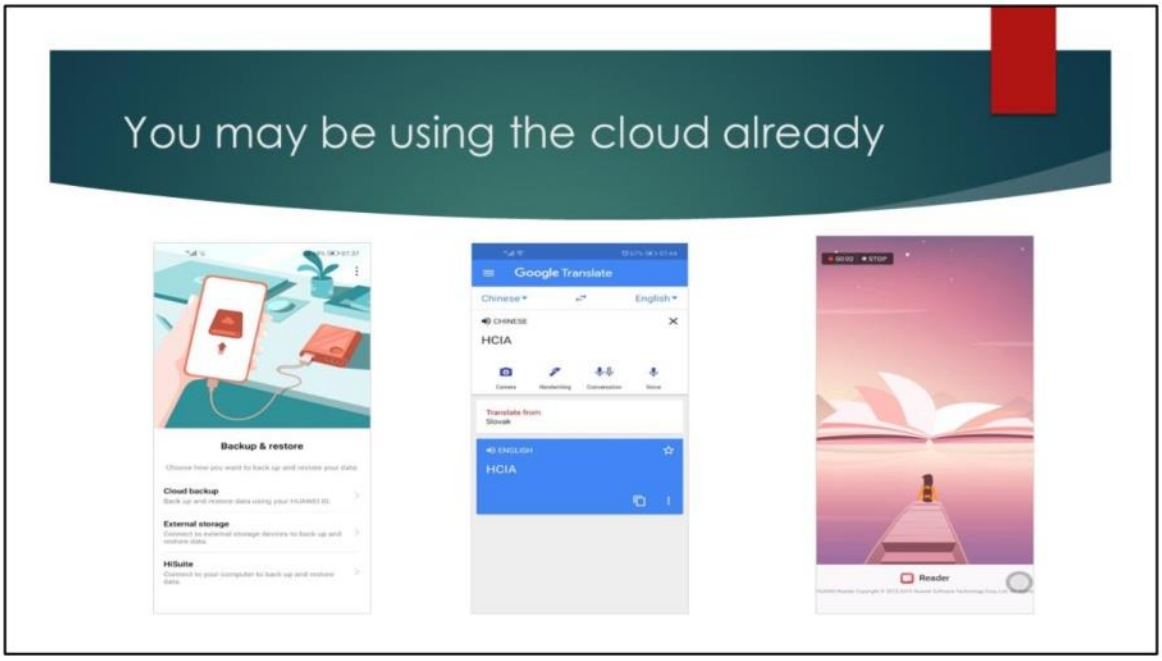
Cloud Computing and Virtualization

Outline

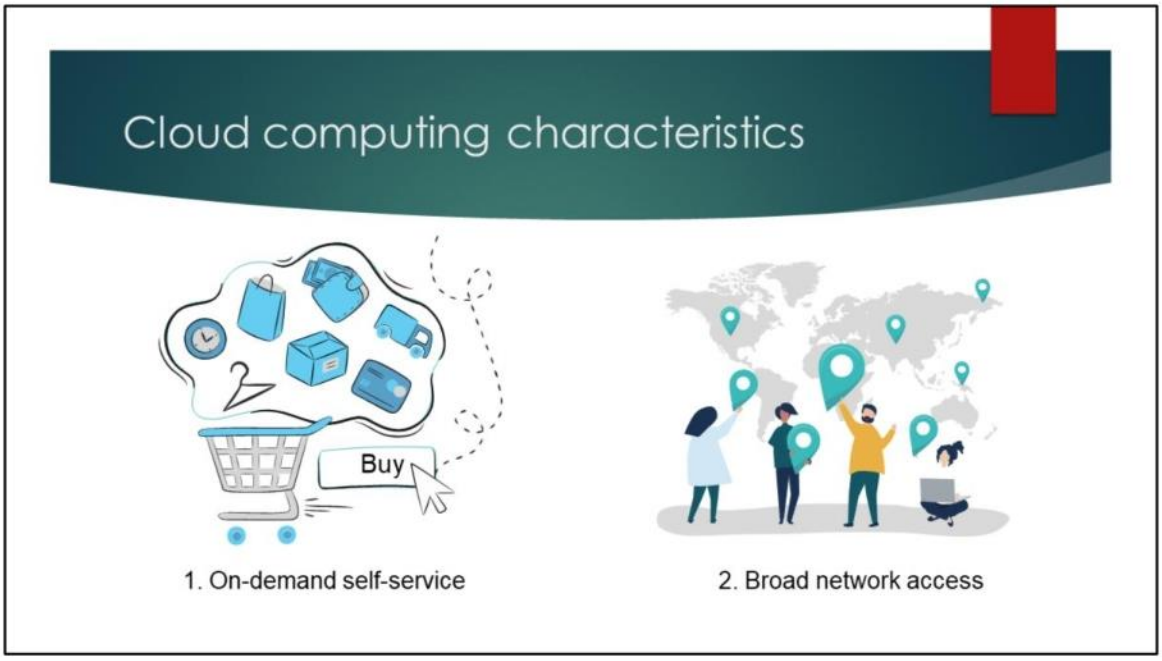
- Cloud computing characteristics
- Cloud computing definition
- Cloud computing models
- Introduction to virtualization



Cloud computing delivers ready-to-use IT services over the Internet.

Here are a few apps you might be using already:

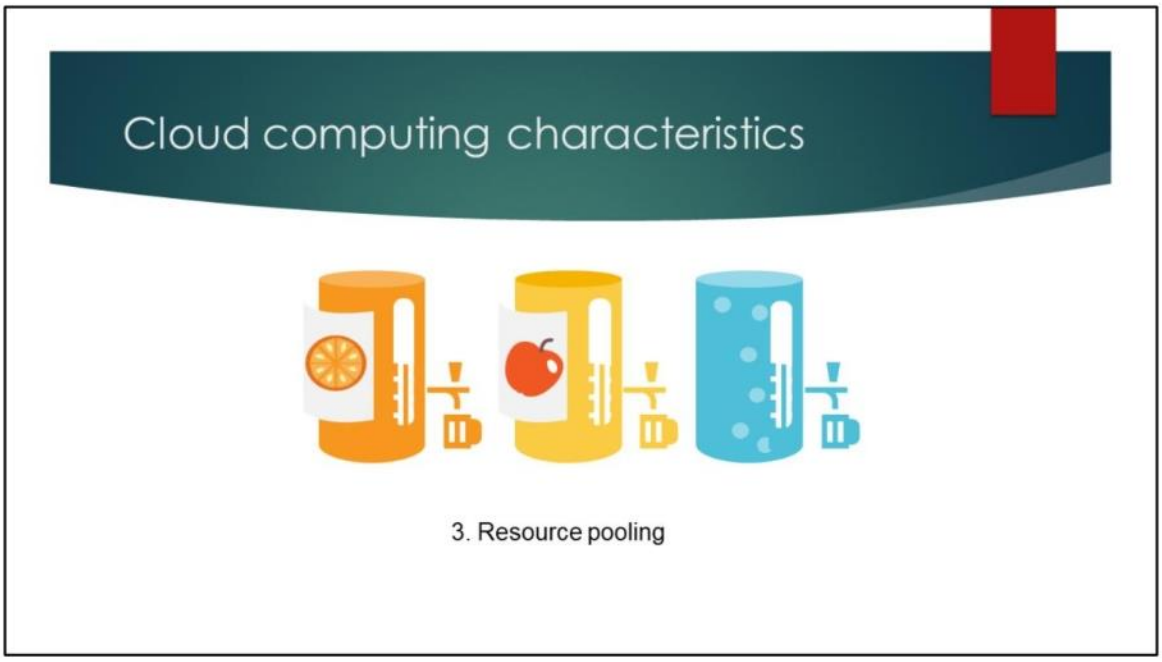
- Backup & Restore allows you to back up the local files on your phone to a remote data center. After you change to a new phone, you can easily restore your data to your new phone using your account and password configured for this service.
- Google Translate is a free service that instantly translates words, phrases, and web pages between English and over 100 other languages.
- iReader is a popular online reading app that gives you access to a huge library of online electronic books.



Cloud computing has the following five well-recognized characteristics:

1. **On-demand self-service**: A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service's provider. The prerequisite for on-demand self-service is for the consumer to understand their own needs and know which product or products can accommodate such needs.

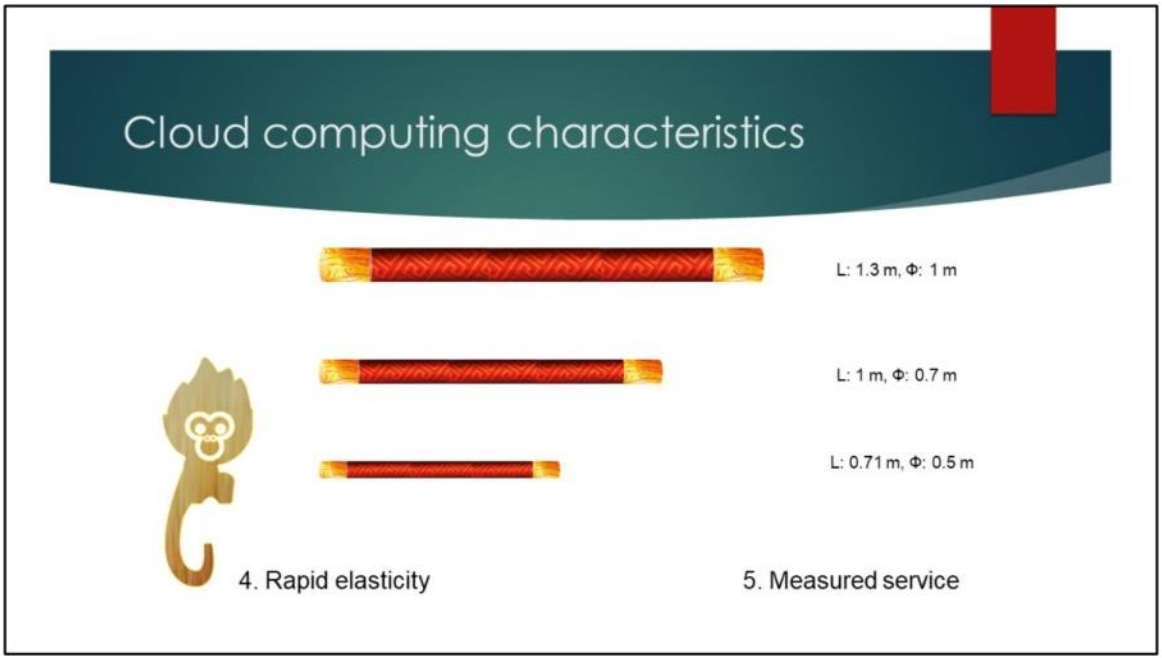
2. **Broad network access**: Cloud computing is computing power over the Internet, so network access is an innate characteristic of cloud computing.



3. **Resource pooling** is one of the prerequisites for on-demand self-service. In a supermarket, we can see that different categories of items are put into different areas, such as the fruit and vegetable area, the fast frozen food area, and so on, so that consumers can quickly find the items they need.

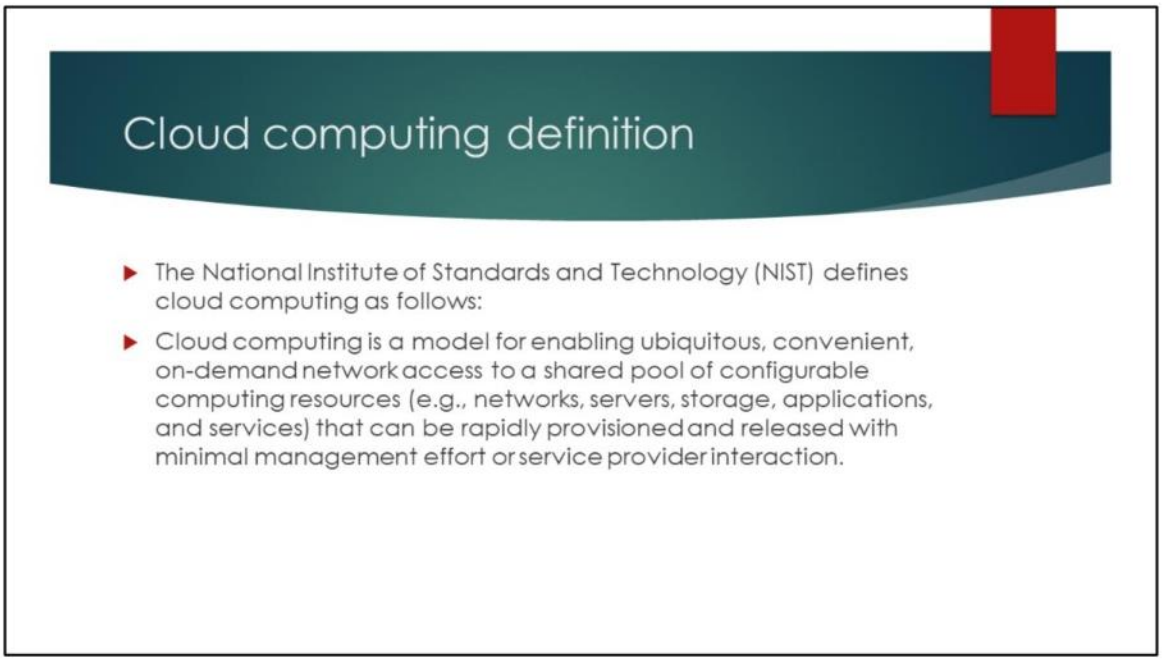
   Resource pooling is not merely putting all resources of the same type onto the same rack, as it is done in supermarkets. Resource pooling is also about breaking down the resources by the finest granularities for flexible, on-demand provisioning.

   Another thing that resource pooling does is to shield the differences in the underlying resources. Resources that can be pooled include compute, storage, and network resources. Consumers have no idea whether they are use AMD or Intel CPUs.



4. **Rapid elasticity** means being able to rapidly and elastically provision computing resources. Rapid elasticity, both scaling in or out, can be achieved manually or based on predefined policies. For example, with this characteristic, a startup company can start by acquiring only small amounts of IT resources and add more as its business grows.

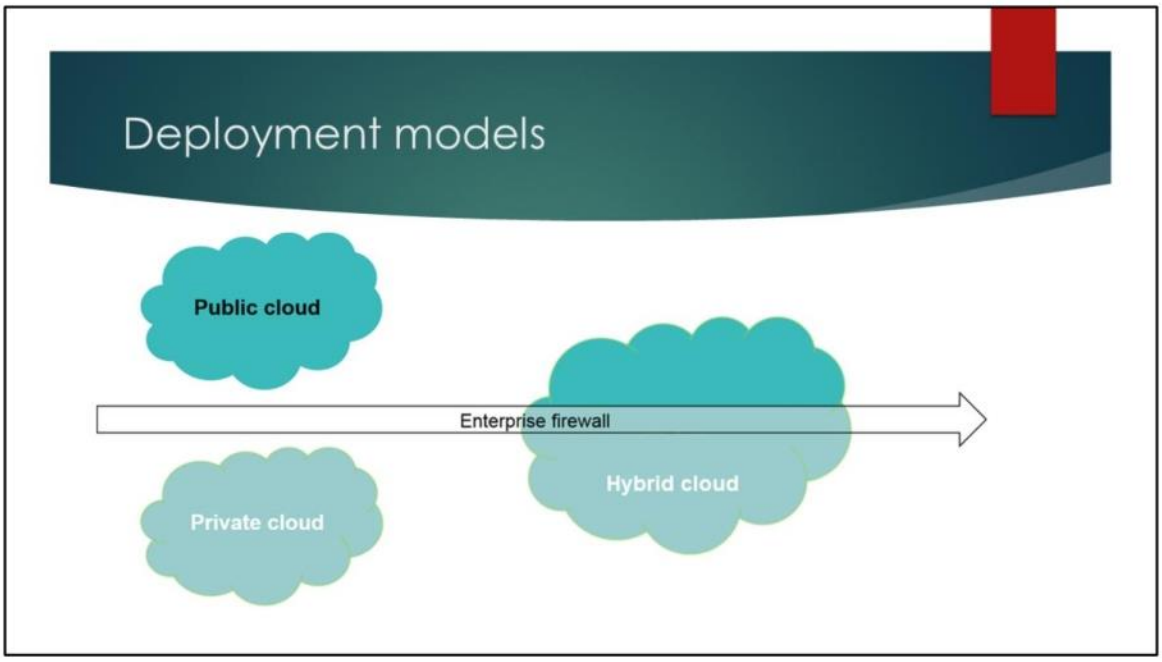
5. **Measured Service** is how cloud systems control a user or tenant's use of resources by leveraging a metering capability. Metering is not billing, although billing is based on metering. Measured service ensures that all resource usage can be accurately measured, based on predefined criteria, which can be the duration of usage, resource quota, or the volume of data transmitted.



- **Cloud computing** is not a technology so much as a **service delivery model**.
- Cloud computing gives users convenient access to IT services, including networks, servers, storage, applications, and services, like using utilities such as water and electricity.
- The prerequisite for convenient, on-demand access to cloud resources is network connectivity.
- Rapid resource provisioning and reclamation fall into the rapid elasticity characteristic of cloud computing, while minimal management effort and service provider interaction the on-demand self-service characteristic.

Cloud computing definition

- The cloud is a metaphor for the Internet. It is an abstraction of the Internet and the infrastructure that underpins it.
- Computing refers to computing services provided by a sufficiently powerful computer capable of providing a range of functionalities, resources, and storage.
- Put together, cloud computing can be understood as the delivery of on-demand, measurable computing services over the Internet.

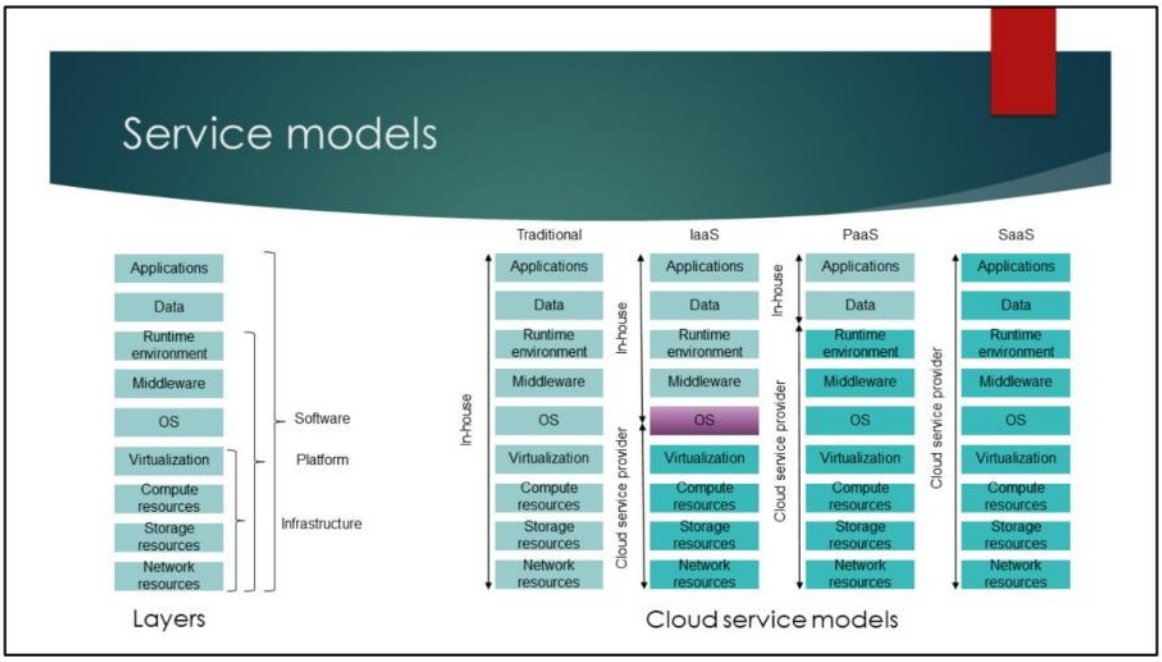


Cloud computing is categorized based on cloud deployment and service models.

Public clouds are usually built and run by cloud service providers. End users access cloud resources or services on a subscription basis while the service provider takes all administration responsibilities.

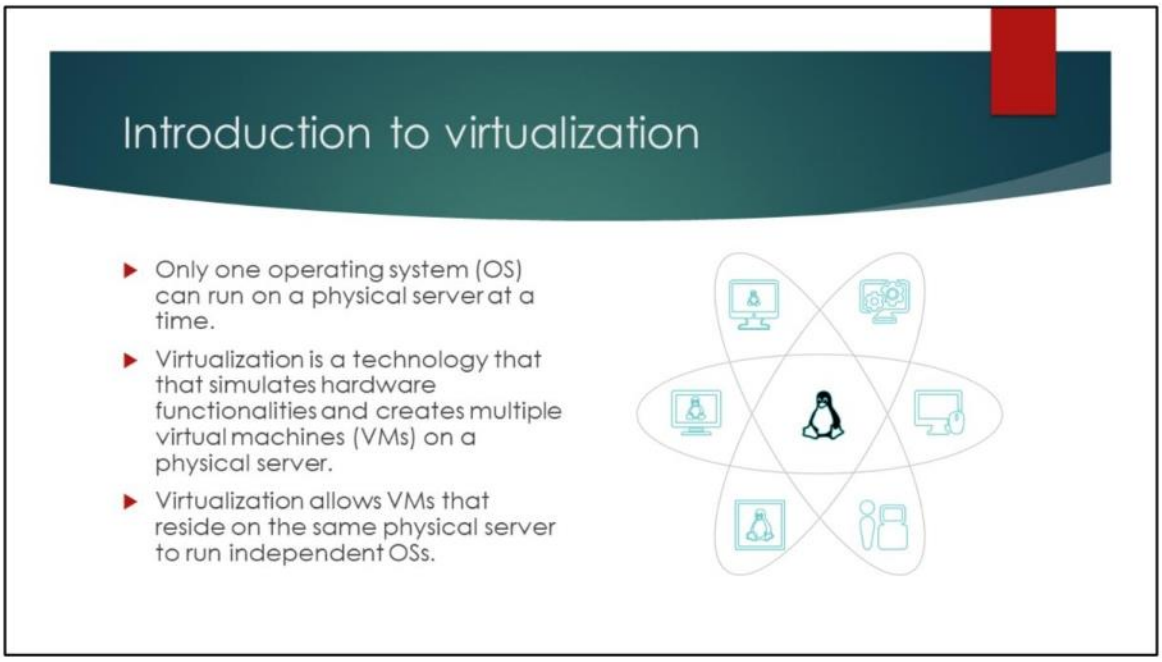
Private clouds are usually deployed for internal use within enterprises. All the data of a private cloud is stored in the enterprise or organization's own data center. The data center's firewalls control access to the data. A private cloud may be able to deliver a higher level of data security and allow reuse of legacy equipment. On the other hand, stricter data access control also means less data sharing.

Hybrid cloud is a flexible cloud deployment model. It may comprise two or more different types of clouds. A hybrid cloud allows users to enjoy the benefits of both public and private clouds. The downside is that hybrid cloud usually requires more complex setup. A hybrid cloud may also need to address compatibility issues between heterogeneous infrastructure platforms.



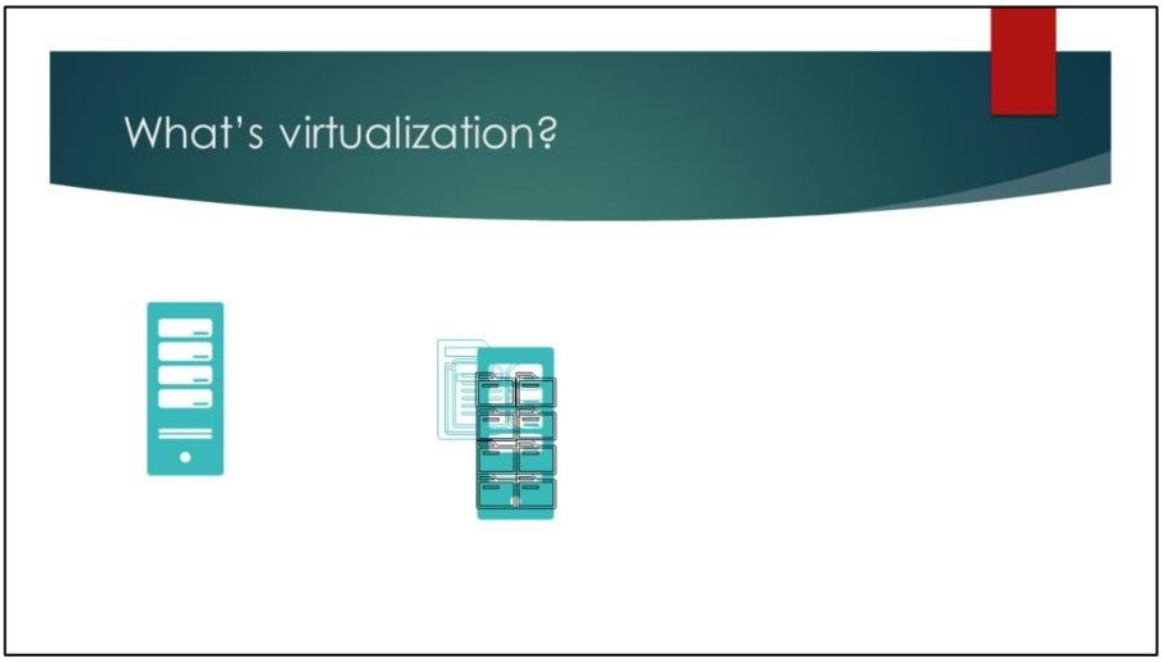
Infrastructure as a service (IaaS) refers to a situation where a cloud service provider provides and manages the infrastructure layer while the consumer the other layers.
Platform as a service (PaaS) refers to a situation where the cloud service provider manages the infrastructure and platform layers while the consumer the application layer.
Software as a service (SaaS) means all three layers are managed by the provider.

Let's explain these three cloud service models using an example of a game.
If we buy a computer of the required specifications, install an OS and then this game, this is not cloud computing. If we buy a cloud server of the same specifications from a public cloud provider, use an image to install an OS, download and then install the game, we're using the IaaS model.
If we buy not only the cloud server but also the ready-to-go runtime environment with the .NET Framework already installed, we're using the PaaS model.
If we buy the cloud server with the game and all the necessary software already installed and all we need to do to start playing the game is to enter our user name and password, we're using the SaaS model.



Virtualization is the process of creating **software-based version** of something (server, compute, storage, networks).

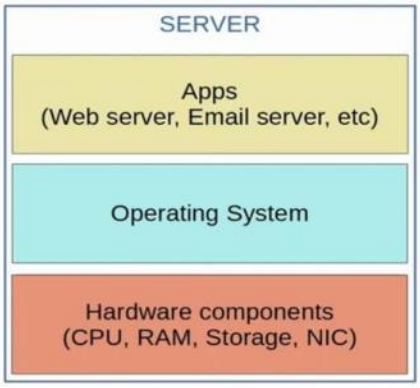
Virtualization is a technology that separates applications from the underlying operating systems(OSs) and hardware and creates multiple VMs on a physical server. Generally, applications need to run inside an OS, and only one OS can run on a physical server at a time. Virtualization allows VMs that reside on the same physical server to run independent OSs. This way, **multiple OSs can concurrently run** on the same physical server.



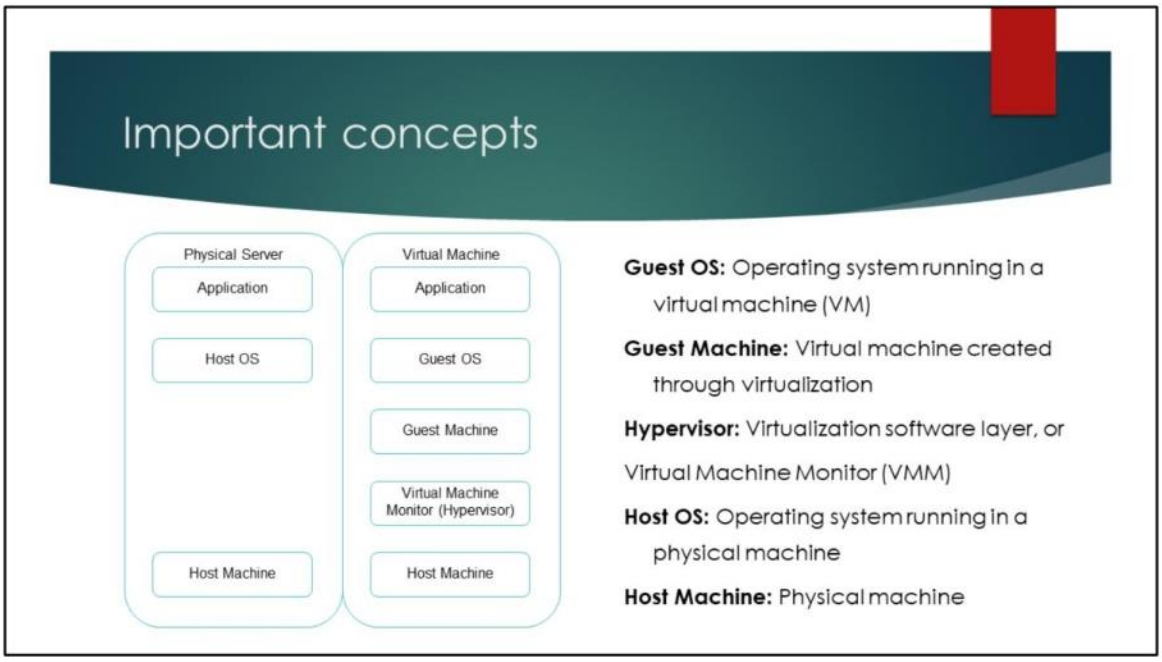
The essence of virtualization is to separate software from hardware by converting "physical“ devices into "logical" folders or files. These folders or files can be divided into two parts: those that store VM configuration information, and those that store user data.

With virtualization, multiple VMs can run on a single physical server, and each VM can run an independent OS. This **improves resource utilization**.

Servers before Virtualization



- Before virtualization, there was a one-to-one relationship between a physical server and an operating system.
- In that operating system, apps providing services such as a web server, email server, etc. would run.
- One physical server would be used for the web server, one for the email server, one for the database server, etc.
- This is inefficient for multiple reasons:
  → Each physical server is expensive and takes up space, power, etc.
  → The resources on each physical server (CPU, RAM, Storage, NIC) are typically under-used.



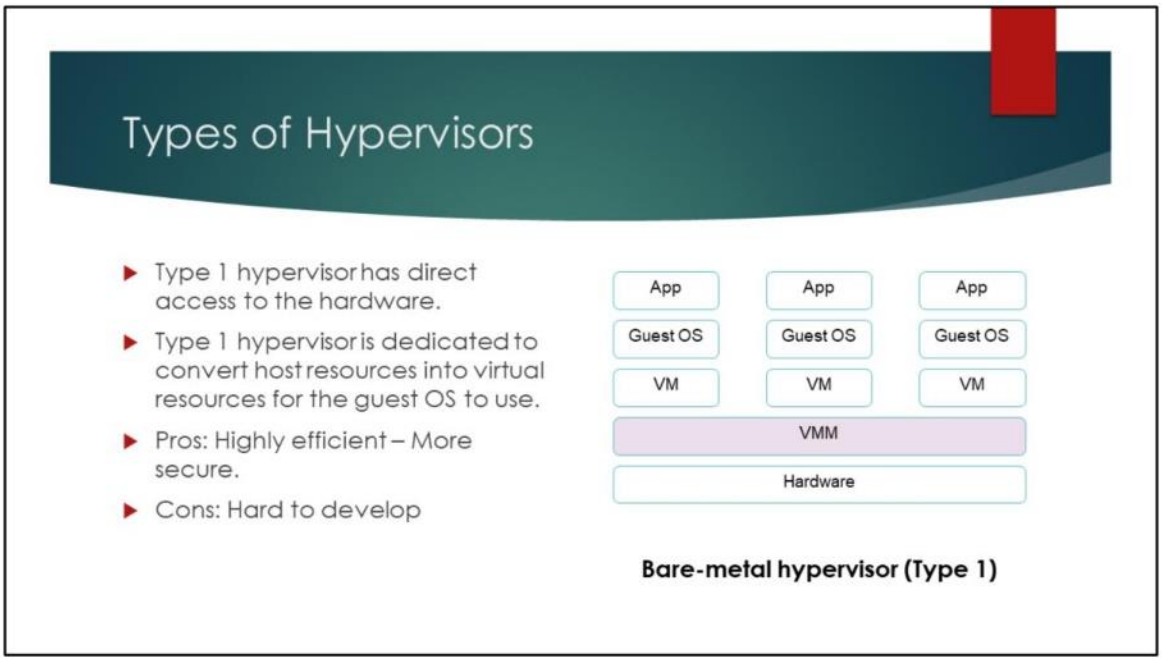
Let's learn some terms that are commonly used in virtualization.

First, a host machine is a physical computer that can run multiple VMs, and an OS installed and running on the host machine is a host OS.

VMs running on a host machine are called guest machines. The OS installed on VMs is called a guest OS.

The core of virtualization technology between the host OS and guest OS is a hypervisor, which is sometimes called virtual machine manager (VMM).

Hypervisor pools the resources from the host machine (physical server) and allocates them to the VMs.

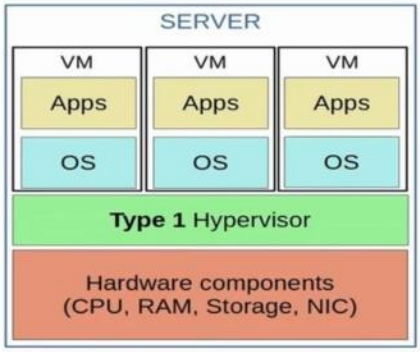


There are two types of hypervisors: Type 1 and Type 2. Many people categorize containers as the third type of hypervisor.

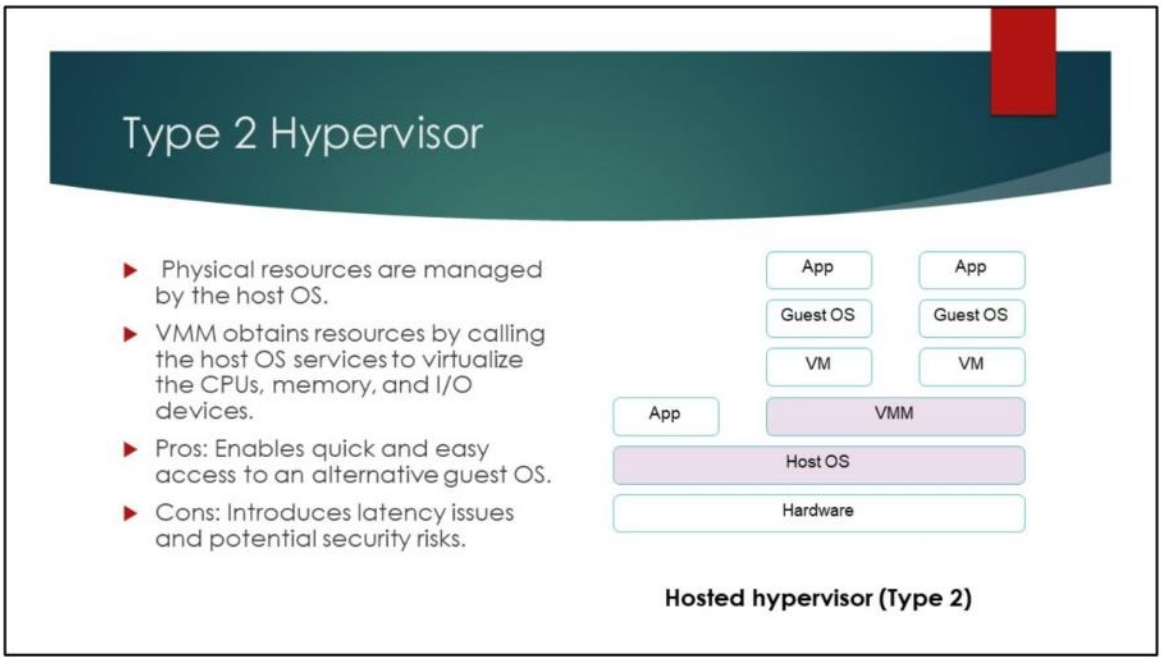
A **Type 1 hypervisor** is also called a **bare-metal hypervisor**. This type of hypervisor has direct access to hardware resources and does not need to access the host OS. The hypervisor can be seen as a **customized host OS**, which merely functions as VMM and does not run other applications.

Pros: Type 1 hypervisors are **highly efficient** because they have direct access to physical hardware. This also **increases their security**, because there is nothing in between them and the CPU that an attacker could compromise.
Cons: The kernel of the virtualization layer is **hard to develop**.

Type 1 Hypervisor



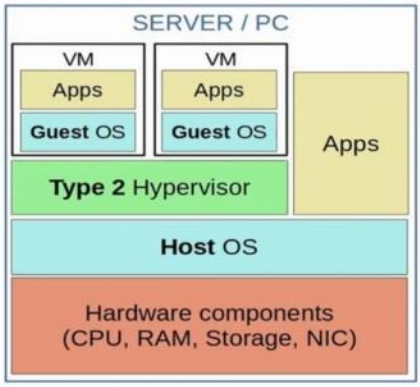
- Virtualization allows us to break the one-to-one relationship of hardware to OS, allowing multiple OS's to run on a single physical server.
- Each instance is called a VM (Virtual Machine).
- A **hypervisor** is used to manage and allocate the hardware resources (CPU, RAM, etc) to each VM.
- Another name for a hypervisor is VMM (Virtual Machine Monitor).
- The type of hypervisor which runs directly on top of the hardware is called a **Type 1** hypervisor.
  → Examples include VMware ESXi, Microsoft Hyper-V, etc.
- Type 1 hypervisors are also called *bare-metal hypervisors* because they run directly on the hardware (metal).
  → Another term is *native hypervisor*
- This is the type of hypervisor used in data center environments.



A **Type 2 hypervisor** is also called a **hosted hypervisor**. Physical resources are managed by the host OS (for example, Windows or Linux). VMM provides virtualization services and functions as a common application in the underlying OS (for example, Windows or Linux). VMM obtains resources by calling the host OS services to virtualize the CPUs, memory, and I/O devices.

Pros: A Type 2 hypervisor enables quick and **easy access to an alternative guest OS** alongside the primary one running on the host system. This makes it great for end-user productivity.
Cons: A Type 2 hypervisor must access computing, memory, and network resources via the host OS. This introduces **latency issues**, affecting performance. It also introduces **potential security risks** if an attacker compromises the host OS.

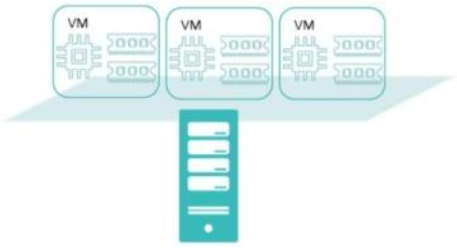
Type 2 Hypervisor



- **Type 2** hypervisors run as a program on an operating system like a regular computer program.
  → Examples include VMware Workstation, Oracle VirtualBox, etc.
- The OS running directly on the hardware is called the **Host OS**, and the OS running in a VM is called a **Guest OS**.
- Another name for a Type 2 hypervisor is *hosted hypervisor*.
- Although Type 2 hypervisors are rarely used in data center environments, they are common on personal-use devices (for example, if a Mac/Linux user needs to run an app that is only supported on Windows, or vice versa).

Characteristics of Virtualization

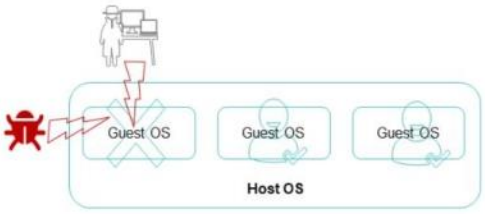
- Partitioning: indicates the VMM capability of allocating server resources to multiple VMs.
- Resource quotas are allocated to each partition to prevent resource overuse by virtualization.



Partitioning

Characteristics of Virtualization

- Isolation: Multiple VMs are logically isolated from each other.
- If one VM crashes due to an OS failure, application breakdown, or driver failure, t should not affect the others on the same server.
- If a VM is infected with worms or viruses, the worms and viruses are isolated from other VMs.



Isolation

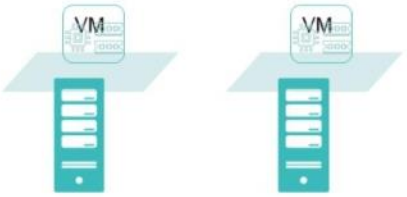
Characteristics of Virtualization

- Encapsulation: Each VM is saved as a group of files.
- You can copy, save, and move a VM by copying only a few files.
- Encapsulation is the most important feature for VM migration.

Encapsulation

Characteristics of Virtualization

- Hardware independence: A VM is completely independent from its underlying hardware.
- The underlying hardware device is shielded by VMM running on it.



HW Independence