# Virtualization in Cloud Computing and Types

Virtualization is a technique of how to separate a service from the underlying physical delivery of that service.

It is the process of creating a virtual version of something like computer hardware.

It involves using specialized software to create a virtual or software-created version of a computing resource rather than the actual version of the same resource.

With the help of Virtualization, multiple operating systems and applications can run on same machine and its same hardware at the same time, increasing the utilization and flexibility of hardware.

One of the main cost effective, hardware reducing, and energy saving techniques used by cloud providers is virtualization.

Virtualization allows to share a single physical instance of a resource or an application among multiple customers and organizations at one time.

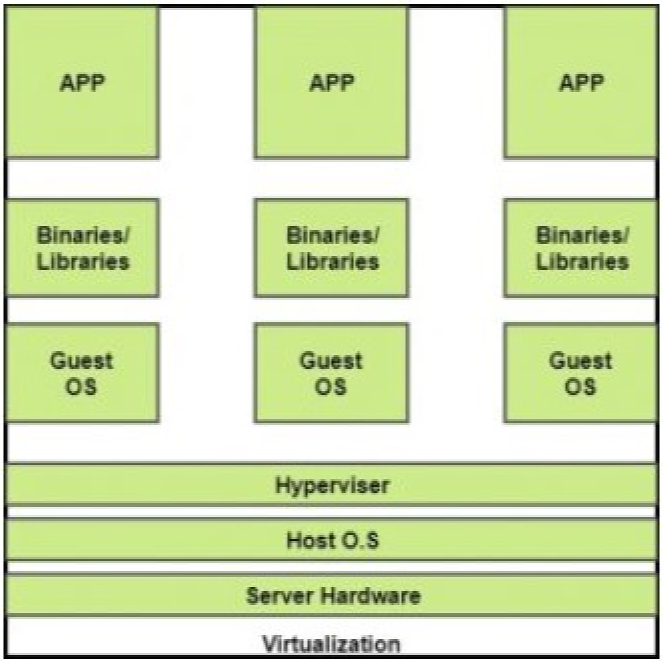
It does this by assigning a logical name to a physical storage and providing a pointer to that physical resource on demand.

The term virtualization is often synonymous with hardware virtualization, which plays a fundamental role in efficiently delivering Infrastructure-as-a-Service (IaaS) solutions for cloud computing.

Moreover, virtualization technologies provide a virtual environment for not only executing applications but also for storage, memory, and networking.

**BENEFITS OF VIRTUALIZATION**

1. More flexible and efficient allocation of resources.
2. Enhance development productivity.
3. It lowers the cost of IT infrastructure.
4. Remote access and rapid scalability.
5. High availability and disaster recovery.
6. Pay per use of the IT infrastructure on demand.
7. Enables running multiple operating systems.



The machine on which the virtual machine is going to be built is known as Host Machine and that virtual machine is referred as a Guest Machine.

Types of Virtualization:

1.Application Virtualization

2.Network Virtualization

3.Desktop Virtualization

4.Storage Virtualization

5.Server Virtualization

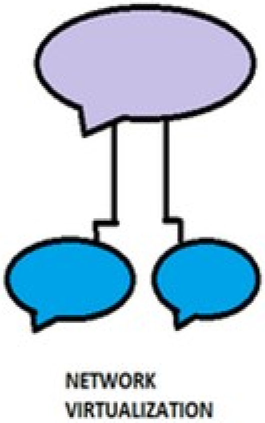
6.Data virtualization

## Types of Virtulisation

### Application Virtualization:

Application virtualization helps a user to have remote access of an application from a server. The server stores all personal information and other characteristics of the application but can still run on a local workstation through the internet. Example of this would be a user who needs to run two different versions of the same software. Technologies that use application virtualization are hosted applications (delivers the app's assets on demand from a web server) and packaged applications (delivers all its content onto a user's device).

### Network Virtualization:



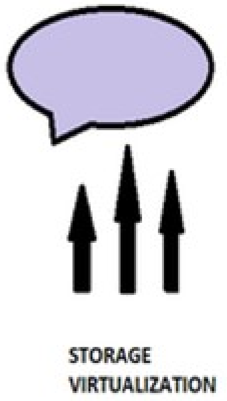
The ability to run multiple virtual networks with each has a separate control and data plan. It co-exists together on top of one physical network. It can be managed by individual parties that are potentially confidential to each other.

Network virtualization provides a facility to create and provision virtual networks—logical switches, routers, firewalls, load balancer, Virtual Private Network (VPN), and workload security.

**Desktop Virtualization:**

Desktop virtualization allows the users' OS to be remotely stored on a server in the data centre. It allows the user to access their desktop virtually, from any location by a different machine. Main benefits of desktop virtualization are user mobility, portability, easy management of software installation, updates, and patches.

**Storage Virtualization:**



Storage virtualization is an array of servers that are managed by a virtual storage system. The servers aren't aware of exactly where their data is stored, and instead function more like worker bees in a hive. It makes managing storage from multiple sources to be managed and utilized as a single repository. Storage virtualization software maintains smooth operations, consistent performance and a continuous suite of advanced functions despite changes, break down and differences in the underlying equipment.

**Server Virtualization:**

This is a kind of virtualization in which masking of server resources takes place. Here, the central-server (physical server) is divided into multiple different virtual servers by changing the identity numbers. So, each system can operate its own operating systems in isolate manner. Where each sub-server knows the identity of the central server. It causes an increase in the performance and reduces the operating cost by the deployment of main server resources into a sub-server resource. It's beneficial in virtual migration, reduce energy consumption, reduce infrastructural cost, etc.

**Data virtualization:**

This is the kind of virtualization in which the data is collected from various sources and managed that at a single place without knowing more about the technical information like how data is collected, stored & formatted, arranged that data logically so that its virtual view can be accessed by its interested people and stakeholders, and users through the various cloud services remotely. Many big giant companies are providing their services like Oracle, IBM, At scale, Cdata, etc.

It can be used to perform various kind of tasks such as:

- Data-integration
- Business-integration
- Service-oriented architecture data-services
- Searching organizational data

---

Virtualization has many practical applications, with two main ones.

- Application testing: Virtualization allows software developers to test their applications in various environments without setting up several different computers. If the application crashes on the virtual machine, they can close and restart the virtual machine to a previous state without damaging the computer.

- Server consolidation: One of the most significant benefits of virtualization is server consolidation. Instead of maintaining multiple servers with unique functions, server virtualization allows you to split a single server's resources for multiple purposes. Often, server resources are underutilized, resulting in businesses spending too much on server upkeep for a little output.

---

**What is a Hypervisor?**

- Software installed on top of hardware that created virtualization layer
- Hosts VMs

- Type 1 Hypervisor – Bare metal hypervisor (VMware ESXi)
- Type 2 Hypervisor – Hosted hypervisor (VMware Workstation)

A **type 1 hypervisor** acts like a lightweight operating system and runs directly on the host's hardware.

VMware ESXi is an enterprise-class, type-1 hypervisor developed by VMware for deploying and serving virtual computers.

As a type-1 hypervisor, ESXi is not a software application that is installed on an operating system; instead, it includes and integrates vital OS components, such as a kernel.

The most commonly deployed type of hypervisor is the type 1 or bare-metal hypervisor, where virtualization software is installed directly on the hardware where the operating system is normally installed.

Because bare-metal hypervisors are isolated from the attack-prone operating system, they are extremely secure.

In addition, they generally perform better and more efficiently than hosted hypervisors. For these reasons, most enterprise companies choose bare-metal hypervisors for data center computing needs.

A **type 2 hypervisor** runs as a software layer on an operating system, like other computer programs.

Hosted hypervisors run on top of the operating system (OS) of the host machine. Additional (and different) operating systems can be installed on top of the hypervisor.

The downside of hosted hypervisors is that latency is higher than bare-metal hypervisors. This is because communication between the hardware and the hypervisor must pass through the extra layer of the OS.

Hosted hypervisors are sometimes known as client hypervisors because they are most often used with end users and software testing, where higher latency is less of a concern.

Both types of hypervisors can run multiple virtual servers for multiple tenants on one physical machine.

**How does a hypervisor work?**

Hypervisors support the creation and management of virtual machines (VMs) by abstracting a computer's software from its hardware. Hypervisors make virtualization possible by translating requests between the physical and virtual resources.

Bare-metal hypervisors are sometimes embedded into the firmware at the same level as the motherboard basic input/output system (BIOS) to enable the operating system on a computer to access and use virtualization software.

## Container vs hypervisor

Containers and hypervisors are both involved in making applications faster and more efficient, but they achieve this in different ways.

Hypervisors:

- Allow an operating system to run independently from the underlying hardware through the use of virtual machines.
- Share virtual computing, storage and memory resources.
- Can run multiple operating systems on top of one server (bare-metal hypervisor) or installed on top of one standard operating system and isolated from it (hosted hypervisor).

Containers:

- *Containers* are packages of software that contain all of the necessary elements to run in any environment.
- Allow applications to run independently of an operating system.
- Can run on any operating system—all they need is a container engine to run.
- Are extremely portable since in a container, an application has everything it needs to run.

Hypervisors are used to create and run virtual machines (VMs), which each have their own complete operating systems, securely isolated from the others.

In contrast to VMs, containers package up just an app and its related services. This makes them more lightweight and portable than VMs, so they are often used for fast and flexible application development and movement.

---

## Load Balancing

Cloud load balancing is defined as the method of splitting workloads and computing properties in a cloud computing.

It enables enterprise to manage workload demands or application demands by distributing resources among numerous computers, networks or servers. It also raises reliability through redundancy.

Cloud load balancing **takes a software-based approach to distributing network traffic across resources**, as opposed to hardware-based load balancing, which is more common in enterprise data centers. A load balancer receives incoming traffic and routes those requests to active targets based on a configured policy.

There are two elementary solutions to overcome the problem of overloading on the servers-

First is a single-server solution in which the server is upgraded to a higher performance server. However, the new server may also be overloaded soon, demanding another upgrade. Moreover, the upgrading process is difficult and expensive.

Second is a multiple-server solution in which a scalable service system on a cluster of servers is built. That's why it is more cost effective as well as more scalable to build a server cluster system for network services.

Cloud-based servers farms can attain more precise scalability and availability using server load balancing.

Load balancing solutions can be categorized into two types –

- Software-based load balancers: Software-based load balancers run on standard hardware (desktop, PCs) and standard operating systems.

- Hardware-based load balancer: Hardware-based load balancers are dedicated boxes which include Application Specific Integrated Circuits (ASICs) adapted for a particular use. ASICs allows high speed promoting of network traffic and are frequently used for transport-level load balancing because hardware-based load balancing is faster in comparison to software solution.

## Major Examples of Load Balancers –

### Direct Routing Requesting Dispatching Technique:

This approach of request dispatching is like to the one implemented in IBM's Net Dispatcher.

A real server and load balancer share the virtual IP address. In this, load balancer takes an interface constructed with the virtual IP address that accepts request packets and it directly routes the packet to the selected servers.

### Dispatcher-Based Load Balancing Cluster:

A dispatcher does smart load balancing by utilizing server availability, workload, capability and other user-defined criteria to regulate where to send a TCP/IP request.

The dispatcher module of a load balancer can split HTTP requests among various nodes in a cluster. The dispatcher splits the load among many servers in a cluster so the services of various nodes seem like a virtual service on an only IP address; consumers interrelate as if it were a solo server, without having an information about the back-end infrastructure.

**Linux Virtual Load Balancer:**

It is an open source enhanced load balancing solution used to build extremely scalable and extremely available network services such as HTTP, POP3, FTP, SMTP, media and caching and Voice Over Internet Protocol (VoIP).

It is simple and powerful product made for load balancing and fail-over. The load balancer itself is the primary entry point of server cluster systems and can execute Internet Protocol Virtual Server (IPVS), which implements transport-layer load balancing in the Linux kernel also known as Layer-4 switching.

## Different Types of Load Balancing Algorithms in Cloud Computing:

### Static Algorithm

Static algorithms are built for systems with very little variation in load. The entire traffic is divided equally between the servers in the static algorithm. This algorithm requires in-depth knowledge of server resources for better performance of the processor, which is determined at the beginning of the implementation.

However, the decision of load shifting does not depend on the current state of the system.

### Dynamic Algorithm

The dynamic algorithm first finds the lightest server in the entire network and gives it priority for load balancing. This requires real-time communication with the network which can help increase the system's traffic. Here, the current state of the system is used to control the load.

The characteristic of dynamic algorithms is to make load transfer decisions in the current system state. In this system, processes can move from a highly used machine to an underutilized machine in real time.

### Round Robin Algorithm

As the name suggests, round robin load balancing algorithm uses round-robin method to assign jobs. First, it randomly selects the first node and assigns tasks to other nodes in a round-robin manner. This is one of the easiest methods of load balancing.

Processors assign each process circularly without defining any priority. It gives fast response in case of uniform workload distribution among the processes. All processes have different loading times. Therefore, some nodes may be heavily loaded, while others may remain under-utilised.

### Weighted Round Robin Load Balancing Algorithm

Weighted Round Robin Load Balancing Algorithms have been developed to enhance the most challenging issues of Round Robin Algorithms. In this algorithm, there are a specified set of weights and functions, which are distributed according to the weight values.

Processors that have a higher capacity are given a higher value. Therefore, the highest loaded servers will get more tasks. When the full load level is reached, the servers will receive stable traffic.

### Opportunistic Load Balancing Algorithm

The opportunistic load balancing algorithm allows each node to be busy. It never considers the current workload of each system. Regardless of the current workload on each node, OLB distributes all unfinished tasks to these nodes.

The processing task will be executed slowly in an OLB, and it does not count the implementation time of the node, which causes some bottlenecks even when some nodes are free.

### Minimum To Minimum Load Balancing Algorithm

Under minimum to minimum load balancing algorithms, first of all, those tasks take minimum time to complete. Among them, the minimum value is selected among all the functions. According to that minimum time, the work on the machine is scheduled.

Other tasks are updated on the machine, and the task is removed from that list. This process will continue till the final assignment is given. This algorithm works best where many small tasks outweigh large tasks.

## Types of Load Balancing

### Network Load Balancing

Cloud load balancing takes advantage of network layer information and leaves it to decide where network traffic should be sent. This is accomplished through Layer 4 load balancing, which handles TCP/UDP traffic. It is the fastest local balancing solution, but it cannot balance the traffic distribution across servers.

### HTTP(S) load balancing

HTTP(s) load balancing is the oldest type of load balancing, and it relies on Layer 7. This means that load balancing operates in the layer of operations. It is the most flexible type of load balancing because it lets you make delivery decisions based on information retrieved from HTTP addresses.

### Internal Load Balancing

It is very similar to network load balancing, but is leveraged to balance the infrastructure internally.

## WHY CLOUD LOAD BALANCING IS IMPORTANT IN CLOUD COMPUTING?

Here are some of the importance of load balancing in cloud computing.

### Offers better performance

The technology of load balancing is less expensive and also easy to implement. This allows companies to work on client applications much faster and deliver better results at a lower cost.

### Helps Maintain Website Traffic

Cloud load balancing can provide scalability to control website traffic. By using effective load balancers, it is possible to manage high-end traffic, which is achieved using network equipment and servers. E-commerce companies that need to deal with multiple visitors every second use cloud load balancing to manage and distribute workloads.

### Can Handle Sudden Bursts in Traffic

Load balancers can handle any sudden traffic bursts they receive at once. For example, in case of university results, the website may be closed due to too many requests. When one uses a load balancer, he does not need to worry about the traffic flow. Whatever the size of the traffic, load balancers will divide the entire load of the website equally across different servers and provide maximum results in minimum response time.

### Greater Flexibility

The main reason for using a load balancer is to protect the website from sudden crashes. When the workload is distributed among different network servers or units, if a single node fails, the load is transferred to another node. It offers flexibility, scalability and the ability to handle traffic better.

Because of these characteristics, load balancers are beneficial in cloud environments. This is to avoid heavy workload on a single server.