




Why is it important?

# Virtualization

Relevant Link:

[https://  
web.stanford.edu/  
class/cs240/old/  
sp2014/readings/  
ASPLOS06-  
Keynote-noan.pdf](https://web.stanford.edu/class/cs240/old/sp2014/readings/ASPLOS06-Keynote-noan.pdf)

- 
- Creating a virtual version of any computing resource
  - Reduced capital and operating cost
  - Minimize downtime
  - Provision resources faster in an automated way
  - + Business continuity and faster disaster recovery

## How it works

- Uses software to simulate existence of hardware and create a virtual computer system that runs on an existing physical server
- This helps us overcome the drawback of a bare metal server - only one application can run on a machine with a fixed OS

# What is a VM ?

- A **tightly isolated software container** with an OS and application inside
- Each self-contained VM is completely independent
- Putting **Multiple VMs on a single physical host** improves efficiency and effective use of available computing resources
- **Hypervisor** is the software that **decouples** the VMs from the host and **dynamically allocates computing resources** to each VM as needed and ensures each VM operates individually **without interference** from the other VMs running on the same host.

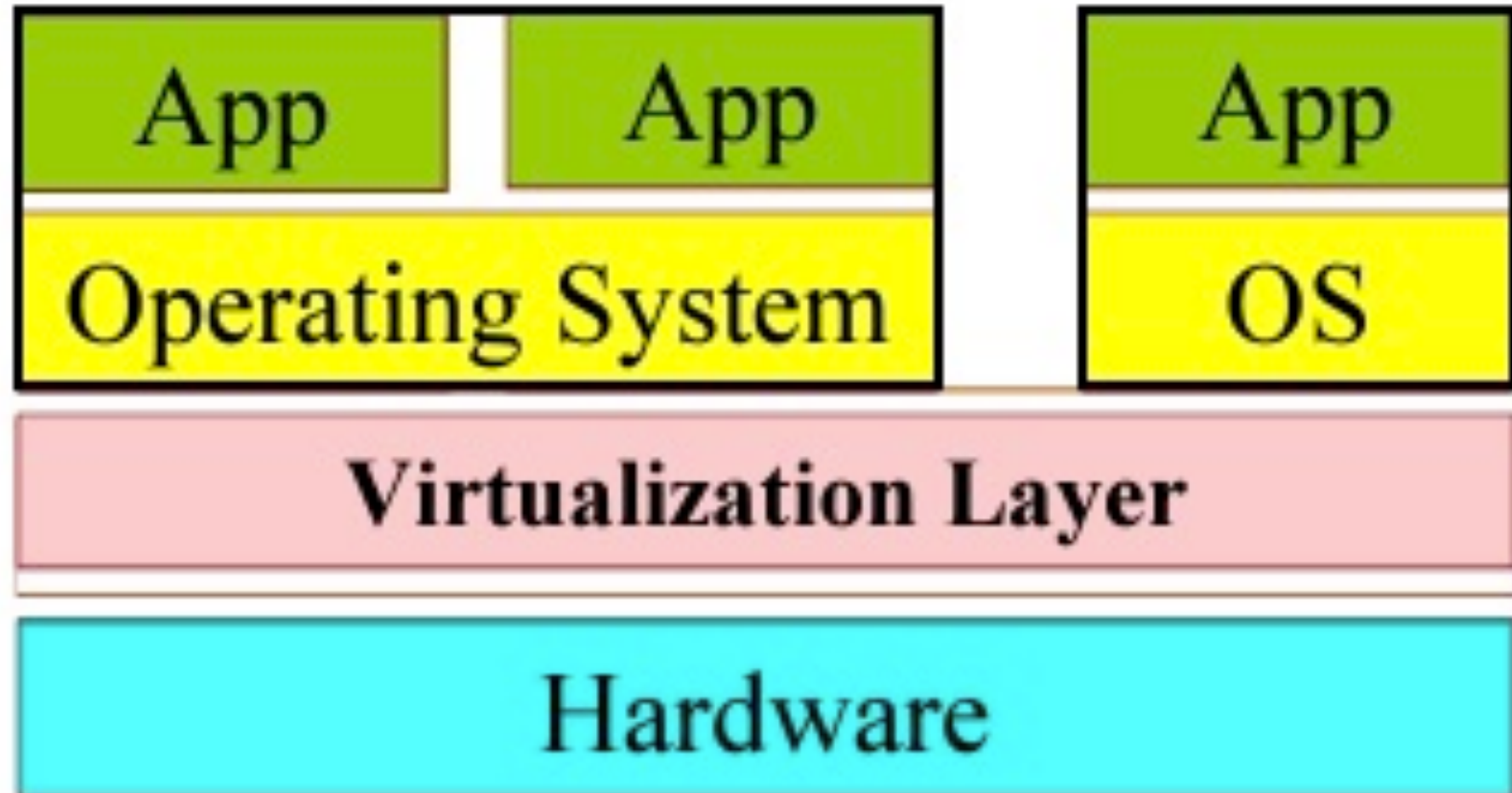
Virtualization as the evolution of Operating Systems:

[https://www.researchgate.net/publication/](https://www.researchgate.net/publication/268349283_Virtualization_as_the_Evolution_of_Operating_Systems)

[268349283\\_Virtualization\\_as\\_the\\_Evolution\\_of\\_Operating\\_Systems](https://www.researchgate.net/publication/268349283_Virtualization_as_the_Evolution_of_Operating_Systems)

**VM-1**

**VM-2**



# Different Types of Virtualization

## Server Virtualization

- Partitions a physical server into multiple virtual servers
- Cost effective way to use hardware resources

## Storage Virtualization

- Combines physical storage devices such as Network Attached Storage (NAS) and Storage Area Network (SAN).
- Pools all your physical data storage and creates a large unit of virtual storage

Application Virtualization pulls out the functions of applications to run on OS other than the OS for which they were designed

- Server based - Users access a remote application from their browser without installing it
- Local - The application code is shipped with its own environment ( Docker )

Desktop Virtualization : Running Windows on Linux machine / Accessing a windows VM through a client

## Virtualization

VS

## Cloud Computing

Virtualization is the technology that makes cloud computing possible

On-demand delivery of computing resources over the internet with pay-as-you-go pricing.

## Server Virtualization

VS

## Containerization

Running multiple instances of isolated and independent software created computers that mimic a physical server having its own operating system and resources like a CPU, RAM and Storage

Containerization is a kind of application virtualization where we pack all the dependencies of the application along with an Operating system so it can be run on any physical / virtual server

# Infra as a Service

Amazon Elastic Compute Cloud & Google Compute Engine

## Elastic Compute Cloud (EC2)

- Provides secure, resizable **compute capacity** in the cloud
- Reduces time to obtain and **boot new server / take down existing instances** to minutes
- We only **pay for the capacity** that we actually use

## Google Compute Engine (GCE)

- Ability to **create VMs** using publicly available Linux and Windows images or custom built images

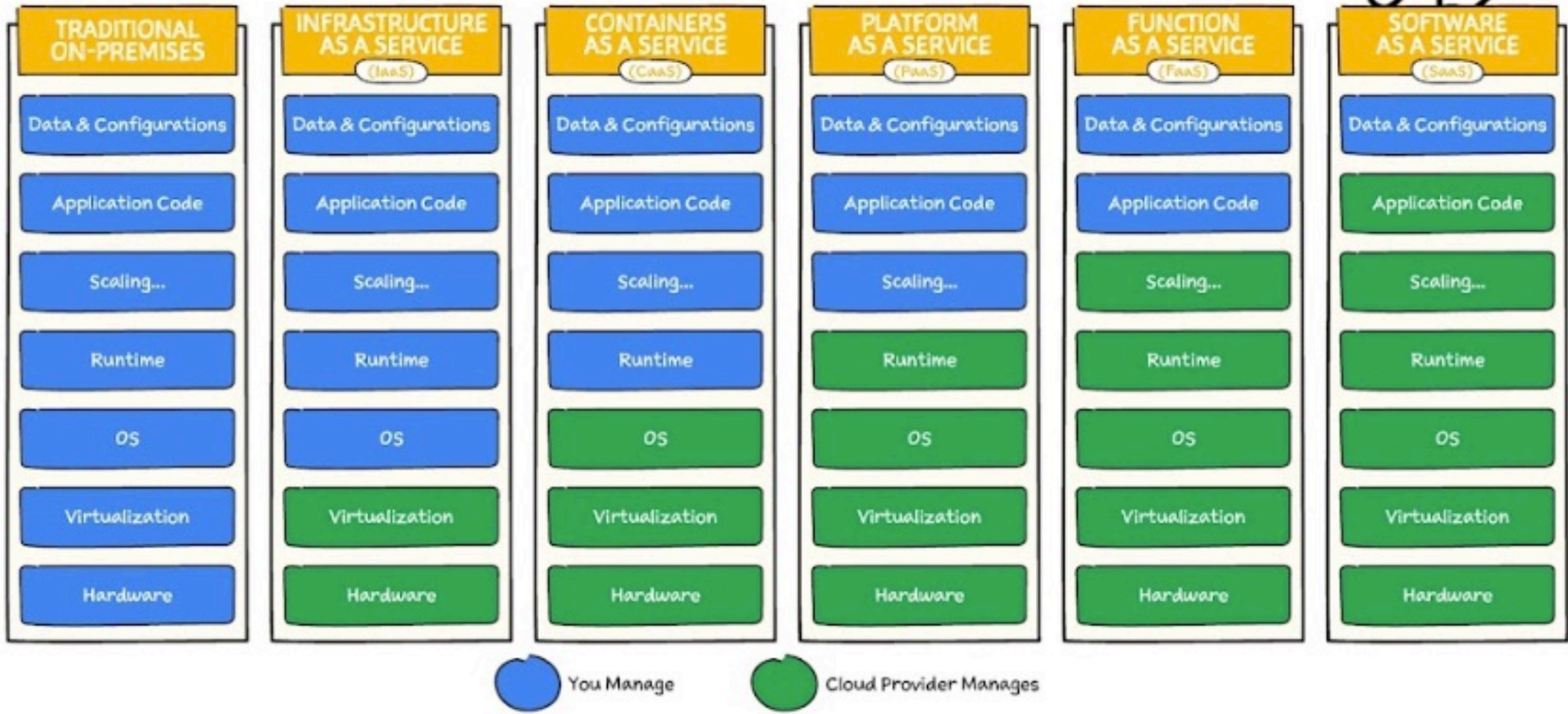


#GCPsketchnote

@PVERGADIA  
THECLOUDGIRL.DEV  
05.11.2021



# Wait... what is Cloud again?







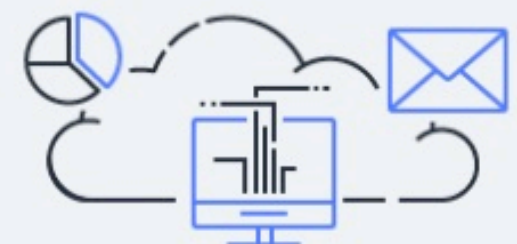
## Infrastructure as a Service (IaaS)

Infrastructure as a Service, sometimes abbreviated as IaaS, contains the basic building blocks for cloud IT and typically provide access to networking features, computers (virtual or on dedicated hardware), and data storage space. Infrastructure as a Service provides you with the highest level of flexibility and management control over your IT resources and is most similar to existing IT resources that many IT departments and developers are familiar with today.



## Platform as a Service (PaaS)

Platforms as a service remove the need for organizations to manage the underlying infrastructure (usually hardware and operating systems) and allow you to focus on the deployment and management of your applications. This helps you be more efficient as you don't need to worry about resource procurement, capacity planning, software maintenance, patching, or any of the other undifferentiated heavy lifting involved in running your application.



## Software as a Service (SaaS)

Software as a Service provides you with a completed product that is run and managed by the service provider. In most cases, people referring to Software as a Service are referring to end-user applications. With a SaaS offering you do not have to think about how the service is maintained or how the underlying infrastructure is managed; you only need to think about how you will use that particular piece of software. A common example of a SaaS application is web-based email where you can send and receive email without having to manage feature additions to the email product or maintaining the servers and operating systems that the email program is running on.

	IaaS	PaaS	SaaS
It's like	<p>Doing all the IT stuff ( fun stuff ) on the cloud, you handle everything</p> <p><small>Instead of buying and maintaining physical hardware and networking infra we rent it from the cloud provider</small></p>	<p>All the “fun stuff” is done. Just deploy the application and run it, scale it as you like, take it down...</p>	<p>You just Just chill and use the end product</p>
Example	<p>Google owns the physical hardware, but we use terraform to create all the VMS and networking stuff to set up our system</p>	<p>Using cloud run to host and run our applications. It</p> <ul style="list-style-type: none"> <li>• automatically scales up based on demand and we link a domain name to it, only worry about latest code being deployed to it from the pipeline.</li> </ul>	<p>I sent the email, but it didn't reach you, what's can I do ? not my problem</p>

# VM Machine Types

1. **General Purpose** : Balance of compute, memory and networking resources, for a variety of diverse workloads. Hosting web applications, code repos
2. **Compute Optimized (CPU)** : High performance processors - batch processing, HPC, scientific computing, media transcoding, gaming servers
3. **Memory Optimized (RAM)**: Fast performance to process large datasets in-memory, in-memory caches, realtime big data analytics
4. **Accelerated Computing (GPU)**: Use hardware accelerators or co-processors (graphic cards) perform functions such as graphics processing, data pattern matching. Generative AI applications, code, video, image generation, speech recognition and more.
5. **Storage Optimized (SSD)** : For large IOPS, transactional database applications

# Instance Purchasing Options

## # Summary

1. On Demand - Short term uninterrupted work loads where you can't predict application behavior
2. Reserved instances - long term steady state applications like a database
3. Flexible reserved instances - long term but attributes can be changed
4. Savings plan - Budgeted instances, usage above budget charged at On Demand rate - locked to instance family and region - flexible in size, os and tenancy
- 5 Spot instances - cheapest, can lose instance any time, fault tolerant, non critical and stateless applications
6. Capacity Reservation - Definitely available on demand at a specific AZ

- **On demand:** coming and staying in resort whenever we like, we pay the full price *for duration of stay*
- **Reserved:** like planning ahead and if we plan to stay for a long time, we may get a good discount.
- **Savings Plans:** pay a certain amount per hour for certain period and stay in any room type (e.g., King, Suite, Sea View, ...)
- **Spot instances:** the hotel allows people to bid for the empty rooms and the highest bidder keeps the rooms. You can get kicked out at any time
- **Dedicated Hosts:** We book an entire building of the resort
- **Capacity Reservations:** you book a room for a period with full price even you don't stay in it

- **On Demand**: Recommended for short term uninterrupted workloads where you can't predict application behavior ? Which case is this
- **Reserved**: Steady state applications like a database - reservation period - cheaper - we can reserve capacity, tenancy, region and OS
- **Spot** - Most cost efficient instances - can lose anytime if spot price goes above max price- spare resources that can be reclaimed by other jobs with higher priority like on demand or reserved loads - recommended for fault tolerant, stateless and time flexible tasks



# Tenancy

Refers to how the hardware is used to support the virtual machines offered by a cloud provider

## Dedicated Host

Get access to the physical server itself for each instance and we get visibility into the lower level hardware

## Dedicated Instance

Multiple EC2 instances of the Account can use the same hardware, instances of other accounts cannot have access to that hardware. Dedicated instances can be used cases that have tenancy restrictions due to security and other compliance restriction.

## Default

Run on hardware that is shared by other instances that belong to other accounts