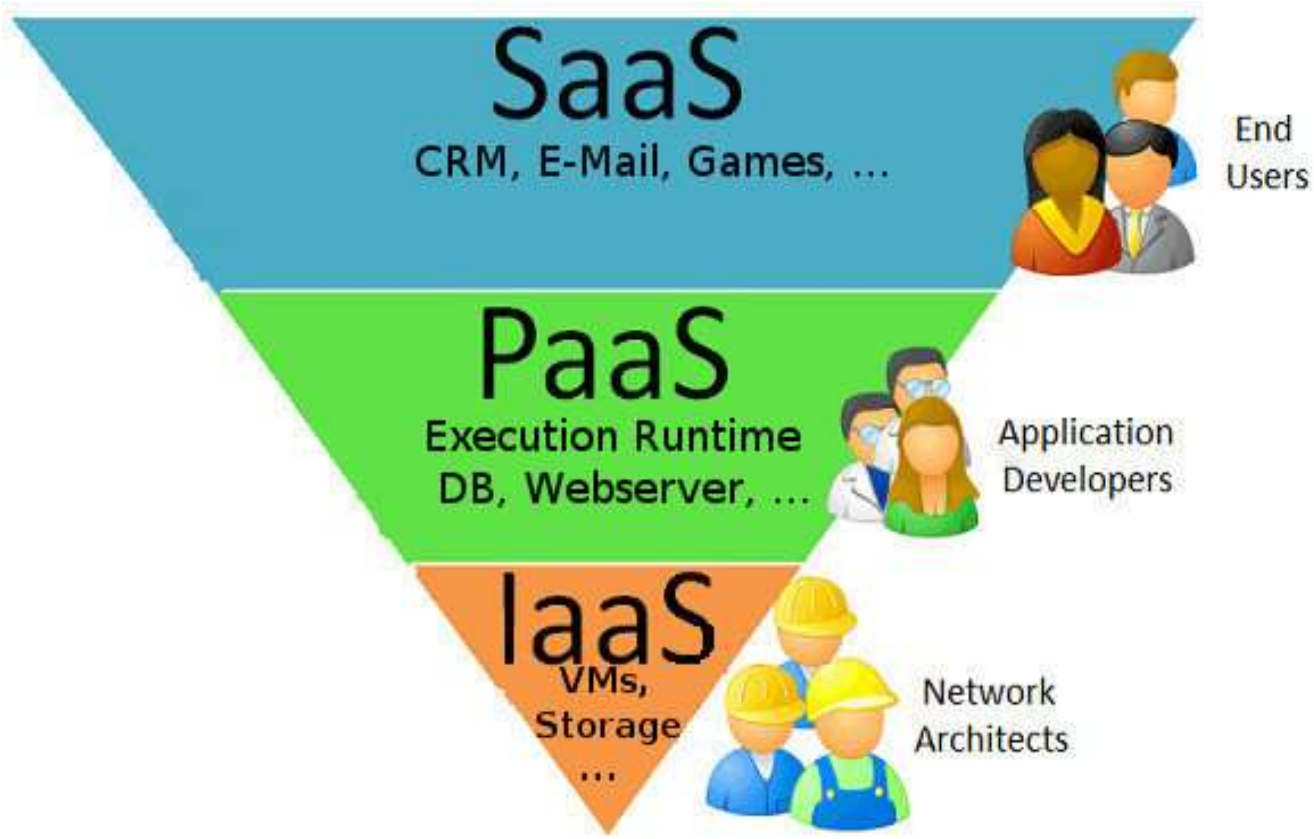
5.Measured Service:

Cloud service usage is metered, and users are billed for the actual resources consumed (e.g., CPU hours, storage used, data transfer).

Cloud Service Models



CLOUD SERVICE MODEL:



Cloud Service Models:

- **Infrastructure as a Service (IaaS)**:

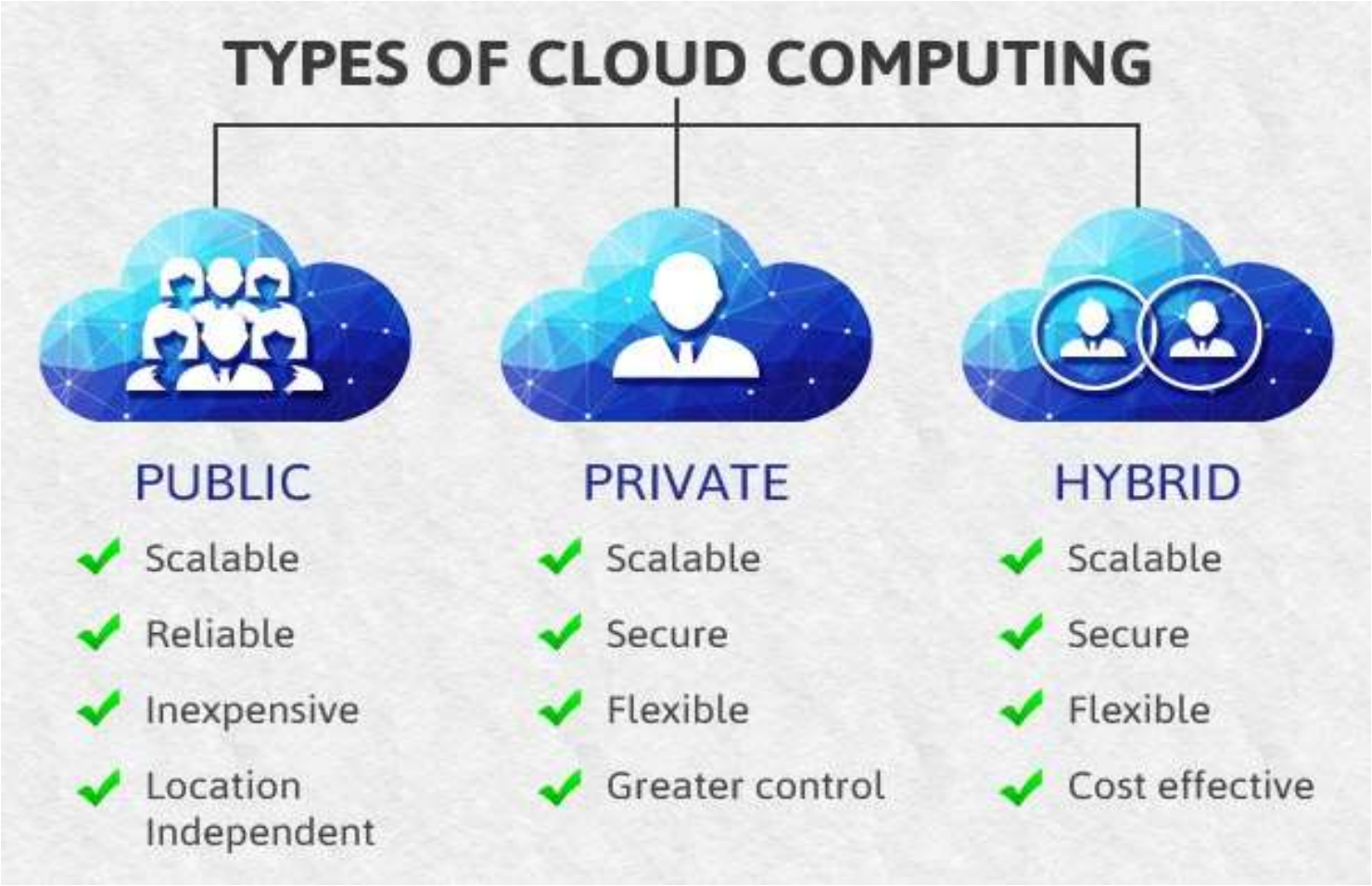
  Provides virtualized computing resources over the internet. Users can rent virtual machines, storage, and networking components.

- **Platform as a Service (PaaS)**:

  Offers a platform and environment to build, deploy, and manage applications without dealing with underlying infrastructure.

- **Software as a Service (SaaS)**:

  Delivers software applications over the internet on a subscription basis. Users access applications through a web browser without the need for installation.


TYPES OF CLOUD COMPUTING
PUBLIC
Scalable
Reliable
Inexpensive
Location Independent
PRIVATE
Scalable
Secure
Flexible
Greater control
HYBRID
Scalable
Secure
Flexible
Cost effective

1.Public Cloud:

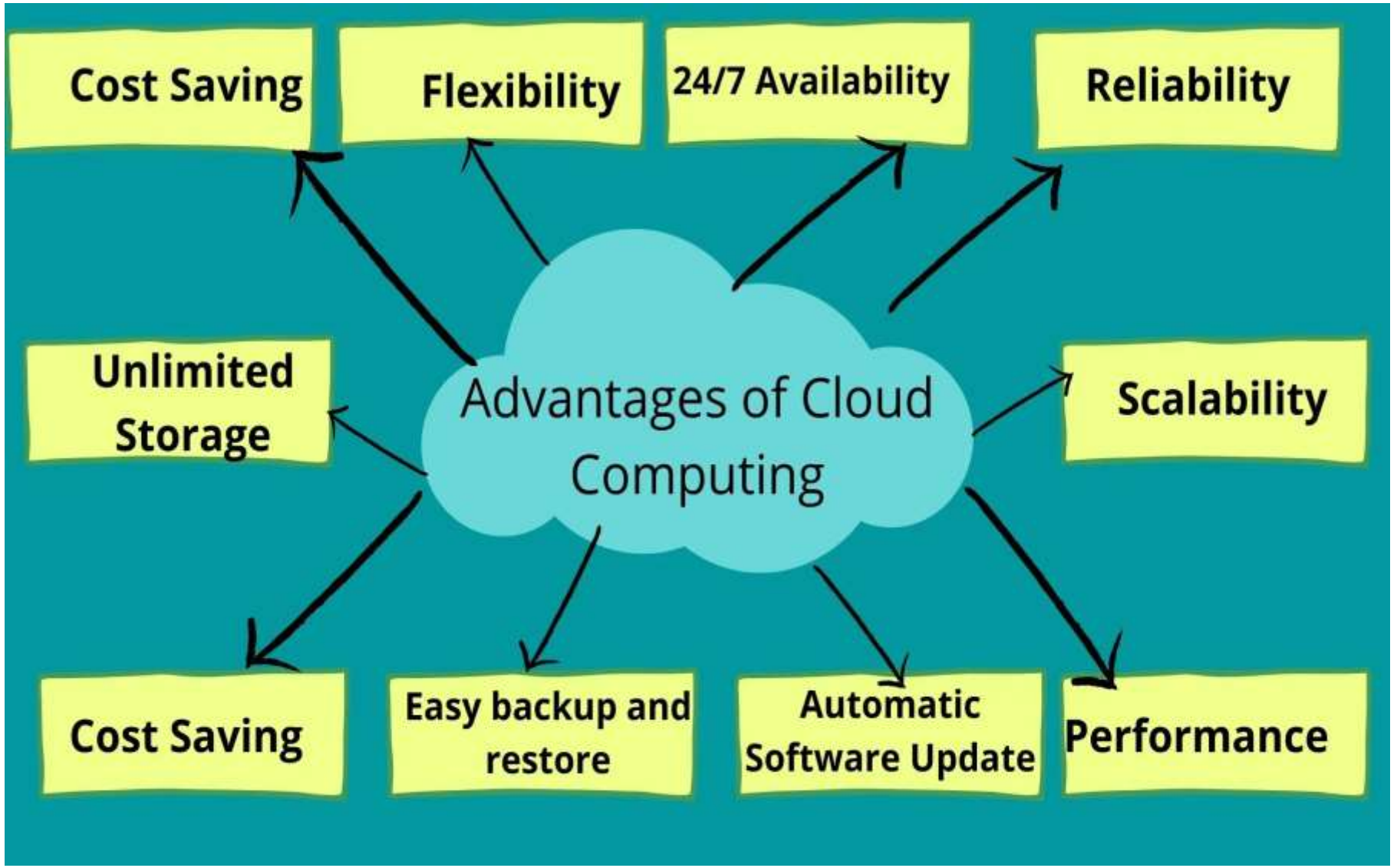
Services are provided over the public internet and are available to anyone who wants to use them. Examples include Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP).

2. Private Cloud:

Services are hosted on a private network and used exclusively by a single organization. It offers more control and security but requires significant infrastructure investment.

3.Hybrid Cloud:

Combines elements of public and private clouds, allowing data and applications to be shared between them.


Cost Saving
Flexibility
24/7 Availability
Reliability
Unlimited Storage
Advantages of Cloud Computing
Scalability
Cost Saving
Easy backup and restore
Automatic Software Update
Performance

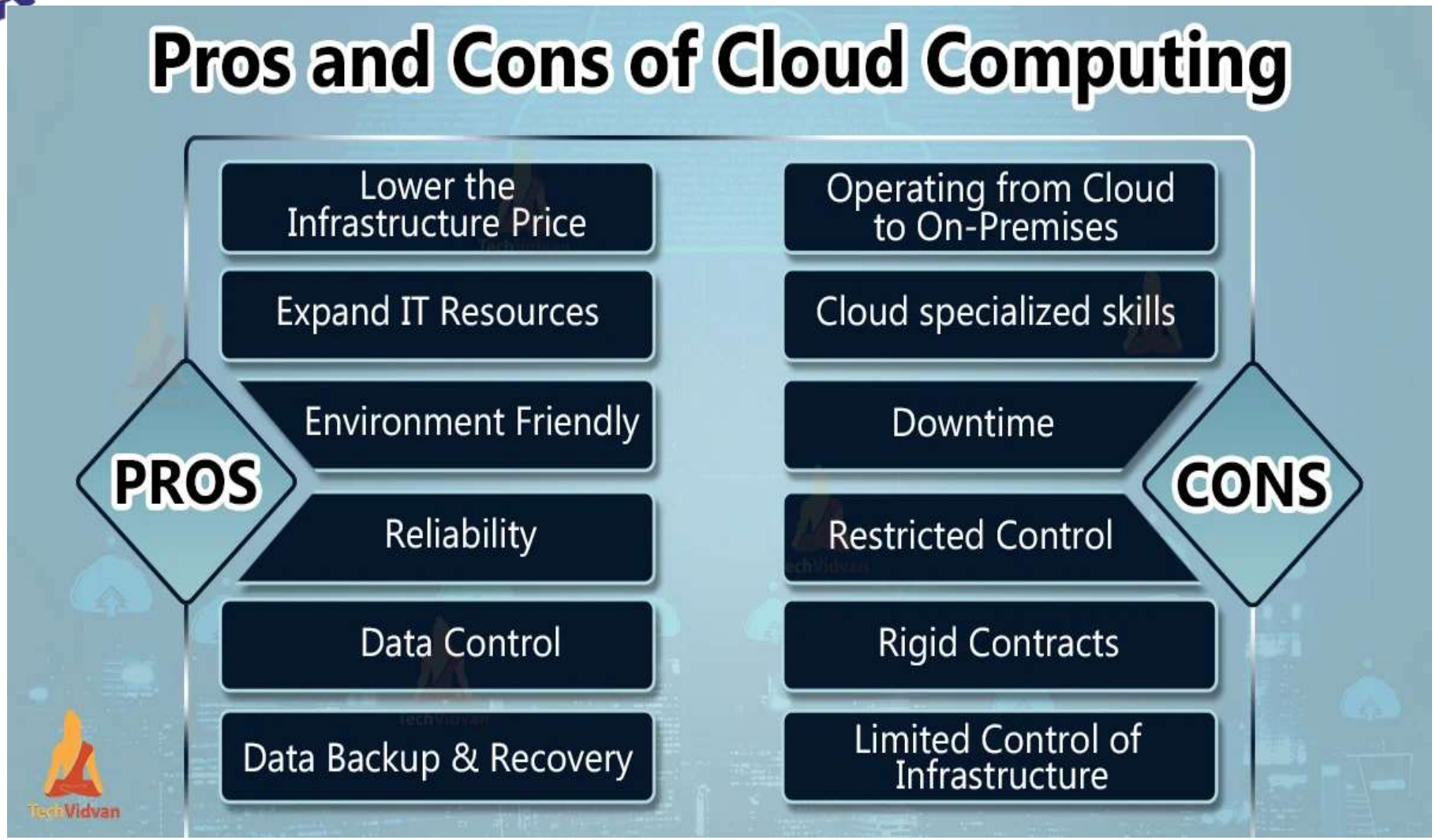
ADVANTAGES:

- **Cost Efficiency**: Cloud computing eliminates the need for upfront infrastructure investments and allows pay-as-you-go billing, reducing overall costs.
- **Scalability**: Cloud resources can be easily scaled up or down based on demand, providing flexibility and cost savings.
- **Reliability**: Leading cloud providers offer high levels of uptime and redundancy, ensuring continuous availability of services.
- **Global Accessibility**: Cloud services are accessible from anywhere with an internet connection, enabling remote collaboration and flexibility.
- **Security**: Cloud providers invest in robust security measures to protect data, often providing better security than individual organizations can implement.


Security Issues
Management Issues
Lack of Expertise
Performance Issues
Migration
Disadvanage of
Cloud
Computing

DISADVANTAGES:

- **Internet Dependency**: Cloud computing heavily relies on internet connectivity. If there is a network outage or poor internet connection, access to cloud services and data may be disrupted, impacting productivity and operations.
- **Security and Privacy Concerns**: Storing sensitive data on remote servers raises security and privacy concerns. Data breaches or unauthorized access to data can have severe consequences.
- **Downtime and Outages**: Even reputable cloud service providers experience occasional downtime or outages due to maintenance, hardware failures, or other issues.
- **Data Transfer and Bandwidth Costs**: Uploading and downloading large amounts of data to and from the cloud can lead to significant data transfer costs.
- **Limited Control and Customization**: Cloud services are usually standardized to cater to a wide range of users. This can result in limitations regarding customization and control over the underlying infrastructure or software.


Pros and Cons of Cloud Computing
PROS
Lower the Infrastructure Price
Expand IT Resources
Environment Friendly
Reliability
Data Control
Data Backup & Recovery
CONS
Operating from Cloud to On-Premises
Cloud specialized skills
Downtime
Restricted Control
Rigid Contracts
Limited Control of Infrastructure
TechVidvan

DEFINITION OF CLOUD

- The "cloud" refers to a vast network of remote servers that store and manage data and applications over the internet.
- Cloud computing allows users to access and utilize these resources from anywhere with an internet connection, eliminating the need for local hardware and infrastructure management.
- It offers on-demand services, scalability, flexibility, and cost-effectiveness, enabling individuals and organizations to efficiently deploy and utilize computing resources without the complexities of maintaining physical servers.

DEFINITION OF CLOUD COMPUTING

- Cloud computing is a technology that enables the delivery of various computing services over the internet.
- It allows users and organizations to access and use a wide range of computing resources, such as servers, storage, databases, software, and applications, without the need to own or manage the physical infrastructure.

- In cloud computing, service providers host and maintain the computing resources in large data centers, and users can access these resources remotely through the internet.
- This remote accessibility and resource management are achieved through virtualization, where physical hardware resources are abstracted and made available as virtual resources to users

EVOLUTION OF CLOUD COMPUTING

Conceptual Origins (1950s - 1990s):

- The seeds of cloud computing were sown in the 1950s when the idea of time-sharing and resource sharing among multiple users on a single mainframe computer was introduced.
- During the 1960s and 1970s, this concept evolved with the development of virtualization techniques.
- In the 1990s, the term "cloud" was used to represent network-based computing.

Internet Boom (Late 1990s - Early 2000s):

- The advent of the internet and the dot-com boom further advanced cloud computing concepts.
- Companies began offering web-based applications and services accessible over the internet.
- This period also witnessed the rise of application service providers (ASPs) who provided software applications remotely.

Utility Computing and Grid Computing (Early 2000s):

- Utility computing and grid computing concepts emerged during this time, offering computing resources and power on a pay-per-use basis.
- These ideas laid the groundwork for the utility-like nature of cloud services we see today.

Amazon Web Services (AWS) Launch (2006):

- Amazon Web Services was officially launched in 2006, providing the first commercially successful cloud computing platform.
- AWS offered Infrastructure as a Service (IaaS), giving developers and businesses access to scalable and flexible computing resources.

Expansion of Cloud Providers (Late 2000s - Early 2010s):

- As cloud computing gained popularity, other major tech companies such as Google (Google Cloud Platform).
- This period marked the start of the cloud computing "war" between these tech giants.

Proliferation of Cloud Services (2010s):

- The 2010s saw a significant increase in the types of cloud services available, including Platform as a Service (PaaS) and Software as a Service (SaaS).
- Cloud providers diversified their offerings to cater to various needs and industries.

Hybrid Cloud and Multi-Cloud Strategies (2010s):

- Hybrid cloud models, combining public cloud services with private on-premises infrastructure, gained popularity as organizations sought to leverage both cloud and local resources.
- Additionally, many companies adopted multi-cloud strategies, using multiple cloud providers for different services or workloads to avoid vendor lock-in.

Serverless Computing (2010s):

- Serverless computing, also known as Function as a Service (FaaS), emerged as a cloud computing model where developers could execute code in response to events without managing servers.
- This model further simplified application development and deployment.

Serverless Computing (2010s):

- Serverless computing, also known as Function as a Service (FaaS), emerged as a cloud computing model where developers could execute code in response to events without managing servers.
- This model further simplified application development and deployment.

Advancements in Artificial Intelligence and Machine Learning (2010s):

- Cloud computing played a crucial role in democratizing access to advanced technologies like artificial intelligence and machine learning.

- Cloud providers started offering pre-built AI/ML services, making it easier for developers to integrate AI capabilities into their applications

Edge Computing (Late 2010s - 2020s):

- Edge computing evolved as a complement to cloud computing, bringing computation closer to the data source and reducing latency.
- It enabled new applications and services that require real-time data processing and analysis.

Focus on Security and Compliance (2020s):

- As cloud adoption continued to grow, there was an increased emphasis on cloud security and regulatory compliance.
- Cloud providers invested in robust security measures and compliance certifications to address data protection concerns.

Quantum Computing (2020s):

Towards the end of the 2010s and into the 2020s, cloud providers started exploring quantum computing services, making this cutting-edge technology more accessible to researchers and developers.

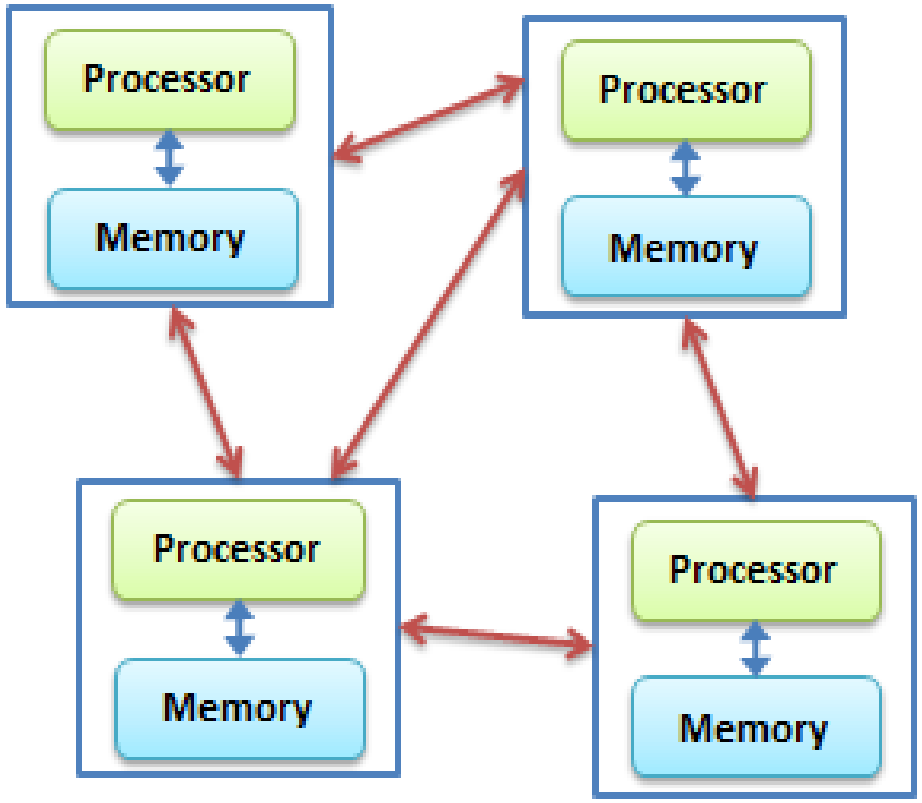
Parallel and distributed computing

- Parallel and distributed computing are two closely related fields
  .

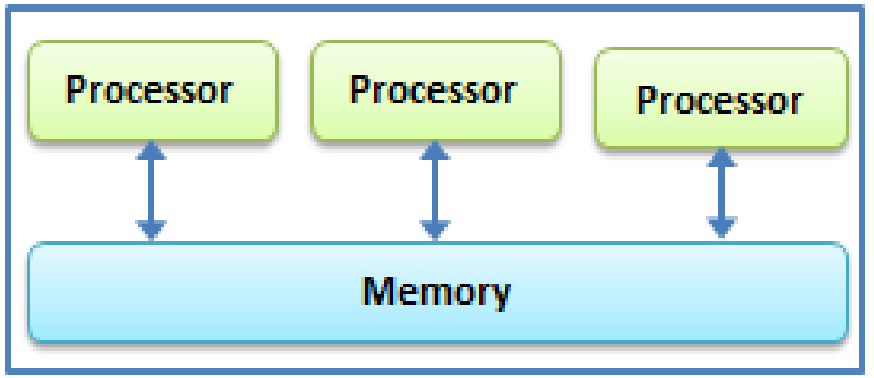
- They aim to solve computational problems by leveraging multiple computing resources.

- While they have some similarities, they also have distinct differences in their architectures and goals.

Distributed Computing



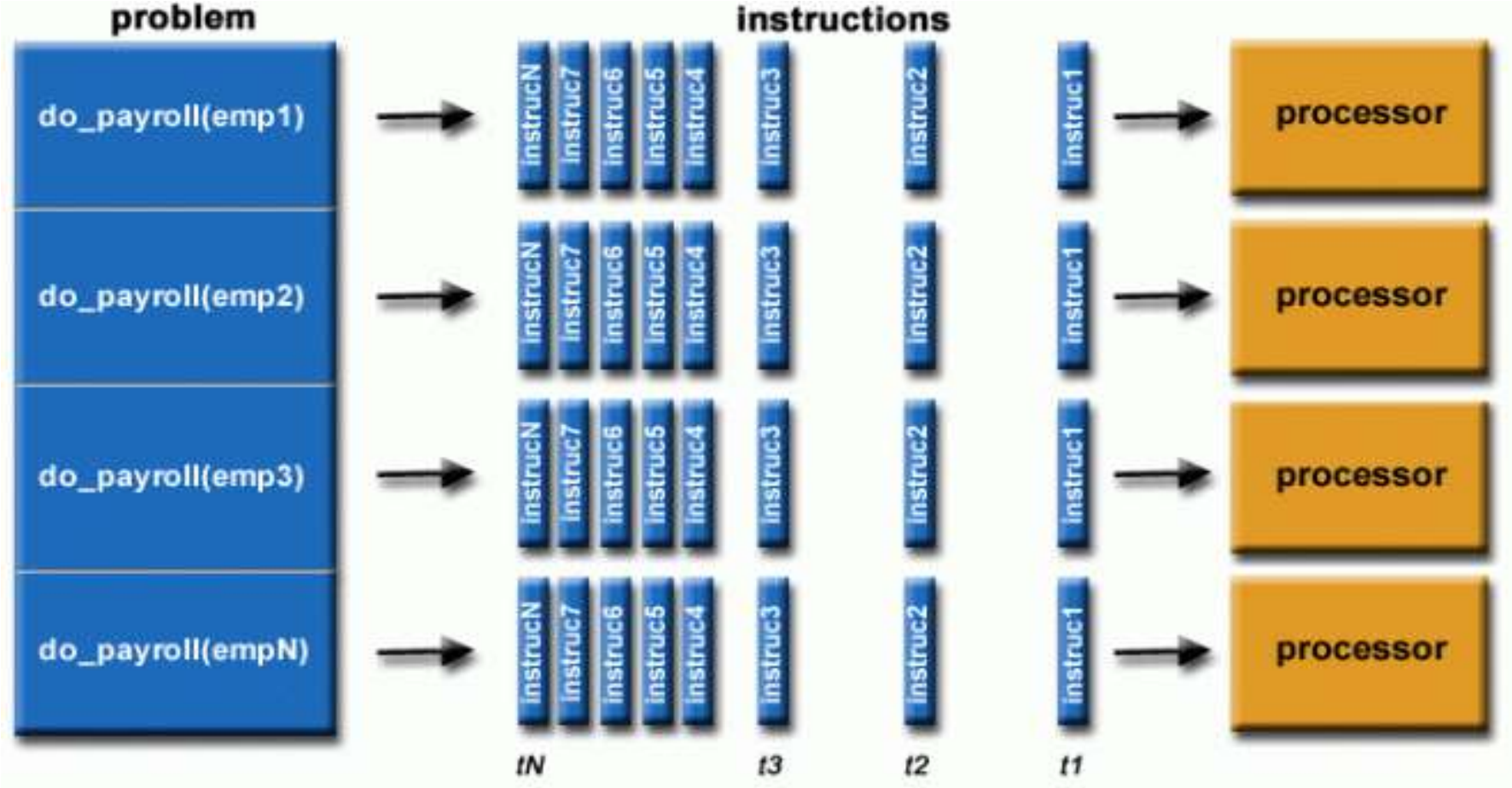
Parallel Computing



PARALLEL COMPUTING

- Parallel computing is a type of computing paradigm in which multiple processors or computing resources work together to solve a computational problem.

- The goal of parallel computing is to break down a large task into smaller sub-tasks that can be processed simultaneously, thereby reducing the overall execution time and improving computational efficiency.


problem
instructions
do_payroll(emp1)
do_payroll(emp2)
do_payroll(emp3)
do_payroll(empN)
instrucN
instruc7
instruc6
instruc5
instruc4
instruc3
instruc2
instruc1
processor
tN
t3
t2
t1

- **Applications for Parallel Processing**
- Science and Engineering
  - Atmospheric Analysis
  - Earth Sciences
  - Electrical Circuit Design
- Industrial and Commercial
  - Data Mining
  - Web Search Engine
  - Graphics Processing

PARALLEL COMPUTING: KEY CONCEPTS

1.Task Decomposition

➢Breaking down a complex problem into smaller, independent tasks that can be executed in parallel.

➢This involves identifying dependencies and ensuring that tasks can be executed without conflicting with each other.

2.Data Independence

- ➢ Managing the flow of data between different parallel tasks.
- ➢ It's crucial to ensure that data dependencies are properly handled to avoid race conditions and ensure correctness.

3.Load Balancing

➢Distributing the workload evenly among the available processing units to maximize resource utilization and minimize idle time.

4.Communication Overhead:

- Minimizing the overhead associated with communication between parallel tasks.
- Efficient communication mechanisms are essential to avoid performance bottlenecks.

5.Scalability:

- Designing parallel systems that can handle increasing workloads by adding more processing units.
- Scalability is crucial to accommodate larger problems efficiently.

DISTRIBUTED COMPUTING

- Distributed computing is a computing paradigm in which a group of interconnected computers work together to solve a computational problem or perform a task.
- Unlike traditional centralized computing, where a single computer handles all processing, distributed computing distributes the workload across multiple computers.
- They often referred to as nodes or servers, that are connected via a network.

Communication Protocols:

- Communication is fundamental in distributed systems.
- Protocols for data exchange, synchronization, and coordination must be implemented efficiently to ensure smooth operation.

Scalability and Elasticity:

Distributed computing systems should be designed to scale horizontally, adding or removing nodes as needed to handle varying workloads.

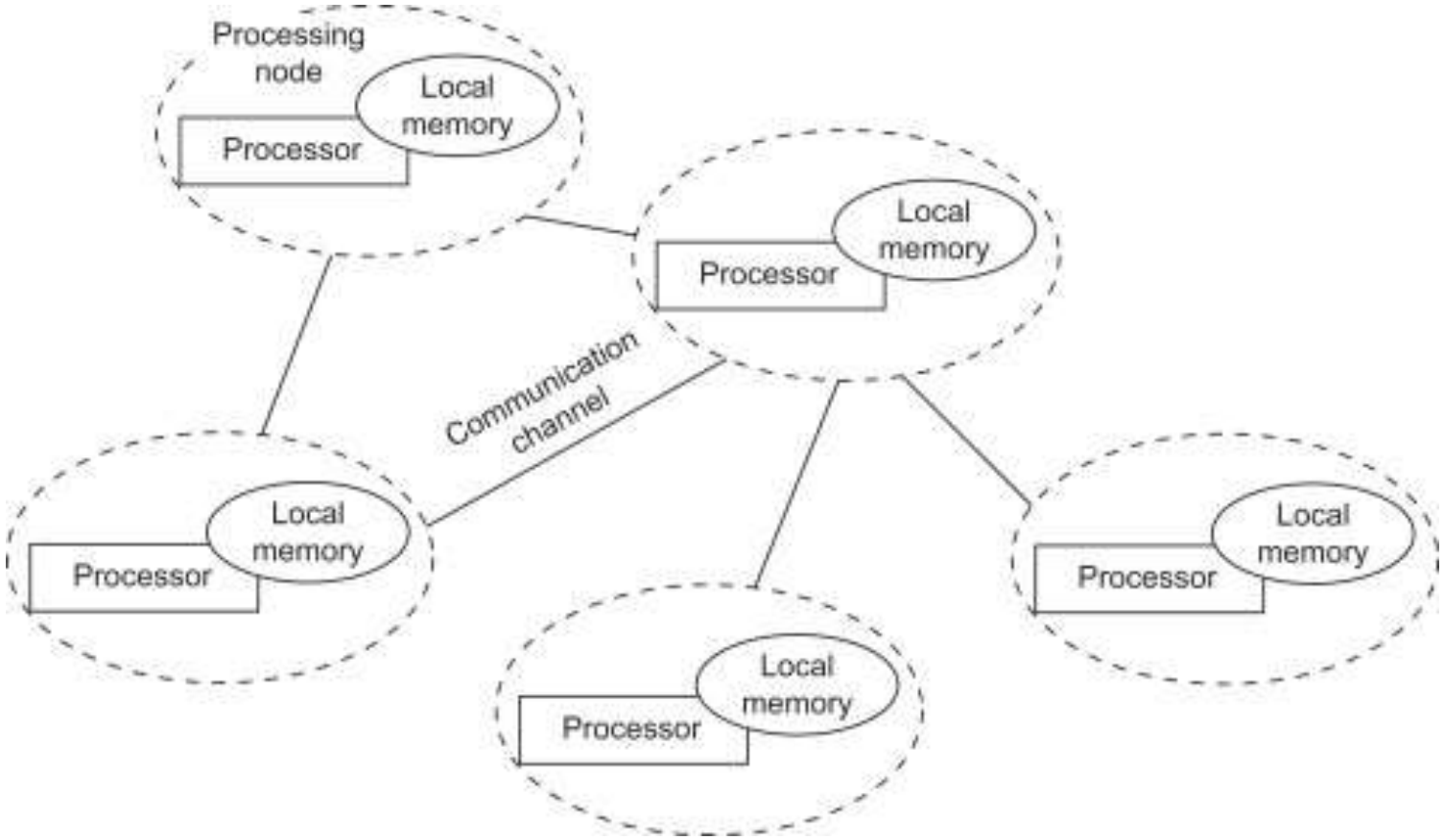
Consistency and Coordination:

Ensuring data consistency and coordinating actions among distributed nodes is a complex challenge, especially in systems with high concurrency.

Shared Principles:

While parallel and distributed computing have distinct characteristics, they share common principles:

- Both aim to solve large-scale computational problems efficiently.
- They rely on breaking down tasks to utilize multiple resources simultaneously.
- Load balancing and efficient communication are crucial for performance optimization.
- Scalability and fault tolerance are essential to handle increasing workloads and ensure system reliability.


Processing node
Local memory
Processor
Local memory
Processor
Communication channel
Local memory
Processor
Local memory
Processor
Local memory
Processor

S.NO	Parallel Computing	Distributed Computing
1.	Single computer is required	Uses multiple computers
2.	Multiple processors perform multiple operations	Multiple computers perform multiple operations
3.	It may have shared or distributed memory	It have only distributed memory
4.	Processors communicate with each other through bus	Computer communicate with each other through message passing.
5.	Improves the system performance	Improves system scalability, fault tolerance and resource sharing capabilities

APPLICATIONS

- Telecommunication networks: telephone networks and cellular networks, ...
- Network applications: World Wide Web and peer-to-peer networks, ...
- Real-time process control: aircraft control systems, ...
- Parallel computation: ...
- Peer-to-peer.

Distributed Computing: Key concepts

Task Distribution:

In distributed computing, the focus is on distributing tasks or parts of a task across multiple interconnected machines (nodes) in a network.

Data Partitioning:

- Data partitioning is critical to distributed computing.
- Large datasets may be split and distributed among nodes, with each node processing its portion of the data.

Communication Protocols:

- Communication is fundamental in distributed systems.
- Protocols for data exchange, synchronization, and coordination must be implemented efficiently to ensure smooth operation.

Scalability and Elasticity:

Distributed computing systems should be designed to scale horizontally, adding or removing nodes as needed to handle varying workloads.

Consistency and Coordination:

Ensuring data consistency and coordinating actions among distributed nodes is a complex challenge, especially in systems with high concurrency.

Cloud Elasticity:

- Elasticity refers to the ability of a cloud to automatically expand or compress the infrastructural resources on a sudden up and down in the requirement so that the workload can be managed efficiently.
- This elasticity helps to minimize infrastructural costs.
- This is not applicable for all kinds of environments, it is helpful to address only those scenarios where the resource requirements fluctuate up and down suddenly for a specific time interval.
- It is not quite practical to use where persistent resource infrastructure is required to handle the heavy workload.

Example:

- Consider an online shopping site whose transaction workload increases during festive season like Christmas.
- So for this specific period of time, the resources need a spike up. In order to handle this kind of situation, we can go for a Cloud-Elasticity service rather than Cloud Scalability.
- As soon as the season goes out, the deployed resources can then be requested for withdrawal.

Cloud Scalability

- Cloud scalability is used to handle the growing workload where good performance is also needed to work efficiently with software or applications.
- Scalability is commonly used where the persistent deployment of resources is required to handle the workload statically.

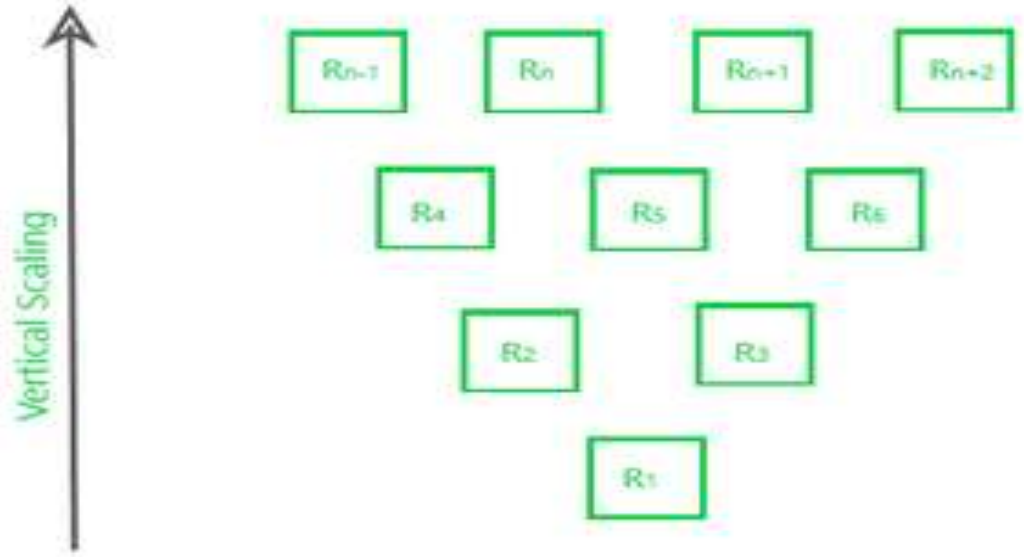
Example:

Consider you are the owner of a company whose database size was small in earlier days but as time passed your business does grow and the size of your database also increases, so in this case you just need to request your cloud service vendor to scale up your database capacity to handle a heavy workload

- Scalability is used to fulfill the static needs while elasticity is used to fulfill the dynamic need of the organization.
- Scalability is a similar kind of service provided by the cloud where the customers have to pay-per-use.
- So, in conclusion, we can say that Scalability is useful where the workload remains high and increases statically.

Types of Scalability:

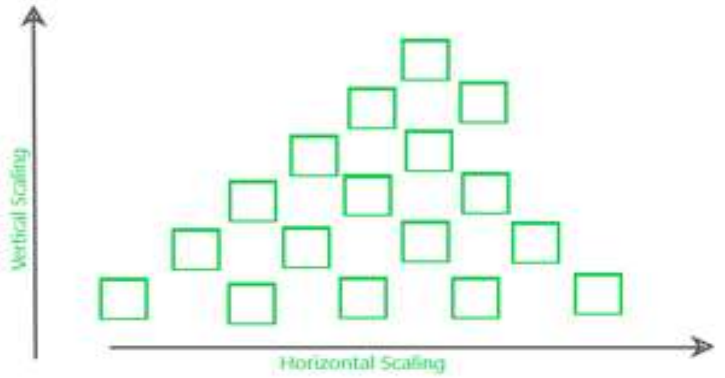
- **1. Vertical Scalability (Scale-up) –**
  In this type of scalability, we increase the power of existing resources in the working environment in an upward direction.



2.Horizontal Scalability: In this kind of scaling, the resources are added in a horizontal row.

3. Diagonal Scalability –

It is a mixture of both Horizontal and Vertical scalability where the resources are added both vertically and horizontally.



	Cloud Elasticity	Cloud Scalability
1.	Elasticity is used just to meet the sudden up and down in the workload for a small period of time.	Scalability is used to meet the static increase in the workload.
2.	Elasticity is used to meet dynamic changes, where the resources need can increase or decrease.	Scalability is always used to address the increase in workload in an organization.
3.	Elasticity is commonly used by small companies whose workload and demand increases only for a specific period of time.	Scalability is used by giant companies whose customer circle persistently grows in order to do the operations efficiently.
4.	It is a short term planning and adopted just to deal with an unexpected increase in demand or seasonal demands.	Scalability is a long term planning and adopted just to deal with an expected increase in demand.

On-Demand Provisioning in Cloud Computing

- On-demand provisioning in cloud computing refers to the ability to rapidly allocate and de-allocate computing resources as needed, without the need for long-term commitments or upfront investments.
- It is one of the fundamental characteristics of cloud computing, enabling organizations to scale their infrastructure based on demand and optimize resource utilization.

Key Concepts:

Resource Elasticity:

- On-demand provisioning allows users to easily scale up or down their resources, such as virtual machines, storage, and networking, according to changing workloads.
- This flexibility ensures that resources are available when needed and can be released when not in use, preventing unnecessary costs.

Pay-as-You-Go:

- With on-demand provisioning, users are typically billed based on actual usage, providing cost savings compared to traditional IT infrastructure where resources may be over-provisioned to handle peak loads.
- This pay-as-you-go model contributes to more efficient resource allocation.

Virtualization:

- Virtualization technologies underpin on-demand provisioning.
- Virtual machines (VMs) and containers enable the abstraction of physical hardware, making it possible to rapidly create, deploy, and manage computing instances.

Auto-scaling:

- Auto-scaling is a mechanism that allows cloud services to automatically adjust the number of resources based on predefined rules or metrics.
- For example, if a web application experiences a sudden increase in traffic, auto-scaling can spin up additional instances to handle the load and then scale down when the demand decreases.

Resource Orchestration:

- On-demand provisioning involves not only creating virtual instances but also managing them effectively.
- Resource orchestration tools and platforms help automate the provisioning, configuration, and management of resources, ensuring consistency and reducing manual intervention.

Agility:

- On-demand provisioning allows organizations to quickly respond to changing business requirements.
- New resources can be provisioned within minutes, enabling faster development, testing, and deployment cycles.

Reduced Management Overhead:

The automation and orchestration of resource provisioning reduce the manual effort required for managing infrastructure, leading to improved operational efficiency.

BENEFITS

Cost Efficiency:

- On-demand provisioning prevents over-provisioning of resources, reducing costs associated with idle resources.
- Organizations can allocate resources when they are needed and release them when they are not, leading to optimal resource utilization.

Scalability:

- Businesses can easily accommodate spikes in demand without the need for significant upfront investments.
- This scalability is especially valuable for applications that experience variable workloads.

CHALLENGES

Cost Management:

- While on-demand provisioning can save costs, improper management of resources can lead to unexpected expenses.
- Organizations need to monitor usage and optimize resource allocation continuously.

Security and Compliance:

- Rapid provisioning can potentially lead to security gaps if not properly managed.
- Ensuring that security measures and compliance requirements are consistently applied to all resources is crucial.

Vendor Lock-In:

Over-reliance on a specific cloud provider's services and proprietary technologies can limit the ability to migrate to another provider or deploy in a hybrid cloud environment.

Performance Variability:

In a multi-tenant environment, the performance of on-demand provisioned resources can be influenced by the activities of other tenants sharing the same physical infrastructure.