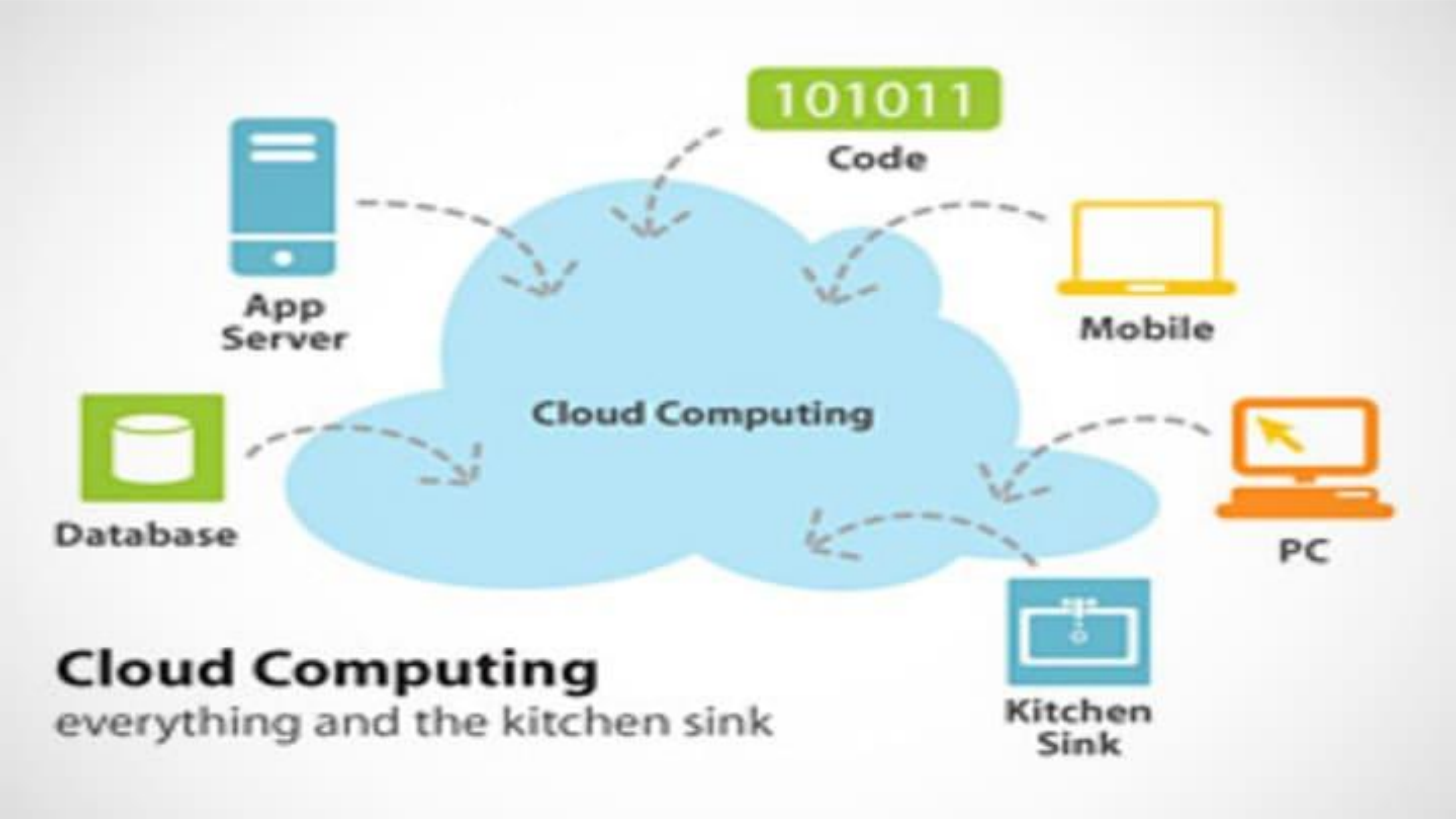
Unit 1: Introduction

Introduction to cloud computing

- Technologies such as cluster, grid, and now, cloud computing, have all aimed at allowing access to large amounts of computing power in a fully virtualized manner, by aggregating resources and offering a single system view.
- In addition, an important aim of these technologies has been delivering **computing as a utility.**
- Utility computing describes a business model for on-demand delivery of computing power; consumers pay providers based on usage ("pay as-you-go").


101011
Code
App
Server
Mobile
Cloud Computing
Database
PC
Kitchen
Sink
Cloud Computing
everything and the kitchen sink

Introduction to cloud computing

- Cloud computing has been coined as an umbrella term to describe a category of sophisticated **on-demand computing services** initially offered by commercial providers, such as **Amazon, Google, and Microsoft.**
- It denotes a model on which a computing infrastructure is viewed as a **“cloud,”** from which businesses and individuals access applications from anywhere in the world on demand
- The main principle behind this model is offering computing, storage, and software **“as a service**.”

History of Cloud computing

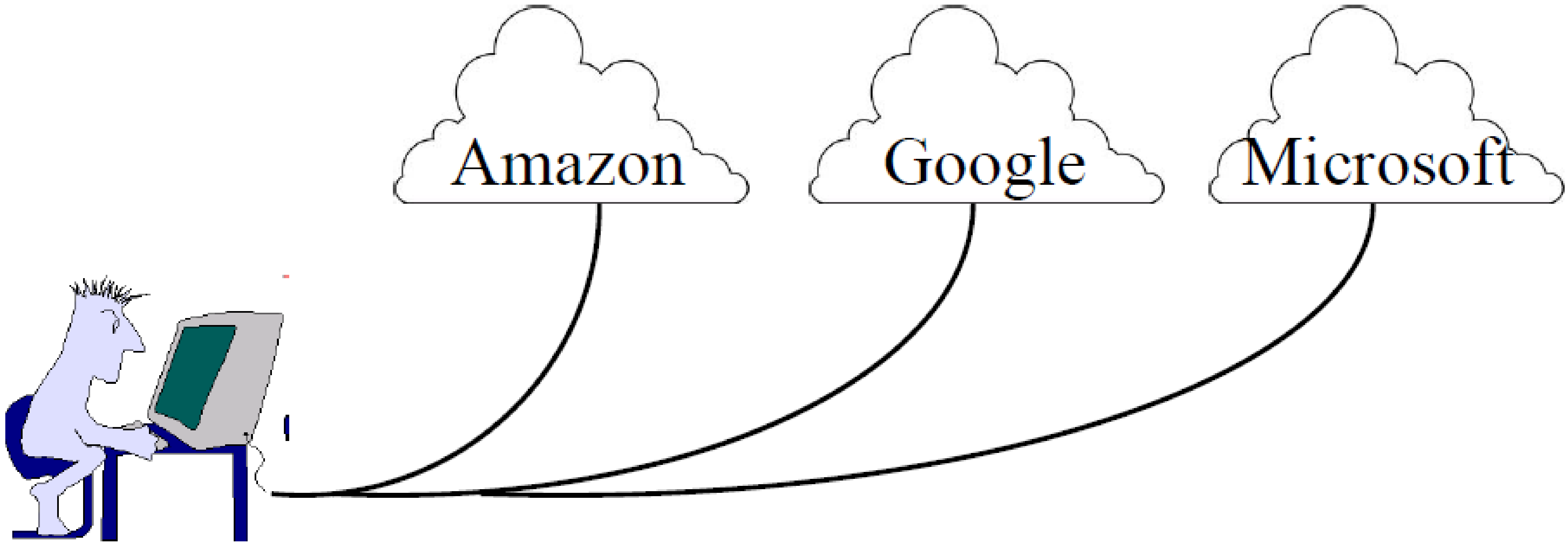
- Before emerging the cloud computing, there **was Client/Server computing** which is basically a centralized storage in which all the software applications, all the data and all the controls are resided on the server side.
- If a single user wants to access specific data or run a program, he/she need to connect to the server and then gain appropriate access, and then he/she can do his/her business.
- Then after, **distributed computing** came into picture, where all the computers are networked together and share their resources when needed.
- On the basis of above computing, there was emerged of cloud computing concepts that later implemented.

History of Cloud computing

- At around in 1961, John MacCharty suggested in a speech at MIT that computing can be sold like a utility, just like a water or electricity.
- It was a brilliant idea, but like all brilliant ideas, it was ahead if its time, as for the next few decades, despite interest in the model, the technology simply was not ready for it

History of Cloud computing

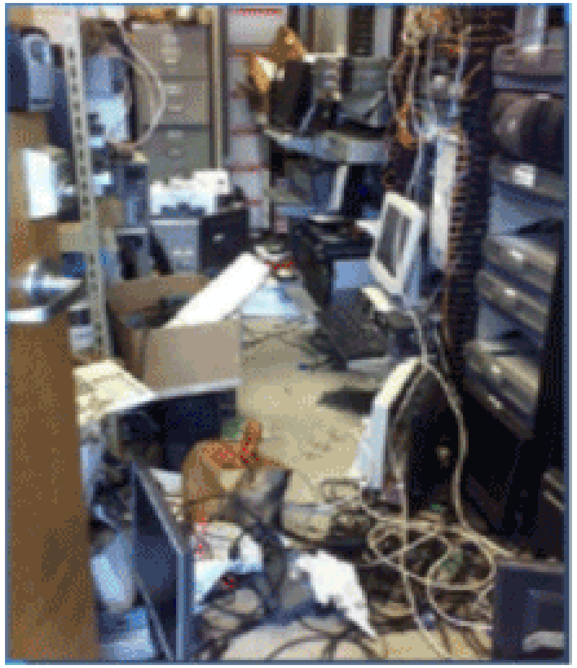
- *In 1999*, **Salesforce.com** *started delivering of applications to users using a simple website*. The applications were delivered to enterprises over the Internet, and this way the dream of computing sold as utility were true.
- *In 2002*, **Amazon** *started Amazon Web Services*, providing services like storage, computation and even human intelligence. However, only starting with the launch of the Elastic Compute Cloud in 2006 a truly commercial service open to everybody existed.
- *In 2009*, **Google Apps** *also started to provide cloud computing enterprise applications*.


Amazon
Google
Microsoft

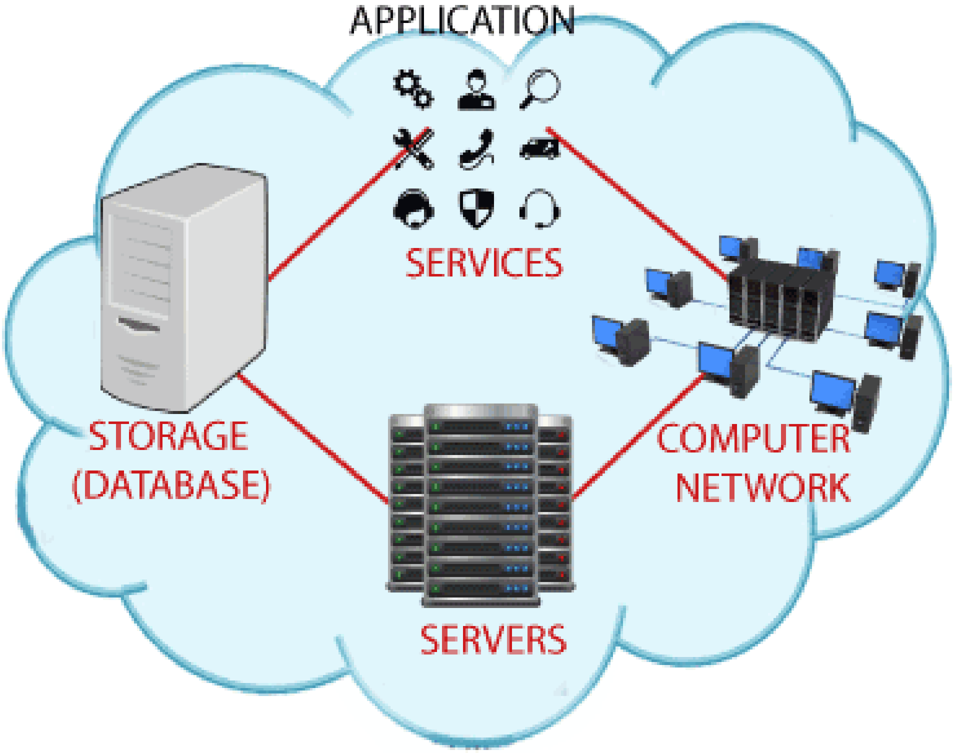
History of Cloud computing

- August 25, 2006: Amazon announced EC2 ⇒ Birth of Cloud Computing in reality (Prior theoretical concepts of computing as a utility)
- Amazon’s CEO was amazed by the number of computers in their data center and their low utilization
- Computing facilities are designed for peak load (Christmas)
- Needed a way to rent unused capacity, like renting their warehouses and other infrastructure ⇒ Develop an application programming interfaces (APIs) to remotely use computers.
- So began the computer rental business that we now call cloud computing.
- Sharing an underutilized resource is good for cloud service customers as well as for the cloud service providers.

Before
Cloud Computing



After
Cloud Computing



Introduction to cloud computing

- **Buyya** have defined it as follows:
  - “Cloud is a parallel and distributed computing system
  - consisting of a collection of inter-connected and virtualised computers that are dynamically provisioned and presented as one
  - or more unified computing resources based on service-level agreements (SLA)
  - established through negotiation between the service provider and consumers.”

Introduction to cloud computing

- **Vaquero** have stated:
- “clouds are a large pool of easily usable and accessible virtualized resources (such as hardware, development platforms and/or services).
- These resources can be dynamically reconfigured to adjust to a variable load (scale), allowing also for an optimum resource utilization.
- This pool of resources is typically exploited by a pay-per-use model in which guarantees are offered by the Infrastructure Provider by means of customized Service Level Agreements.

Introduction to cloud computing

- A recent **McKinsey and Co**. report claims that
- “Clouds are hardware based services offering compute, network, and storage capacity
- where: Hardware management is highly abstracted from the buyer, buyers incur infrastructure costs as variable OPEX, and infrastructure capacity is highly elastic.”

Introduction to cloud computing

- While there are countless other definitions, there seems to be common characteristics between the most notable ones listed above, which a cloud should have:
  - **(i) pay-per-use (no ongoing commitment, utility prices)**
  - **(ii) elastic capacity and the illusion of infinite resources**
  - **(iii) self-service interface**
  - **(iv) resources that are abstracted or virtualised**
- In addition to raw computing and storage, cloud computing providers usually offer a broad range of software services. They also include APIs and development tools that allow developers to build seamlessly scalable applications upon their services.
- The ultimate goal is allowing customers to run their everyday IT infrastructure “in the cloud.”

Attributes of Cloud Service

- **Off-Premise**: Out-side the company firewall. Connected via Internet Belongs to a service provider
- **Elasticity**: Scalable up or down rapidly.
- **Flexible Billing**: Usage or flat rate. Payment or advertising-paid (as in Google docs)
- **Multi-tenancy**: Cost optimization of sharing
- **Universal Access**: Available to anyone
- **Virtualization**: Easier allocation of resources
- **Service**: Provides both management and computational APIs.
- **Service level agreement**: Different levels

LAYERS AND TYPES OF CLOUDS

- Cloud computing services are divided into three classes:
- according to the abstraction level of the capability provided and the service model of providers, namely:
- **(1) Infrastructure as a Service (IaaS)**
- **(2) Platform as a Service, and (PaaS)**
- **(3) Software as a Service (SaaS)**

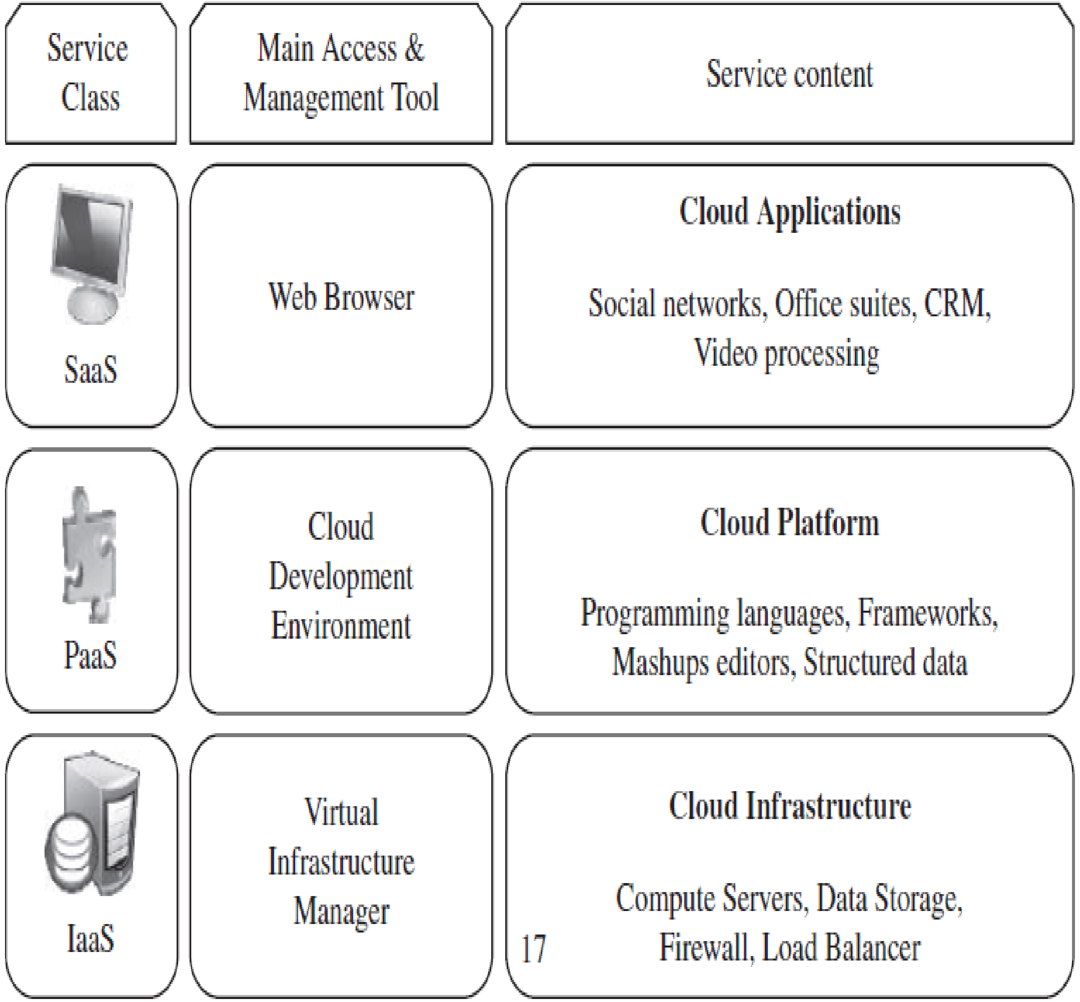
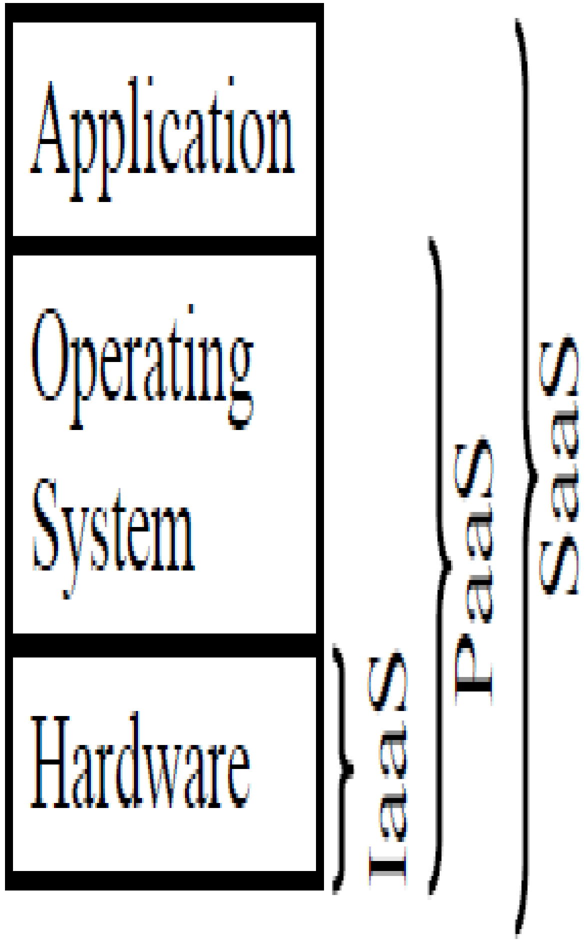
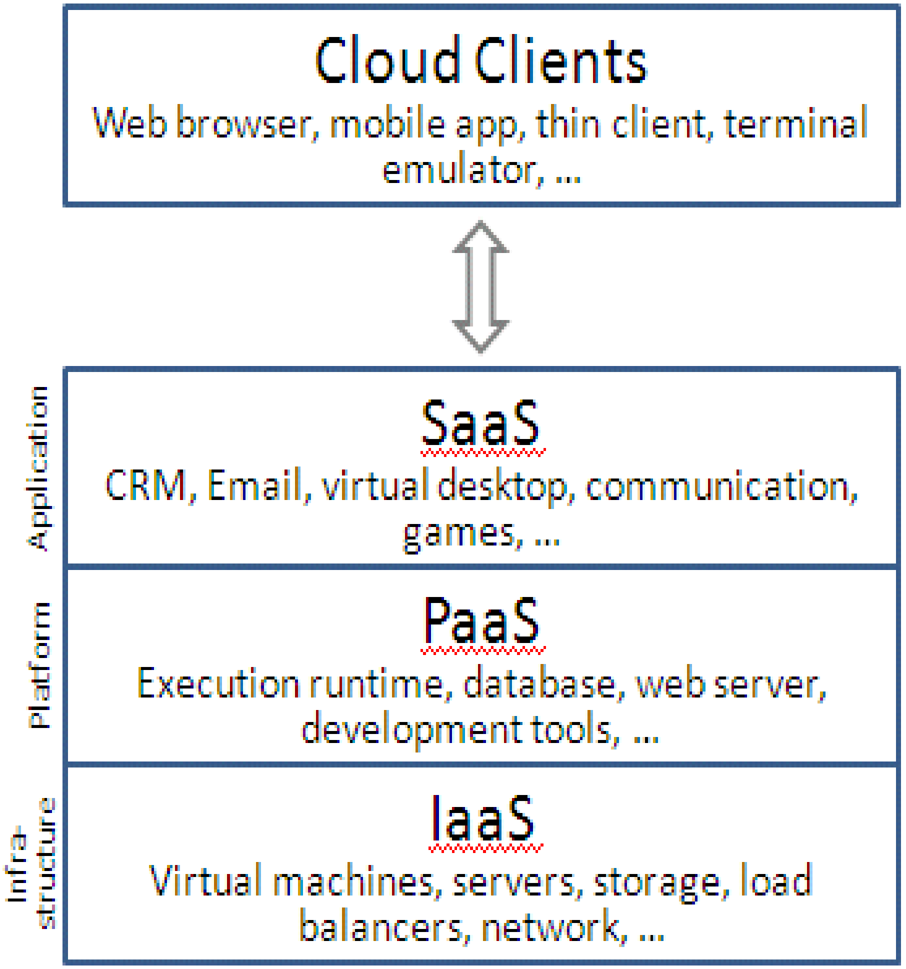


FIGURE 1.3. The cloud computing stack.




Cloud Clients
Web browser, mobile app, thin client, terminal emulator, ...
SaaS
CRM, Email, virtual desktop, communication, games, ...
PaaS
Execution runtime, database, web server, development tools, ...
IaaS
Virtual machines, servers, storage, load balancers, network, ...
Application
Platform
Infra-structure

- These abstraction levels can also be viewed as a layered architecture where services of **a higher layer can be composed from services of the underlying layer**
- A **core middleware manages physical resources and the VMs** deployed on top of them
- in addition, it provides the required features (e.g., accounting and billing) to offer multi-tenant pay-as-you-go services.
- Cloud development environments are built on top of infrastructure services to offer application development and deployment capabilities
- in this level, various programming models, libraries, APIs, and mashup editors enable the creation of a range of business, Web, and scientific applications.
- Once deployed in the cloud, these applications can be consumed by end users.

Infrastructure as a Service

- Offering **virtualized resources (computation, storage, and communication)** on demand is known as Infrastructure as a Service (IaaS) .
- A cloud infrastructure enables **on-demand provisioning of servers** running several choices of operating systems and a customized software stack.
- Infrastructure services are considered to be the **bottom layer of cloud computing systems .**
- Amazon Web Services mainly offers IaaS, which in the case of its EC2 service means offering VMs with a software stack that can be customized similar to how an ordinary physical server would be customized.

- Users are given privileges to perform numerous activities to the server, such as:
  - **starting and stopping it**
  - **customizing it by installing software packages**
  - **attaching virtual disks to it**
  - **configuring access permissions and firewalls rules.**

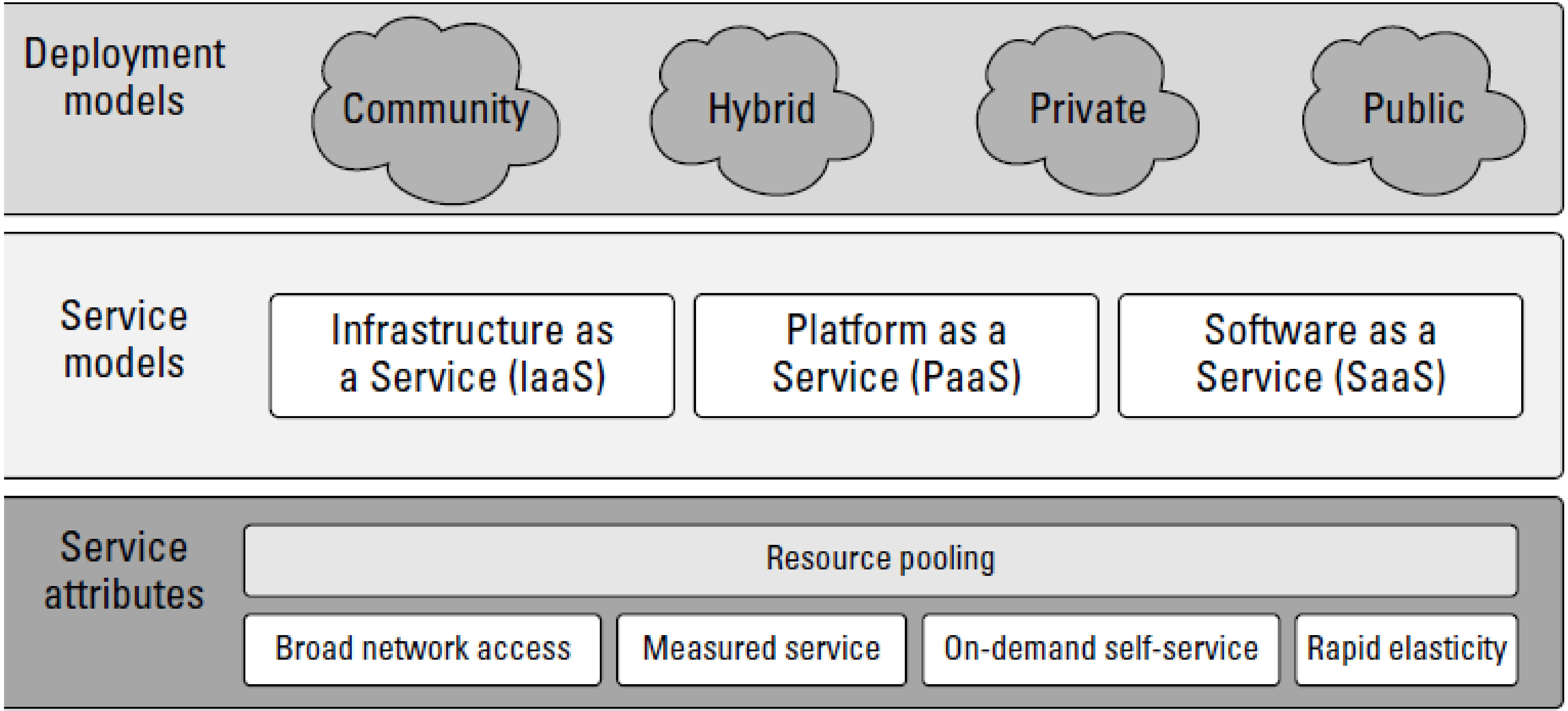
Platform as a Service

- In addition to infrastructure-oriented clouds that provide raw computing and storage services, another approach is to offer **a higher level of abstraction** to make a cloud easily programmable, known as Platform as a Service (PaaS).
- A cloud platform offers an environment on which **developers create and deploy applications** and do not necessarily need to know how many processors or how much memory that applications will be using.
- In addition, multiple programming models and specialized services (e.g., data access, authentication, and payments) are offered as building blocks to new applications .

- Google AppEngine, an example of Platform as a Service, offers a scalable environment for developing and hosting Web applications, which should be written in specific programming languages such as Python or Java, and use the services' own proprietary structured object data store.
- Building blocks include an in-memory object cache (memcache), mail service, instant messaging service (XMPP), an image manipulation service, and integration with Google Accounts authentication service.

Software as a Service (SaaS)

- Applications reside on the top of the cloud stack.
- Services provided by this layer can be accessed by end users through **Web portals**.
- Therefore, consumers are increasingly shifting from locally installed computer programs to on-line software services that offer the same functionally.
- Traditional desktop applications such as word processing and spreadsheet can now be accessed as a service in the Web.
- This model of delivering applications, known as Software as a Service (SaaS), alleviates the burden of software maintenance for customers and simplifies development and testing for providers.
- Salesforce.com, which relies on the SaaS model, offers business productivity applications (CRM) that reside completely on their servers, allowing costumers to customize and access applications on demand.


Deployment models
Community
Hybrid
Private
Public
Service models
Infrastructure as a Service (IaaS)
Platform as a Service (PaaS)
Software as a Service (SaaS)
Service attributes
Resource pooling
Broad network access
Measured service
On-demand self-service
Rapid elasticity

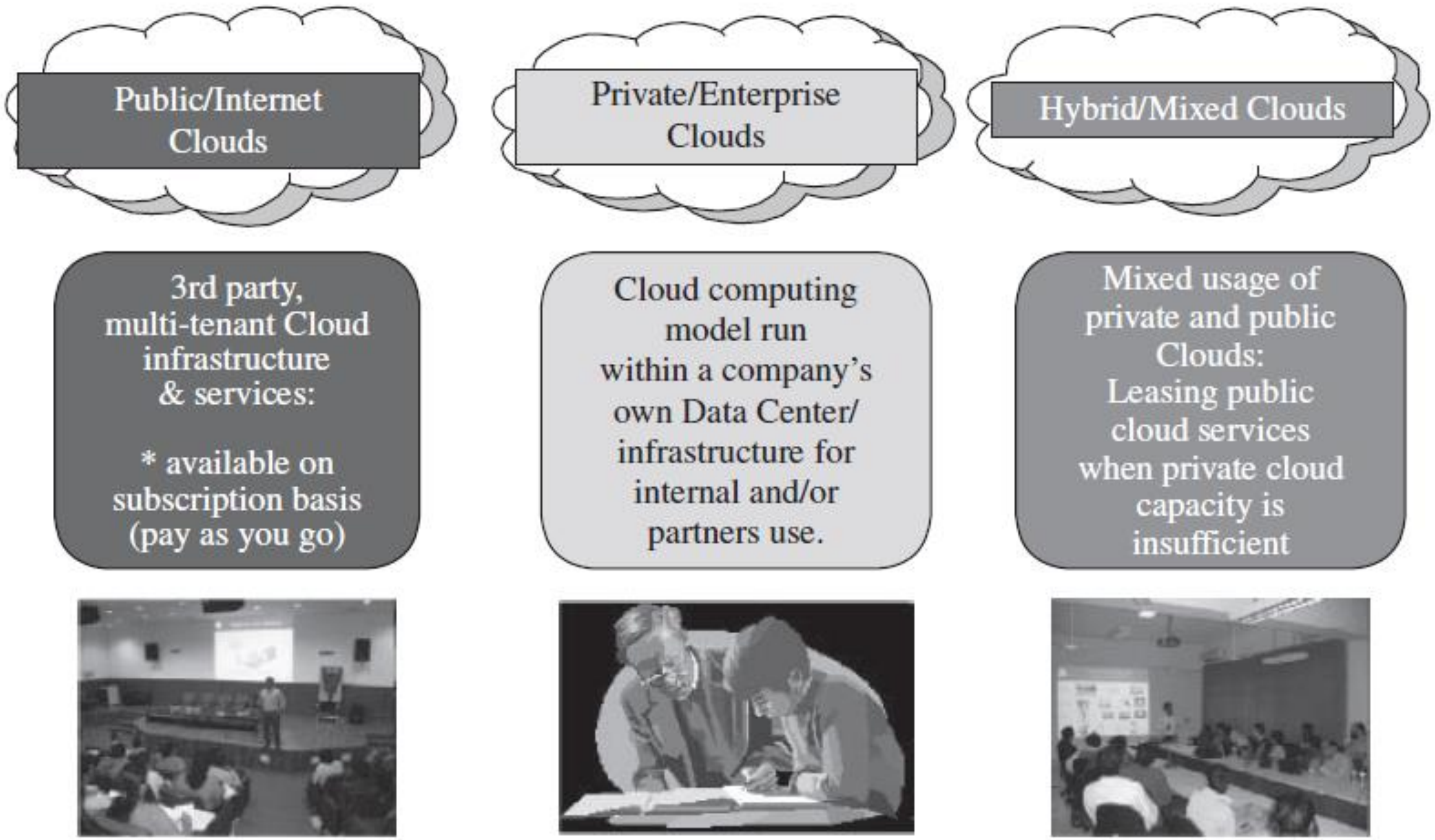
Deployment Models

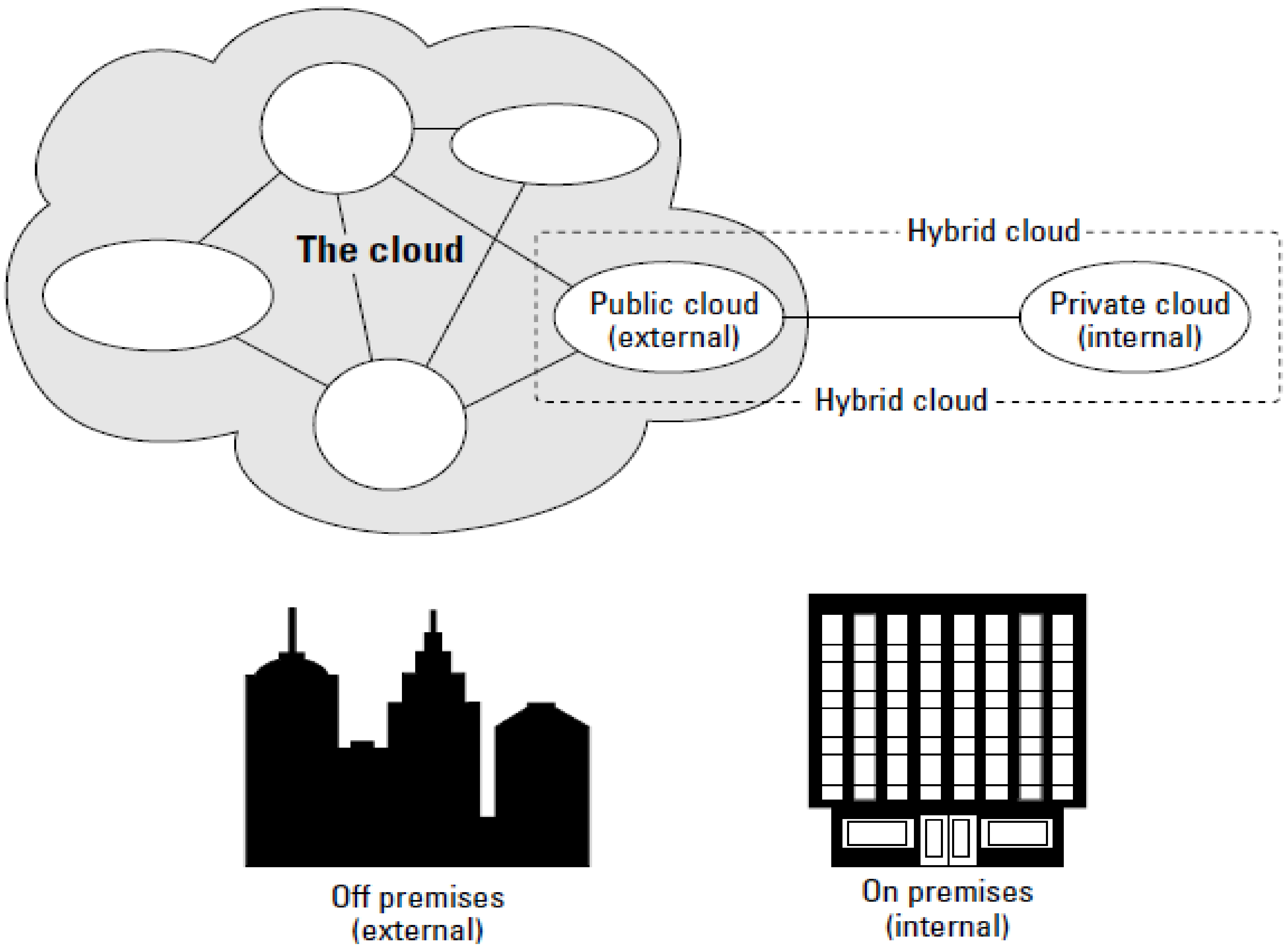
- Although cloud computing has emerged mainly from the appearance of public computing utilities, other deployment models, with **variations in physical location and distribution**, have been adopted.
- In this sense, regardless of its service class, a cloud can be classified as **public, private, community, or hybrid** based on model of deployment

- **A Public cloud** is a “cloud made available in a pay-as-you-go manner to the general public”
- The public cloud infrastructure is available for public use alternatively for a large industry group and is owned by an organization selling cloud services.
- **A Private cloud** is a “internal data centre of a business or other organization, not made available to the general public.”
- The private cloud infrastructure is operated for the exclusive use of an organization.
- The cloud may be managed by that organization or a third party.
- Private clouds may be either on- or off-premises.

- **A hybrid cloud** takes shape when a private cloud is supplemented with computing capacity from public clouds .
- The approach of temporarily renting capacity to handle spikes in load is known as **“cloud-bursting”.**
- A hybrid cloud **combines multiple clouds** (private, community of public)
- where those clouds retain their unique identities, but are bound together as a unit.
- A hybrid cloud may offer standardized or proprietary access to data and applications, as well as application portability.

- **A community cloud** is "shared by several organizations and supports a specific community that has shared concerns (e.g., mission, security requirements, policy, and compliance considerations)"
- A community cloud is one where the cloud has been organized to serve a common function or purpose.
- It may be for one organization or for several organizations, but they share common concerns
- A community cloud may be managed by the constituent organization(s) or by a third party.


Public/Internet Clouds
Private/Enterprise Clouds
Hybrid/Mixed Clouds
3rd party, multi-tenant Cloud infrastructure & services:
* available on subscription basis (pay as you go)
Cloud computing model run within a company's own Data Center/ infrastructure for internal and/or partners use.
Mixed usage of private and public Clouds:
Leasing public cloud services when private cloud capacity is insufficient


The cloud
Hybrid cloud
Public cloud
(external)
Private cloud
(internal)
Hybrid cloud
Off premises
(external)
On premises
(internal)

Virtualization in Cloud Computing

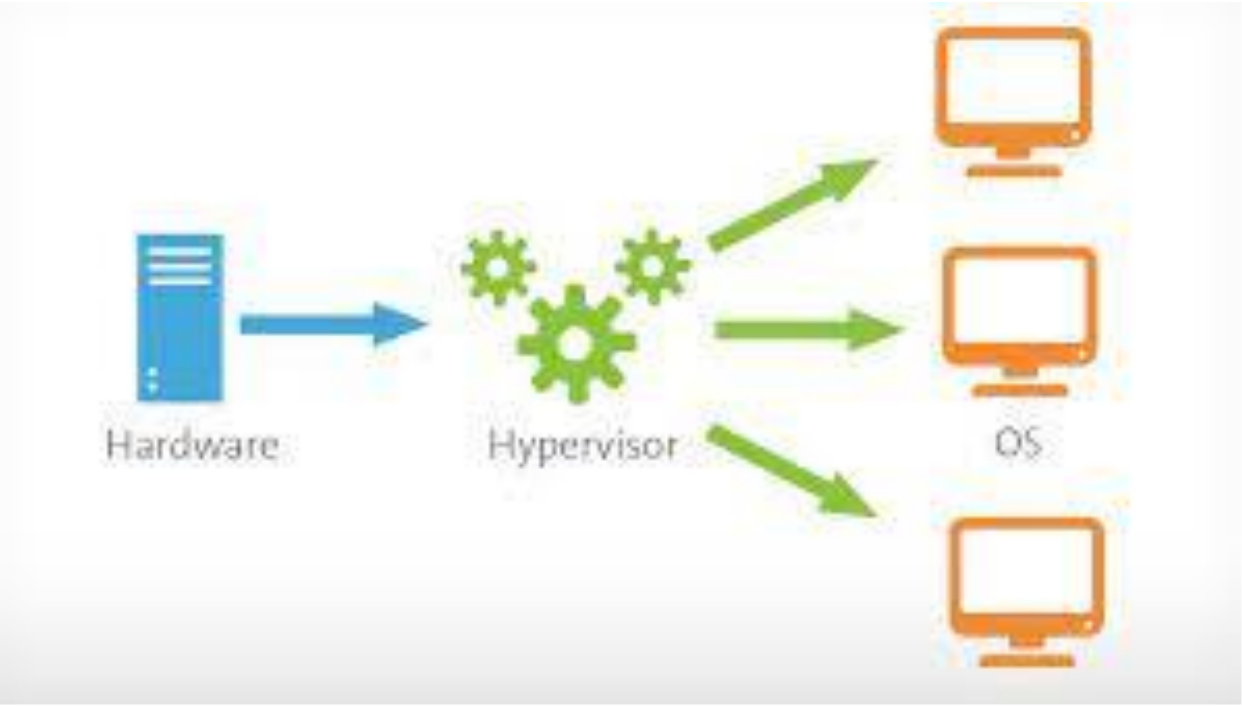
- Virtualization in Cloud Computing is making a virtual platform of **server operating system and storage devices.**
- This will help the user by providing multiple machines at the same time it also allows sharing a single physical instance of resource or an application to multiple users.
- Cloud Virtualizations also manage the workload by transforming traditional computing and make it more scalable, economical and efficient.
- Virtualizations in Cloud Computing rapidly integrating the fundamental way of computing.
- One of the important features of virtualization is that it allows **sharing of applications** to multiple customers and companies.

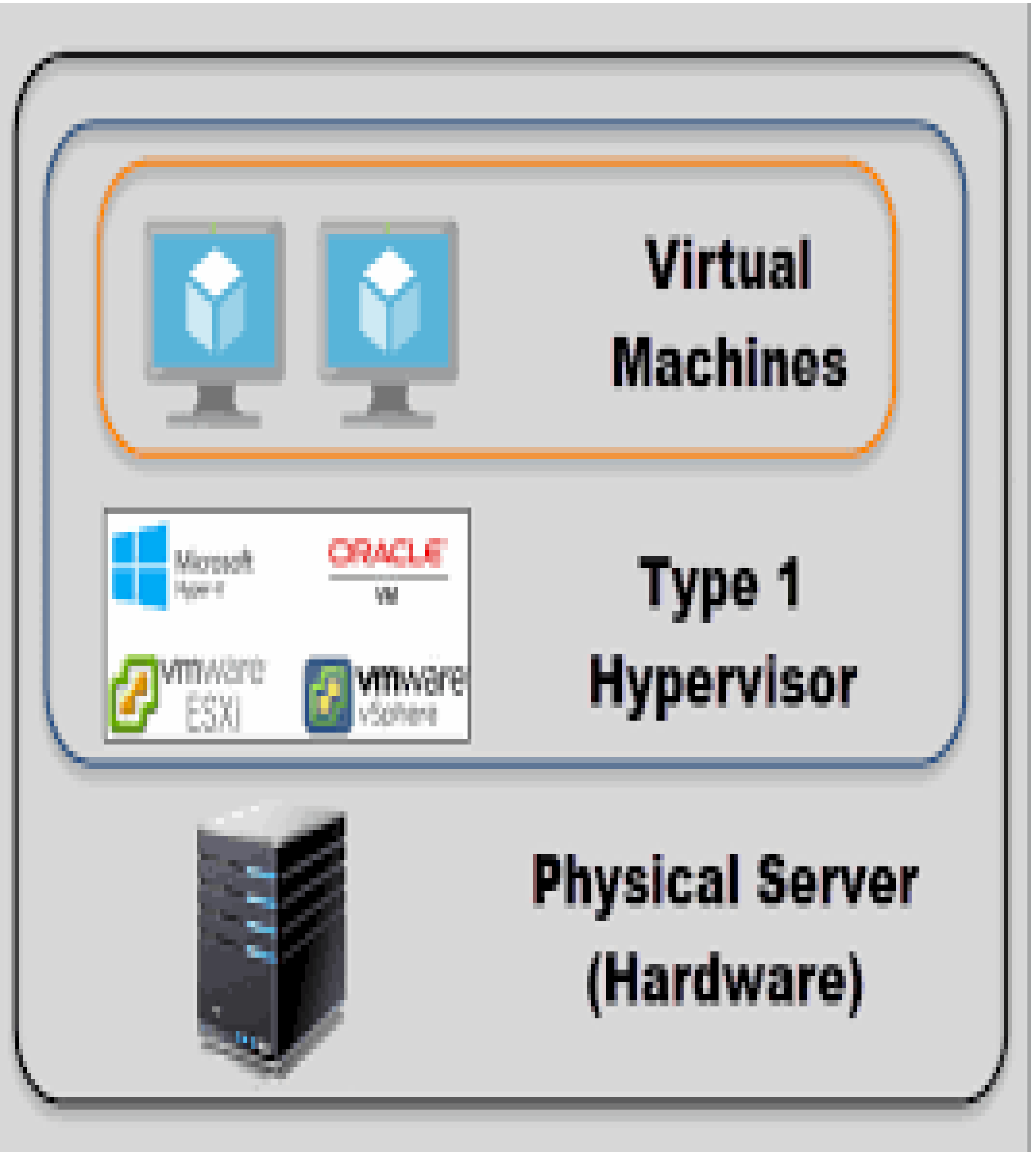
Virtualization in Cloud Computing

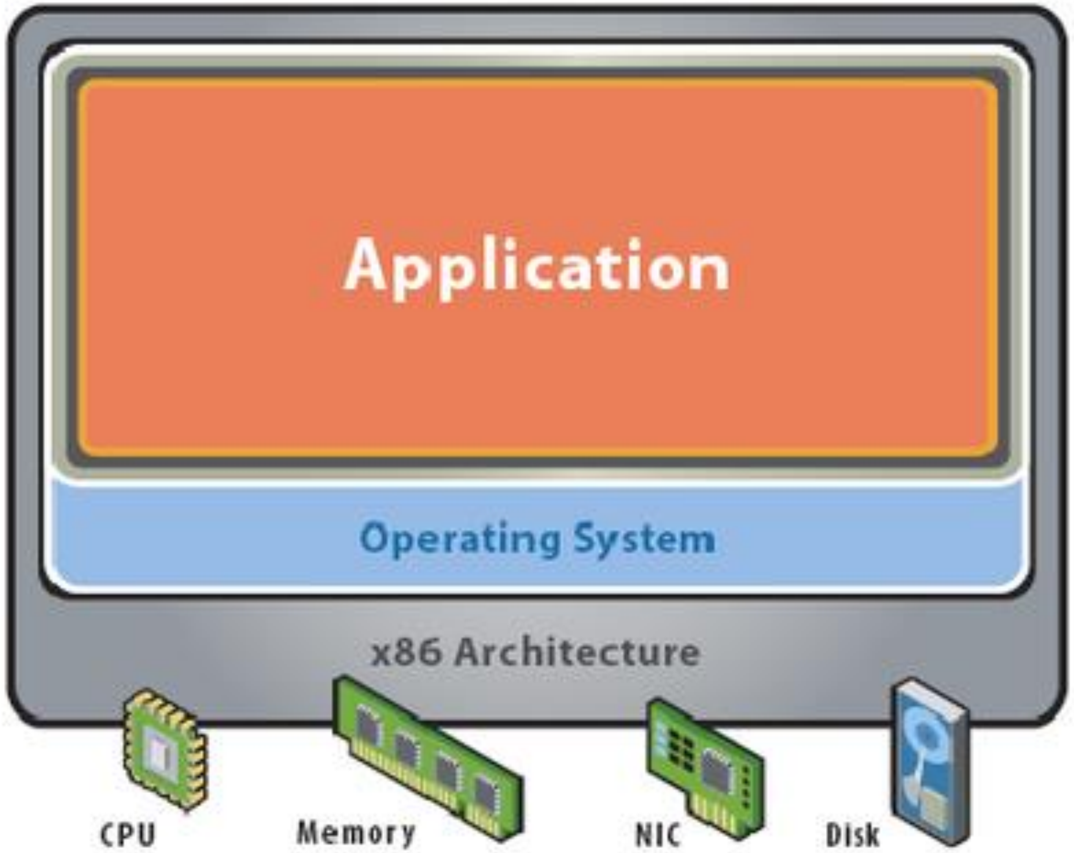
- Cloud Computing can also be known as services and application delivered to help the virtualized environment.
- This environment can be either **public** or **private**.
- With the help of virtualization, the customer can maximize the resources and reduces the physical system which is in need.
- Virtualization is achieved by assigning a logical name to a physical storage and providing a pointer to that physical resource when demanded.

Virtualization in Cloud Computing

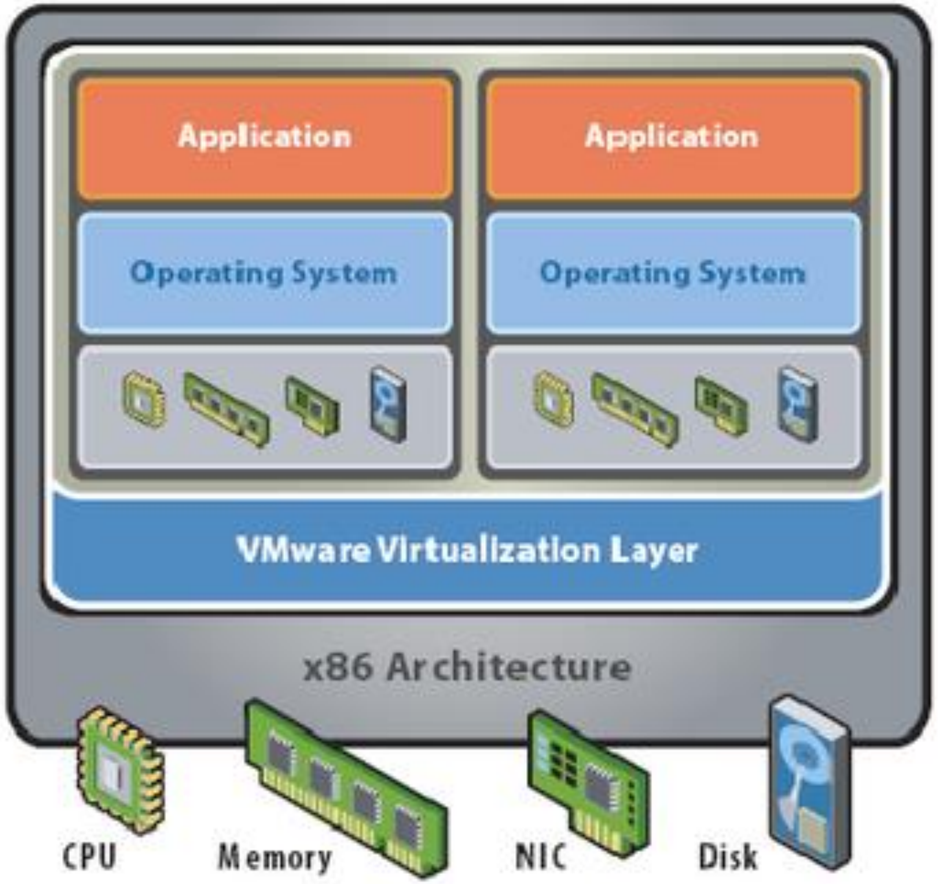
- Creation of a virtual machine over existing operating system and hardware is known as Hardware Virtualization.
- **A Virtual machine** provides an environment that is logically separated from the underlying hardware.
- The machine on which the virtual machine is going to create is known as **Host Machine** and that virtual machine is referred as a **Guest Machine**
- It involves using **specialized software (Hypervisor)** to create a virtual or software-created version of a computing resource rather than the actual version of the same resource


Hardware
Hypervisor
OS


Virtual Machines
ORACLE
vmware ESXi
vmware vSphere
Type 1 Hypervisor
Physical Server (Hardware)



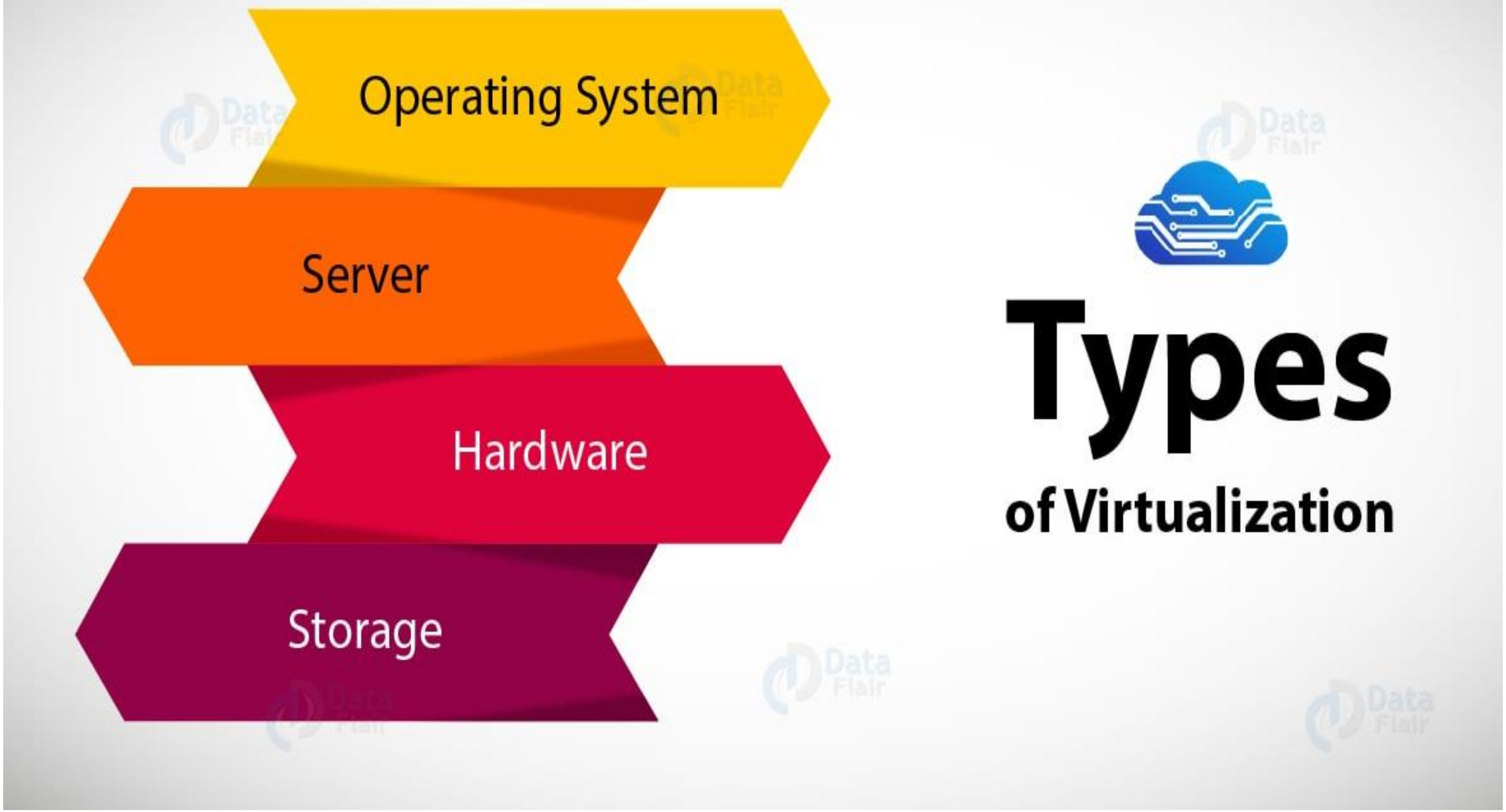
Before Virtualization:



After Virtualization:

Types of Virtualization in Cloud Computing

- Operating System Virtualization
- Hardware Virtualization
- Server Virtualization
- Storage Virtualization


Operating System
Server
Hardware
Storage
Types
of Virtualization

Operating System Virtualization

- When the **virtual machine software or virtual machine manager *(VMM)*** *is installed on the Host operating system* instead of directly on the hardware system is known as operating system virtualization.
- Operating System Virtualization is mainly used for testing the applications on different platforms of OS.

Hardware Virtualization

- When the virtual machine software or virtual machine manager *(VMM)* ***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.
- After virtualization of hardware system we can install different operating system on it and run different applications on those OS.

Server Virtualization

- When the virtual machine software or virtual machine manager *(VMM)* ***is directly installed on the Server system*** is known as server virtualization.
- Server virtualization is done because a single physical server can be divided into multiple servers on the demand basis and for balancing the load.

Storage Virtualization

- In **storage virtualization** in Cloud Computing, a grouping is done of physical storage which is from multiple network storage devices this is done so it looks like a single storage device.
- It can implement with the help of software applications and storage virtualization is done for the backup and recovery process.
- It is a sharing of the physical storage from multiple storage devices.

CLOUD INFRASTRUCTURE MANAGEMENT

- A key challenge IaaS providers face when building a cloud infrastructure is **managing physical and virtual resources**, namely servers, storage, and networks, in a holistic fashion .
- The orchestration of resources must be performed in a way to rapidly and dynamically provision resources to applications.
- The software toolkit responsible for this orchestration is called a **virtual infrastructure manager (VIM).**
- This type of software resembles a traditional operating system—but instead of dealing with a single computer, it aggregates resources from multiple computers, presenting a uniform view to user and applications.
- The term **"cloud operating system"** is also used to refer to it .
- Other terms include **"infrastructure sharing software "** and **"virtual infrastructure engine ."**

CLOUD INFRASTRUCTURE MANAGEMENT

- two categories of tools used to manage clouds.
- **The first category—cloud toolkits**—includes those that “expose a remote and secure interface for creating, controlling and monitoring virtualize resources,” but do not specialize in VI management.
- **Tools in the second category—the virtual infrastructure managers**—provide advanced features such as automatic load balancing and server consolidation, but do not expose remote cloud-like interfaces.
- cloud toolkits can also manage virtual infrastructures, although they usually provide less sophisticated features than specialized VI managers do.

Advantages of Cloud Computing:-

- On-demand self-service
- Broad network access
- Resource pooling
- Rapid elasticity
- Measured service
- Lower costs
- Quality of Service
- Ease of utilization
- Reliability
- Outsourced IT management
- Simplified maintenance and upgrade
- Low Barrier to Entry

Challenges and Risks

- Despite the initial success and popularity of the cloud computing paradigm and the extensive availability of providers and tools, a significant number of challenges and risks are inherent to this new model of computing.
- Providers, developers, and end users must consider these challenges and risks to take good advantage of cloud computing.
- Issues to be faced include user privacy, data security, data lock-in, availability of service, disaster recovery, performance, scalability, energy-efficiency, and programmability.

1. Security, Privacy, and Trust

- current cloud offerings are essentially public exposing the system to more attacks.
- For this reason there are potentially additional challenges to make cloud computing environments as secure as in-house IT systems.
- At the same time, existing, well understood technologies can be leveraged, such as data encryption, VLANs, and firewalls.
- Security and privacy affect the entire cloud computing stack, since there is a massive use of third-party services and infrastructures that are used to host important data or to perform critical operations.
- In this scenario, the trust toward providers is fundamental to ensure the desired level of privacy for applications hosted in the cloud

- Legal and regulatory issues also need attention. When data are moved into the Cloud, providers may choose to locate them anywhere on the planet. The physical location of data centres determines the set of laws that can be applied to the management of data. For example, specific cryptography techniques could not be used because they are not allowed in some countries.
- Similarly, country laws can impose that sensitive data, such as patient health records, are to be stored within national borders.

2. Data Lock-In and Standardization

- A major concern of cloud computing users is about having their data locked-in by a certain provider. Users may want to move data and applications out from a provider that does not meet their requirements.
- However, in their current form, cloud computing infrastructures and platforms do not employ standard methods of storing user data and applications. Consequently, they do not interoperate and user data are not portable.
- The answer to this concern is **standardization.**
- In this direction, there are efforts to create open standards for cloud computing.

- The **Cloud Computing Interoperability Forum (CCIF)** was formed by organizations such as Intel, Sun, and Cisco in order to “enable a global cloud computing ecosystem
- whereby organizations are able to seamlessly work together for the purposes for wider industry adoption of cloud computing technology.” The development of the **Unified Cloud Interface (UCI)** by CCIF aims at creating a standard programmatic point of access to an entire cloud infrastructure

3. Availability, Fault-Tolerance, and Disaster Recovery

- It is expected that users will have certain expectations about the service level to be provided once their applications are moved to the cloud.
- These expectations include availability of the service, its overall performance, and what measures are to be taken when something goes wrong in the system or its components.
- In summary, users seek for a warranty before they can comfortably move their business to the cloud.
- SLAs (Service Level Agreements), which include QoS requirements, must be ideally set up between customers and cloud computing providers to act as warranty.

- An SLA specifies the details of the service to be provided, including availability and performance guarantees.
- Additionally, metrics must be agreed upon by all parties, and penalties for violating the expectations must also be approved.

4. Resource Management and Energy-Efficiency

- One important challenge faced by providers of cloud computing services is the efficient management of virtualized resource pools.
- Physical resources such as CPU cores, disk space, and network bandwidth must be sliced and shared among virtual machines running potentially heterogeneous workloads.
- The multi-dimensional nature of virtual machines complicates the activity of finding a good mapping of VMs onto available physical hosts while maximizing user utility.
- Dimensions to be considered include: number of CPUs, amount of memory, size of virtual disks, and network bandwidth.
- Dynamic VM mapping policies may leverage the ability to suspend, migrate, and resume VMs as an easy way of pre-empting low-priority allocations in favour of higher-priority ones.

- Migration of VMs also brings additional challenges such as detecting when to initiate a migration, which VM to migrate, and where to migrate.
- In addition, policies may take advantage of live migration of virtual machines to relocate data center load without significantly disrupting running services.
- In this case, an additional concern is the trade-off between the negative impact of a live migration on the performance and stability of a service and the benefits to be achieved with that migration