* Overview of Cloud Computing.

Cloud computing refers to the delivery of computing services - including servers, storage, databases, networking, software, analytics, and more over to the Internet also called cloud to offer faster innovation, flexible resources, and economies of scale.

Types of Cloud Computing

* Based On Deployment
  - ↳ Private
  - ↳ Public
  - ↳ Hybrid

* Based on Service Model
  - ↳ Infrastructure as a Service (Iaas)
  - ↳ Platform as a Service (Paas)
  - ↳ Software as a Service (Saas)

* Characteristics of Cloud Computing

1) On Demand Self Service

→ Users can provision computing resources, such as server time and storage, as needed without requiring human interaction with each service provider.

2) Broad Access Network

→ Services are available over the Internet and accessed through standard mechanisms that promote use of by heterogenous client platforms (e.g. mobile phones, tablets, laptops, workstations).

3) Resource Pooling

The providers computing resources are pooled to serve multiple consumers using a multi-tanent model, with different physical and virtual resources dynamically assigned and reassigned according to demand.

4) Rapid Elasticity

Resources are easily, rapidly and elastically provisioned and released to scale rapidly outward and inward with demand. To the consumer, the capabilities available for the provisioning often appear to be unlimited and can be purchased in any quantity at any time.

5) Measured Service

Resource usage can be monitored, controlled and reported, providing transparency for both the provider and the consumer of the utilized service.

* Benefits of Cloud Computing.

1) Cost Efficiency

Pay as you go pricing models and economics of scale can result in cost savings compared to traditional on premises infrastructure.

2) Scalability

Resources can be scaled up and down quickly and easily to accomodate changing demands

3) Flexibility

Users can access resources from anywhere with an internet connection and from any device.

4) Reliability

Cloud providers typically offer robust infrastructure and redundency to ensure high availability

* Challenges of Cloud Computing

1) Security Concerns

Cloud environments face security risks such as data breaches, cyber attacks, and unauthorized access due to shared infrastructure and potential vulnerabilities in cloud systems.

2) Data privacy and compliance

Cloud computing involves storing and processing data in shared environments, rising concerns about data privacy, regulatory compliance, and adherance to industry standards that vary across regions.

3) Vendor lock-in

Adopting specific cloud providers proprietary technologies and servies may lead to dependency, making it challenging and costly to migrate to alternative providers or deploy applications to premises

4) Downtime and Service outages

Despite efforts to ensure high availability, cloud services can experience downtime due to hardware failures, software bugs, maintenance activities, or external factors like network disruptions, impacting business continuity and user experience.

5) Performance and latency fluctuations

The performance of cloud applications and services may vary due to factors such as network latency, resource contention, and geographic distance between users and data centers, leading to inconsistent performs and user dissatisfaction.

* Cloud Storage

Cloud storage refers to the storage of data on remote servers accessed over the internet rather than on local storage devices like hard drives or servers. It allows users to store, access and manage data from anywhere with an internet connection.

* Key points in Cloud Storage

1) Scalability
2) Cost-effectiveness
3) Accessibility
4) Redundency and Reliability
5) Security
6) Backup and disaster recovery
7) Integration and Compatibility.

* Cloud Services Requirements

Cloud services have varying requirements depending on the specific service and its intended use. However, here are some common requirements for utilizing cloud services effectively.

1) Internet Connectivity
2) Compatible Devices and Browsers
3) Account and Credentials
4) Compliance and legal requirements
5) Security Measures
6) Resource allocation and Budgeting
7) Data Migration and Integration
8) Training and Support

* Cloud and Dynamic Infrastructure

* Cloud Infrastructure

Cloud Infrastructure refers to the collection of hardware, software and network resources that are provided as services over the Internet.

Instead of maintaining physical servers and infrastructure on premises, organizations can leverage cloud computing services to store data, run applications, and deliver various computing resources on demand.

Cloud Infrastructure is typically managed by cloud servile providers (e.g. Amazon Web Services, Microsoft, Azure, Google Cloud platform to run on a single physical server.)

* Dynamic Infrastructure

Dynamic Infrastructure refers to an IT infrastructure that can adapt and scale based on varying workload demands.

It leverages cloud computing principles and technologies to dynamically allocate computing resources, optimize performance, and meet changing requirem

The key features of dynamic Infrastructure

1) Scalability
2) Elasticity
3) Automation
4) Self-service Provisioning
5) Orchestration
   - ensures smooth interactions between different components of the infrastructure and helps automate complex workflows

* Cloud Reference Model

A cloud reference model is a conceptual framework that provides a structural approach for understanding and categorizing the components, layers, and interactions within a cloud computing environment. It serves as a reference guide for designing, implementing and managing cloud-based solutions.

* Infrastructure as a service (Iaas)

Infrastructure as a service (Iaas) is a cloud computing model that provides virtualized computing resources over the Internet, storage and networking resources on-demand. Here's an overview of Infrastructure as a service

* Key features :-

1) Virtualization

- Iaas provides use virtualization technologies to abstract physical hardware resources and create virtualized instances of servers, storage, and networking components, which can be dynamically provisioned and managed by users.

2) Scalability

Iaas platforms offer scalability, allowing users to scale resources up or down based on demand. Users can easily add or remove virtual machines, storage volumes, or network resources as needed to accomodate changing workloads.

3) Self-Service Provisioning

Iaas platforms provides self-service interfaces and APIs that allows users to provision and manage infrastructure resources autonomously, without requiring manual intervention from IT administrators.

4) Pay-Per-Use-Billing

Iaas follows a pay-per-use billing model, where users are billed based on their actual usage of resources, such as compute instances, storage capacity, and data transfer. This allows for a cost-effective resource utilization and eliminates upfront capital expenditures.

* Examples of Iaas Providers

- Amazon Web Services: Offers a comprehensive suite of Iaas services, including Amazon EC2, Amazon S3, and Amazon VPC.

- Microsoft Azure provides a wide range of Iaas offerings, such as Azure Virtual Machines, Azure Blob Storage and Azure Virtual Network.

- Google Cloud Platform offers Iaas services like Google Compute Engine, Google cloud Storage, etc.

* Platform as a Service (Paas)

Paas is a cloud computing model that provides a platform allowing customers to develop, deploy, and manage applications without dealing with the underlying infrastructure complexities.

* Key Features:-

1) Development Tools

Paas platforms typically include a range of development tools such as integrated development environments (IDEs), code editors, and version control systems to support application development.

2) Middleware

Paas offerings provide middleware services such as databases, messaging queues, caching and application servers, abstracting the complexity of managing these components.

3) Runtime Environments

Paas platforms offers runtime environments for executing applications, including support for multiple programming languages and framework

4) Monitoring and Management

Paas platforms offers tools for monitoring applications performance, managing deployment and troubleshooting issues, simplifying operational tasks for developers and administrators.

5) Scalability and Elasticity

Paas platforms enables automatic scaling of resources to handle fluctuations in application demand, ensuring optimal performance and cost efficiency.

* Examples of Paas Platforms

- Microsoft Azure App service offers a fully managed platform for building, deploying and scaling web and mobile applications.

- Google App Engine - Provides a platform for developing and hosting web applications using Google's infrastructure, with support for multiple programming languages.

- Heroku : A cloud platform that enables developers to deploy, manage and scale applications written in various programming languages, including Ruby, Python and Node.js.

* Software as a Service (Saas)

Saas is a service based model in cloud computing which allows users to access and use software applications through a web browser or dedicated client applicatio without needing to install or maintain the software locally on their devices. The software is centrally hosted and managed by the saas provider, who handles maintenance, updates and security.

* Key features

1) Accessibility

Saas features are accessible from any device with an internet connection, enabling users to access their data and applications from anywhere, at any time.

2) Subscription - based Pricing

- Saas offers subscription based pricing model, where users pay a reurring fee based on usage or number of users. This eliminates the need for upfront software licensing costs and allows for predictable budgeting.

3) Automatic updates and maintenance

Saas providers handle software updates, patches, and maintenance tasks, ensuring that users always have access to the latest features and security enhancements without needing to perform manual updates.

4) Scalability

Saas applications are designed to scale effortlessly to accomodate growing user bases and increasing workloads. Users can easily add or remove users, upgrade plans or adjust resources.

5) Multi-tenancy

Saas applications typically follow a multi-tenant architecture, where multiple users or organizations share a single instance of the software while maintaining data isolation and security.

6) Customization and Integration

Saas applications may offer customization options and integration capabilities to tailor the software to specific business requirements and integrate with other cloud-service provider / services.

* Examples of Saas Applications:

1) Salesforce is a leading Customer Relationship Management (CRM) platform that helps businesses manage sales, marketing, and customer service operations

2) Microsoft Office 365 is a productivity tool including Word, Excel, PowerPoint, and Outlook, delivered as a service with collaboration and communication features.

3) Google Workspace (formally G Suite) :-
A collection of cloud based productivity and collaboration tool including gmail, google drive, google docs, google meet, etc.

* Cloud Deployment Models

Cloud Deployment Models refers to the different ways in which cloud computing services can be deployed to meet the needs of users and organizations. There are four main cloud deployment models:-

1) Public Cloud

- In the public cloud deployment model, cloud services and resources are provided by the third party cloud servie providers over the public internet.

- These services are available to multiple users and organizations on a pay-as-you-go basis, allowing users to access computing resources, such as virtual machines, storage and applications without upfront investment in hardware

- Public cloud providers own and operate the underlying infrastructure and are responsible for managing and maintaining the hardware, software and network components.

- Examples of public cloud providers include Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP).

* Advantages	Disadvantages
1) High Scalability	1) loss of control over data
2) Cost Reduction	2) Data security & privacy
3) Reliability & Flexibility	3) Limited visibility
4) Disaster Recovery	4) Unpredictable cost

2) Private Cloud

- In the private cloud, cloud services and resource are provisioned and hosted on dedicated infrastructure that is exclusively used by a single organization.

- Private clouds can be deployed on-premises within an organization's data center or hosted by a third party cloud provider. They offer greater control, security, and customization compared to public clouds.

- Private clouds environments are ideal for organizations with strict security and compliance requirements, sensitive data or specific performance and customization needs.

- Examples of private cloud solutions include VMware vSphere, Openstack, and Microsoft Azure Stack.

* Advantages of Private Cloud

- Customer Information Protection
- Infrastructure Ensuring SLAs
- Compliance with standard procedures and guidelines

↳ SLAs (Service level Agreements) refer to the specific commitments or guarantees made by the cloud provider regarding the performance and availability of the services offered within the private cloud infrastructure.

* Disadvantages of using Private Clouds

- The restricted area of operations
- Expertise requirement.

* Hybrid Cloud

- The hybrid cloud deployment model combines elements of both public and private clouds, allowing organizations to leverage resources and services from multiple cloud environments.

- In the hybrid cloud setup, workloads can be distributed between on-premises infrastructure, private clouds, and public clouds based on factors such as performance, security, compliance, and cost.

- Hybrid cloud architectures enable seamless integration and workload probability between different cloud environments, providing flexibility and scalability to meet changing busmess requirements.

- Examples of hybrid cloud computing solutions include AWS outputs, Azure Hybrid Benefits, and Google Authos.

* Advantages of using a Hybrid Cloud

1) Cost - available at cheap cost
2) Speed - reduces the latency of data transfer
3) Security - totally safe and secure because it works on the distributed system network

* Disadvantages of using a Hybrid Cloud

- Managing security can be challenging
- Requires intensive knowledge of the system

* Community Clouds

- The community cloud deployment model involves sharing cloud resources and servies among multiple organizations with common interests, requirements or compliance obligations.

- Community clouds are typically hosted and managed by a third-party service provider and serve specific industries, such as healthcare, finance or government.

- Organizations in the community cloud share the costs, risks and benefits of the cloud infrastructure while maintaining data isolation and security.

- Example of Cloud Computing can be FedRAMP.

* Advantages of Cloud Computing Community

- Cost effective
- Adaptable and Scalable
- Sharing of resoure between different enterprises.

* Disadvantages of Community cloud Computing

- Not all businesses choose community cloud
- Gradual adoption of data
- Challenges for corporations to share duties.

* Virtualization

- Virtualization is the process of creating a virtual version of something, such as a operating system, a server, a storage device, or network resources.

- It allows multiple systems to run on a single physical machine, by abstracting the underlying hardware and presenting it as one or more "virtual" devices.

- This enables IT organizations to maximize the utilization of resources, reduce costs, and increase flexibility and agility.

- The machine on which the virtual machine is going to create is known as Host Machine and that virtual machine is refered as a guest machine.

* Advantage of Virtualization

1) Resource Utilization

Virtualization allows effecient utilization of hardware resources by enabling multiple virtual machines (VMs) to run on a single physical server. This reduces hardware costs and maximize resource utilization.

2) Cost Savings

By consolidating multiple servers onto a single physical machine, organization can save on hardware, maintenance and energy costs. Additionally, virtualization reduces the need for physical space, cooling and power consumption in data science.

3) Scalability

Virtualization provides the flexibility to scale IT infrastructure up or down rapidly in response to changing business needs. New virtual machines can be added quickly without purchasing

4) Improved Disaster Recovery

Virtualization enables easy backup, replication, and migration of virtual machines, making disaster recovery processes more efficient.

5) Enhanced Testing and Development

Virtualization allows developers and testers to create isolated environments for software development, testing and debugging without affecting the production environment.

6) Centralized Management

Virtualization platforms often include centralized management tools that streamline the administration of virtualized infrastructure. This simplifies tasks such as deployment, monitoring and resource allocation.

* Types of Virtualization

1) Server Virtualization

- This involves creating one or more virtual servers on a single physical machine.

- Each virtual server runs its own operating system and can have its own set of applications and servers.

- This allows multiple servers to run on a single piece of hardware, reducing the need for physical servers and associated costs.

- Popular server virtualization software includes VMware, vSphere, Microsoft Hyper-V and Xen.

2) Desktop Virtualization

- This enables users to access their desktop environment and applications from anywhere, using any device that has a remote-access client installed.

- The desktop environment is hosted on a virtual machine, which runs on a remote server, and the user interacts with it using a remote-access protocol such as Remote Desktop Protocol (RDP) or Virtual Networking Computing (VNC).

- Citrix XenDesktop and VMware Horizon are examples of desktop virtualization software.

3) Application Virtualization

- This allows multiple versions of an application to run on the same system, without conflicts. Applications are encapsulated in a virtual environment and are isolated from the underlying operating system.

- This allows different versions of the same applications to be used by different users, or even by the same user, or the same system.

- Examples of application virtualization software include Microsoft App-V, Citrix XenApp and VMware ThinApp.

4) Storage Virtualization

- This abstracts the physical storage devices, such as hard drives and storage area networks (SAN's), and presents them as a single logical storage device.

- This makes it easy to manage and allocate storage resources, and to move data between physical storage devices.

- Storage Virtualization is the practice of grouping physical storage servers from multiple network storage devices into a single virtual storage device managed from a central console.

- Storage virtualization can be done in software (e.x. Openfiler) or in hardware (ex: Fibre Channel SAN).

5) Network Virtualization

- This creates a virtualized version of a network, allowing multiple virtual networks to coexist on the same physical infrastructure.

- Each virtual network can be configured and managed independently of the others, allowing different network environments to be created for different purpose.

- There are various technologies and approaches used for network virtualization, including virtual LAN's (VLANs), virtual private networks (VPNs), and software-defined networking (SDN)

6) Hardware Virtualization

- Hardware Virtualization is accomplished by abstracting the physical hardware layer by use of a hypervisor or VMM (Virtual Machine Manager).

- When the virtual machine software or virtual machine manager (VMM) or hypervisor software is directly installed on the hardware system is known as hardware virtualization.

- The main job of hypervisor is to control and monitoring the processor, memory and other hardware resources.

- Hardware virtualization is mainly done for the server platforms, because controlling virtual machines is much easier than controlling a physical server.

*Implementation levels of Virtualization.

1) Instruction Set Architecture level

- At the ISA level, virtualization is performed by emulating a given ISA by the ISA of the host machine.

- For example, MIPS binary code can run on x86-based host machine with the help of ISA emulation.

- With this approach, it is possible to run on a large amount of legacy binary code written to a various process on any given new hardware host machine.

2) Hardware Abstraction level

- Hardware level virtualization is performed right on top of the bare hardware.

- On the other hand, this approach generates a virtual hardware environment for a VM.

- On the one hand, the process manages the underlying hardware through virtualization

- The idea is to virtualize a computer's resources, such as its processors, memory and I/O devices

3) Operating System Level

- This refers to an abstraction layer between traditional OS and user applications.

- To allocate hardware resources among a large number of mutually distrusting users, OS-level virtualization is often utilized in the creation of virtual environments.

- Every user gets their environment with their virtual hardware resources.

4) Library Support level

- Instead of the OS's long system calls, most programs use APIs revealed by user-level libraries as most systems have well documented APIs.

- Virtualization using library interfaces is achieved by using API hooks to regulate communicate channel

5) User - Application level

- When you simply want to virtualize a single program, application virtualization needs to be considered.

- In this case, the entire environment of the platform doesnot need to be virtualized

- As applications runs on a single process on a computer's operating system this can be referred to as a process level virtualization.

- The most popular approach is to deploy high level language (HLL).

* Virtual Infrastructure Requirements

1) Hardware

- A virtualize infrastructure requires a set of physical servers and other hardware components such as storage devices and networks switches, to host the virtual machines and provide the underlying resources.

- These servers should have enough CPU, memory and storage capacity to meet the needs of the virtual machines running on them.

2) Hypervisor

- A virtualized infrastructure requires a hypervisor to manage and run the virtual machines. The hypervisor sits on top of the physical servers and abstracts the underlying hardware resources.

3) Virtual Machines

- Virtual Machines are the foundations of a virtualized infrastructure and are used to host the various applications and servies that make up the infrastructure.

- Virtual Machines can be created, started, stopped, moved, and deleted on-demand, as needed.

4) Virtualized Storage

- A virtualized infrastructure requires storage resources to store the virtual machines to each other and to external networks.

- This can be done using software-based or hardware-based network virtualization solutions.

5) Security

- Virtualized infrastructure requires additional security measures to protect the virtual machines and data from unauthorised access and attacks.

- This includes implementing security best practices and implementing security solutions such as firewalls, intrusion/prevention systems and encryption technologies.

6) Monitoring and monitoring tools

- To ensure availability, reliability and stability of the virtualized infrastructure it is important to monitor it regularly and be able to troubleshoot in case of any issue.

* Service Oriented Architecture (SOA)

↳ SOA is essentially a collection of service, that are provided to end users to make application through Internet.

↳ These services are communicated each other. The communication can involve either simple data passing or it could involve two or more services coordinating some activity.

↳ A cloud has many characteristics: Elasticity self-service provisioning and pay as you go. These types of functionalities has to be engineered into software. Service Oriented Architecture helps to make this possible.

↳ SOA is much more than a technological approach and methodology for creating IT system. It's also a business methodology.

* Advantages of SOA are:-

i) Service reusability

Single service can be used by different users in same or different time & location.

ii) Easy Maintenance

Services are independent to each other, so it is easy to maintain, easy to solve it.

iii) Promote Interaction

With Service oriented architecture (SOA), a standard form of communication take place that helps to promote interaction.

iv) Platform Independent

There are many services which are independent of each other. So, customers can work from anywhere by taking any services.

v) Reliability
SOA services are complete and self-containing program

vi) Reduce Cost

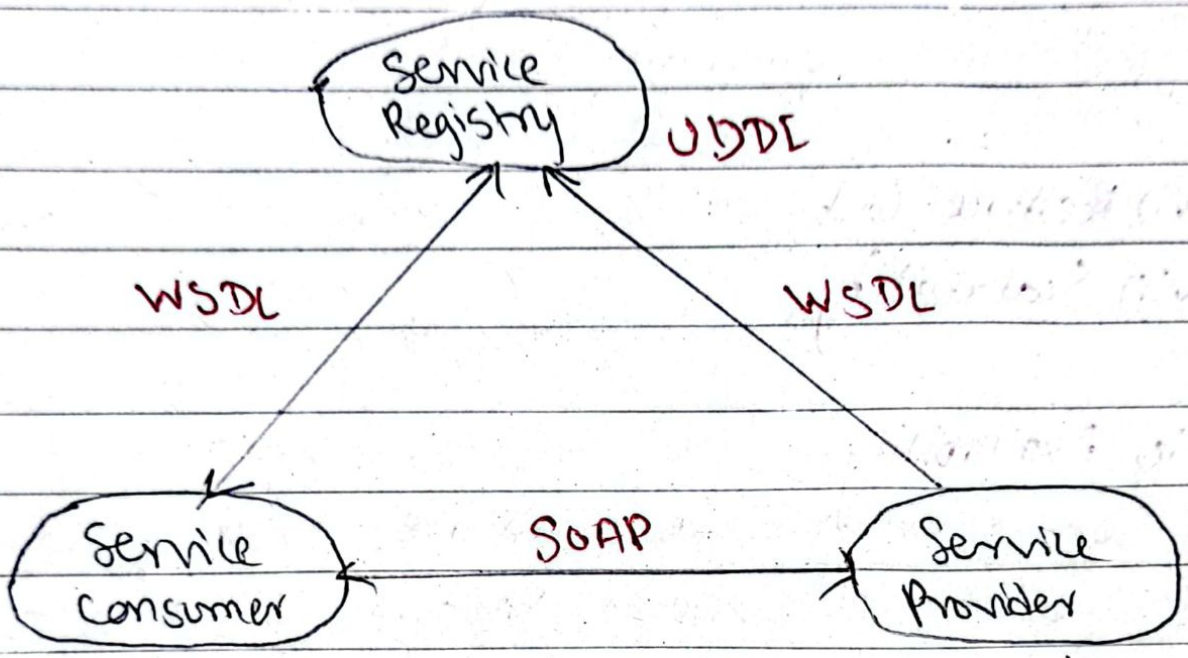
vii) Scalability

viii) Availability
SOA services are available to any requester that needs them.

ix) High quality Services

x) Increased Productivity

* Components of SOA



- UDDI → Universal Description, Discovery and Integration
- SOAP → Simple Object Access Protocol
- WSDL → Web Services Description Language

* Cloud Security Challenges & Risks:-

1) Data loss

Storing data on someone else's servers can raise concerns about it's safety, leading to potential data breaches.

2) Downtime

If the cloud servile goes down, your access to data and applications could be interrupted, causing business disruptions.

3) Compliance

Ensuring that your cloud usage complies with legal and regulatory requirements can be complex and failing to do so can lead to legal consequences.

4) Data Integration

Integration data from different cloud services can be challenging and may lead to data inconsistency or loss.

* Software-as-a-service Security

1) Data Encryption
Sensetive data should be encrypted both in transit and at rest to prevent unauthorized access.

2) Identity and Access Management
Implement strong authentication mechanisms like multi-factor authentication (MFA) to ensure only authorised users can access the Saas application.

3) Regular Updates and Patch Management
Ensure the Saas provider regularly updates and patches their software to address vulnerabilities and mitigate potential security risks.

4) Data Segregation
Ensure that eden customer's data is logically segregated from others to prevent unauthorised access.

5) Security Monitoring and Logging
Implement robust monitoring and logging mechanisms to detect and respond to security incidents in real-time.

6) Compliance and Auditing
Ensure that SaaS provider complies with relevant security standards and regulations, and conduct regular security audits to assess compliance.

7) Backup and Disaster Recovery
Implement backup and disaster recovery plans to ensure data availability and integrity in case of unexpected events or outages.

8) Vendor Security Assessment
Before selecting a SaaS provider, perform a thorough security assessment of their infrastructure, policies and practices to ensure they meet your organizations security

* Security Monitoring in Cloud Computing

Security Monitoring in Cloud Computing involves the continuous observation of cloud resources, services and data to detect and respond to security threats and vulnerabilities.

1) Real Time Monitoring

Security Monitoring Tools continuously monitor cloud environments in real-time, analyzing logs, events and network traffics for signals of suspicious activity.

2) Alerting and Notification

When potential security threats or anomalies are detected, security monitoring tools generate alerts and notifications to inform security teams so they can investigate and respond promptly.

3) Logging and Auditing

Cloud service providers typically offers logging and auditing features that record activity and events within the cloud environment. Security monitoring leverages these logs to track user actions, system changes, and access attempts for security analysis and compliance purposes.

4) Behavioral Analysis

System monitoring tools ~~are~~ use behavioral analysis tools to establish a baseline of normal behaviour for cloud resources and users. Deviations from this baseline can indicate potential security incidents, such as unauthorised access or malicious activity.

5) Threat Intelligence Integration

Security monitoring solutions often integrate with threat intelligence feeds to enhance detection capabilities.

* Server Virtualization

Server virtualization is a technology that allows multiple virtual instances of operating systems (OS) to run on a single physical server. Instead of each physical server running a single OS, virtualization enables the creation of multiple virtual machines (VMs) on a single physical server, each running its own OS and applications.

* Types of Server Virtualization

1) Full Virtualization

In full virtualization, a hypervisor is used to create and manage multiple VMs on a physical server. Each VM runs its own complete OS, which is unaware that its running in virtualized environment. Examples of hypervisors for full virtualization include VMware vSphere / ESXi, Microsoft Hyper-V, and KVM (Kernel Based Virtual Machine).

2) Para Virtualization

Unlike full virtualization, para virtualization requires modifications to the guest OS to be aware of the virtualization layer. This allows for more efficient communication between the guest OS and the hypervisor, leading to better performance compared to full virtualization. Xen is an example of a para virtualization.

3) Hardware Assisted Virtualization

This type of virtualization takes advantage of hardware features to improve virtualization performance. Processors with virtualization extensions (such as Intel VT-x or AMD-V) allow the hypervisor to run more efficiently, reducing overhead.

4) Container-based virtualization

Containers provide a lightweight form of virtualization where applications and their dependencies are packed together. Unlike

traditional VMs, container share the host OS kernel, which makes them more lightweight and faster to start up. Docker is a popular tool for containerization.

* Advantages of Virtualization

1) Resource Utilization

Virtualization allows for better utilization of physical server resources by running multiple VMs on a single server.

2) Isolation

Virtualization provides isolation between VMs, which enhances security and reduces risks of one application or OS application or OS affecting others on the same physical server.

3) Flexibility and Scalability

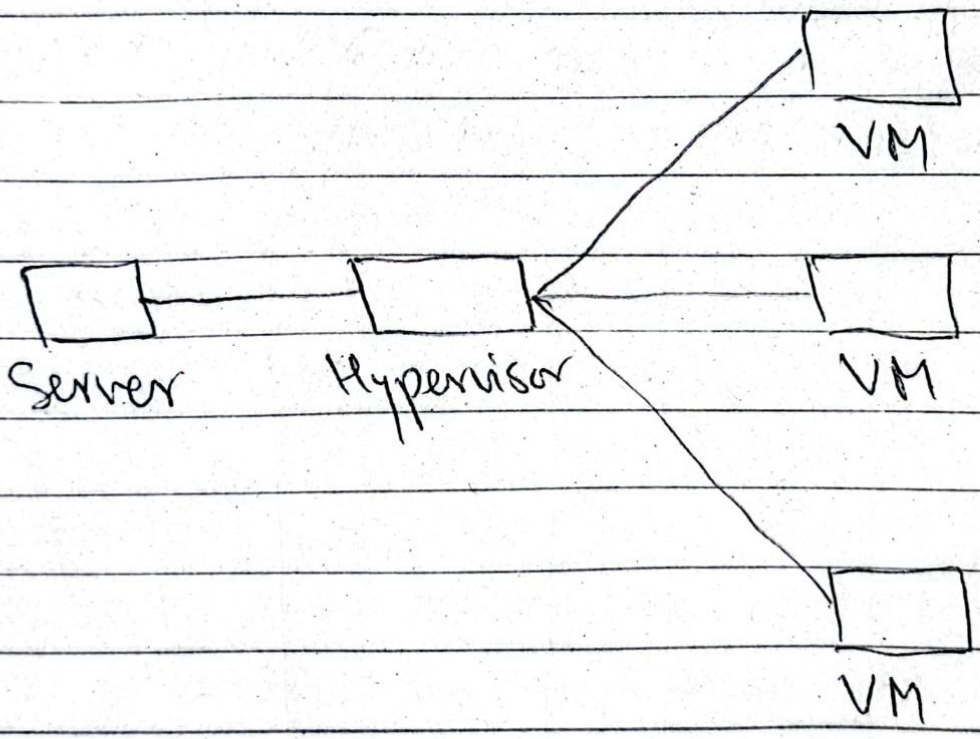
Virtualized environments are highly flexible and scalable. VMs can be easily provisioned, cloned or moved between physical servers,

allowing for dynamic allocation of resources based on demand.

4) Disaster recovery and high availability
Virtualization enables easy backup, replication and migration of VMs, making disaster recovery and high availability strategies more feasible and efficient.

★ Hypervisor

A hypervisor is a piece of hardware, software or firmware that can create virtual computers, manage them, and allocate resources to them.



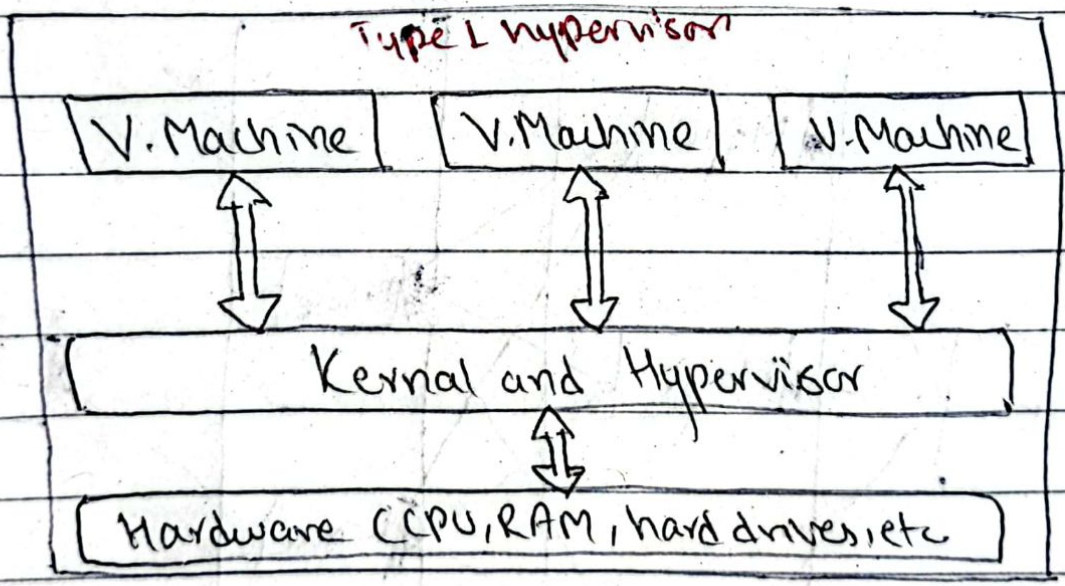
- Hypervisor is also known as virtual machine monitor or VMM, is a software that creates and manages virtual machines (VMs).

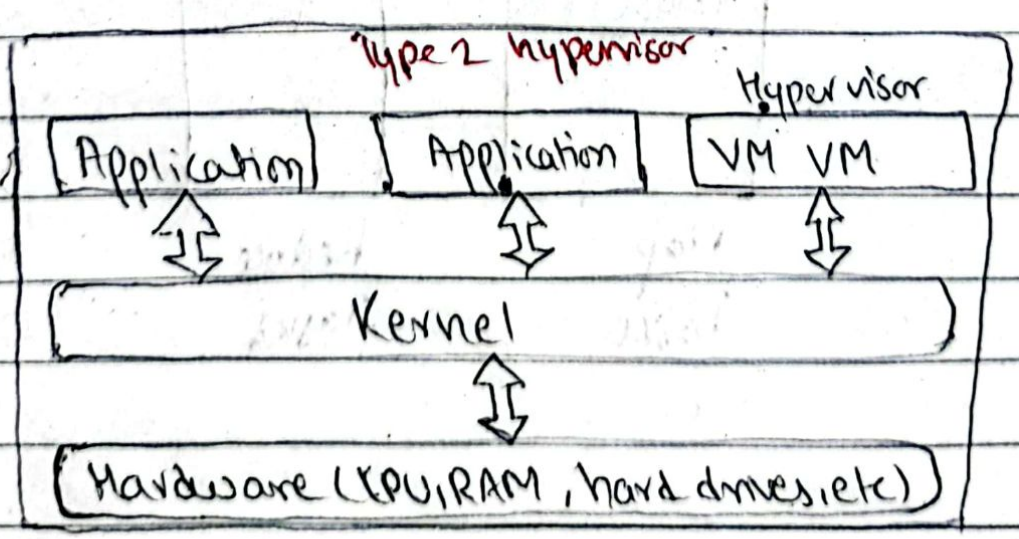
- Types of Hypervisors

1) Type 1 (Bare metal) hypervisor 2) Type 2 (Hosted) hypervisor

1) Type 1 hypervisor (Bare metal hypervisor)
2) Type 2 hypervisor (Hosted hypervisor)

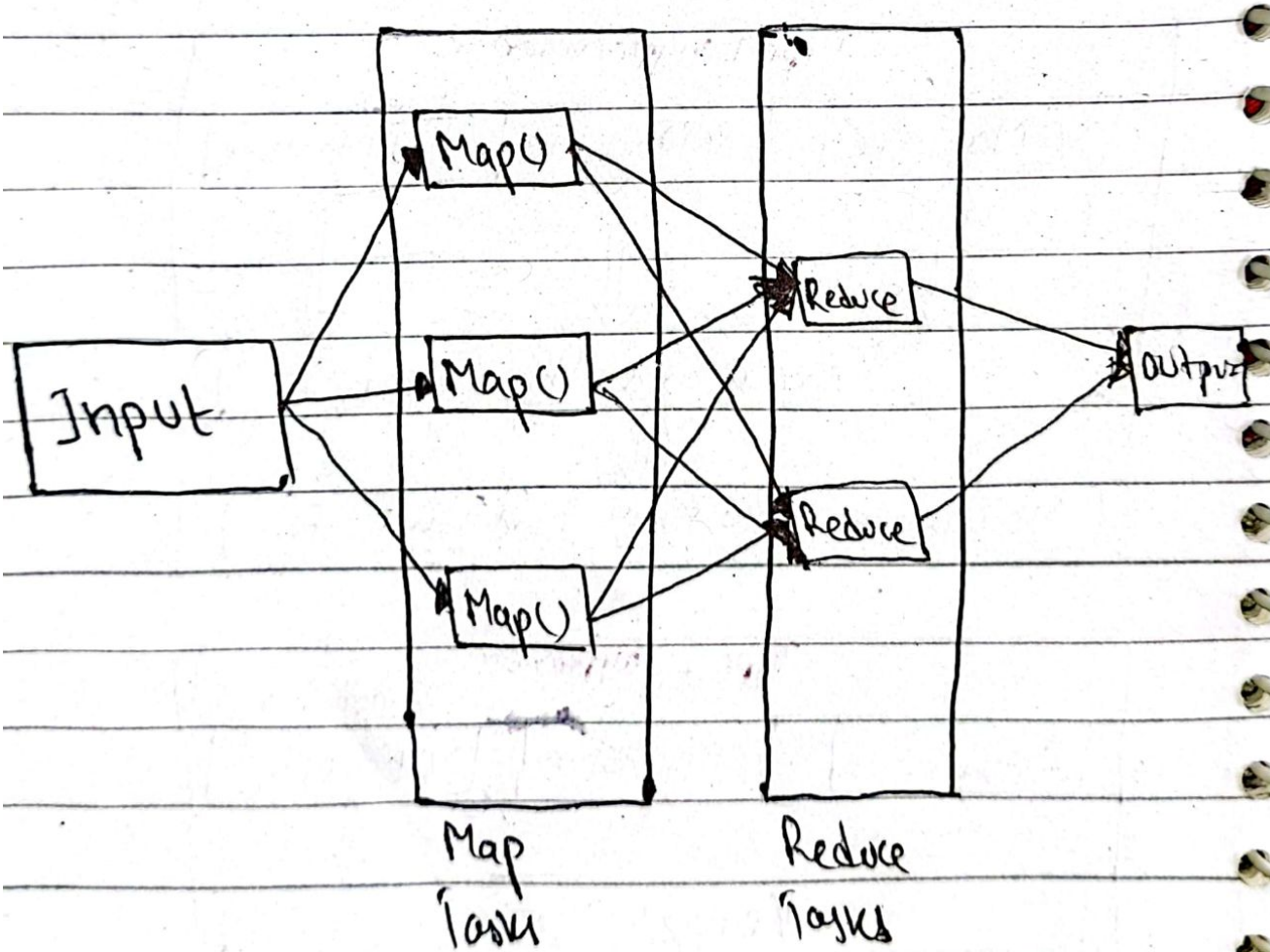
A type 1 hypervisor acts as a lightweight operating system and runs directly on the host's hardware, while type 2 hypervisor runs as a software layer on an operating system, like other computer programs.





* Map Reduce

Map Reduce is a framework using which we can write applications to process large amounts of data, in parallel, on large clusters of commodity hardware in a reliable manner.



*Phases of MapReduce programming model

1) Map Phase

In this phase, the input data is divided into smaller chunks and distributed across the nodes in the cluster.

Each node processes its chunk independently by applying a specified function called the "mapper" to the input data.

The mapper function produces intermediate key-value pair.

2) Shuffle Phase

In this phase, the intermediate key-value pairs generated by the mappers are shuffled and sorted based on their keys.

The goal is to gather all values associated with the same key across different nodes.

3) Reduce Phase

In this phase, each node processes a subset of key-value pairs.

The reducer applies a specified function, called the "reducer", to the values associated with each unique key.

The reducer function typically aggregates or combines these values to produce the final output.

* Applications of Map Reduce

1) Data Analysis and Algorithm

Map Reduce is often used for processing and analyzing large volumes of data to perform tasks such as data aggregation, filtering, sorting and summarization.

2) Log Processing and Analysis

Map Reduce is frequently employed for analyzing log data generated by web servers, applications or network devices.

3) Text Processing and Mining

Map Reduce is used for processing and analyzing large collections of text data, such as documents, web pages or social media.

4) Recommendation Systems

Map Reduce is employed in building recommendation systems for personalized content delivery in e-commerce, streaming platforms & social networks.