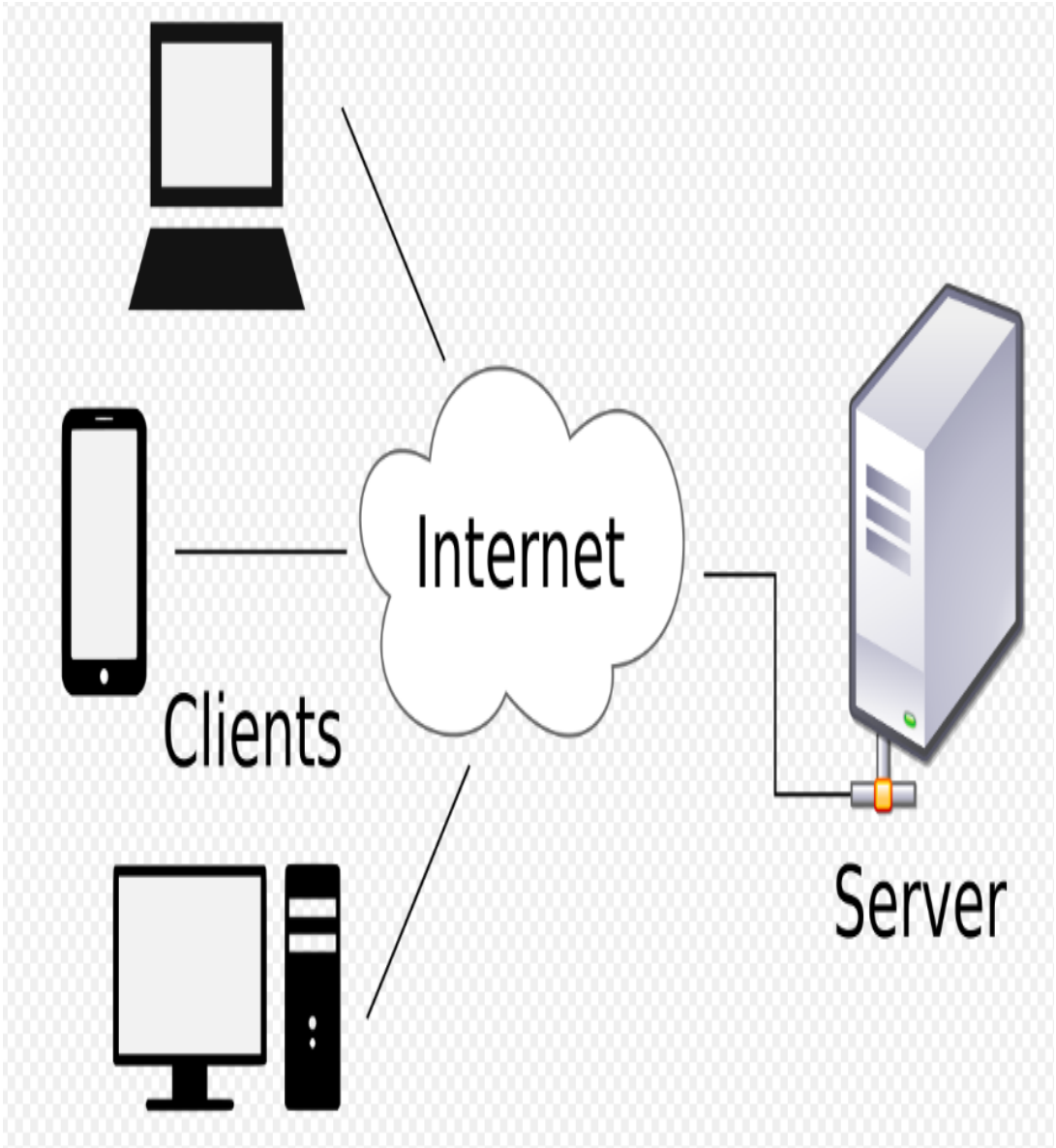
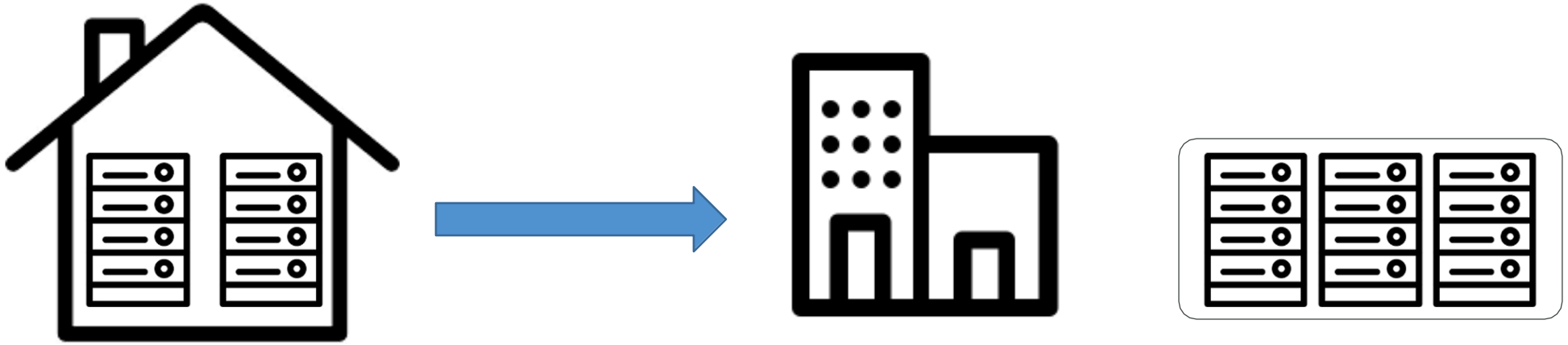
Virtualization & Cloud Computing

Life before Virtualization

- To run App/Service we need servers
- Physical computers(Servers in Datacenter)
- One Service – One Server(for Isolation)
- Servers are always overprovisioned
- Server Resources mostly underutilized
- Huge Capital Expenditure & Operational Expenditure



Traditionally, how to build Infrastructure



Home or Garage

Office

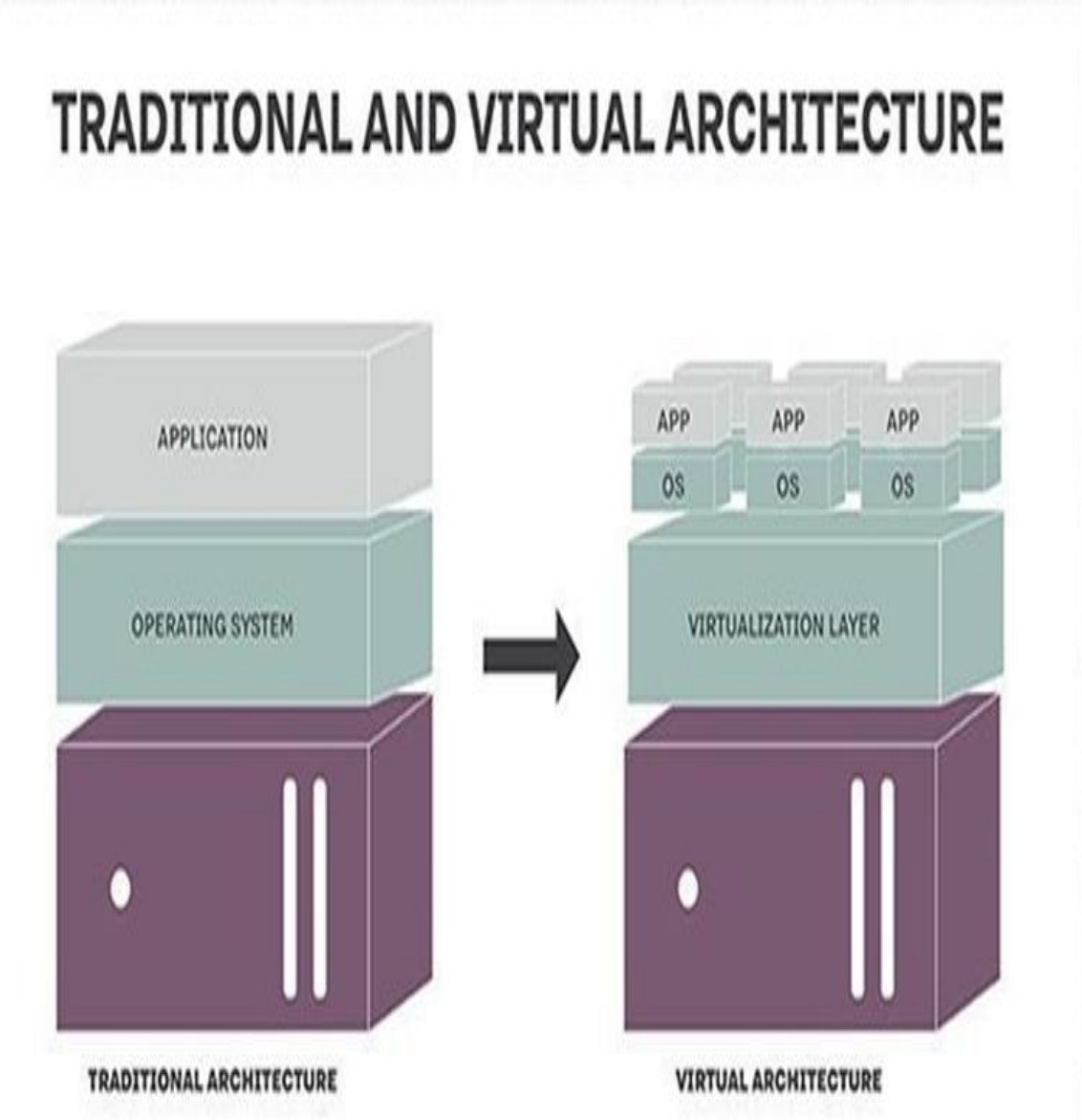
Data center

Problems with Traditional IT approach

- Pay for the rent for the data center
- Pay for power supply, cooling, maintenance
- Adding and replacing hardware takes time
- Scaling is limited
- Hire 24/7 team to monitor the infrastructure
- How to deal with disasters? (earthquake, power shutdown, fire…)

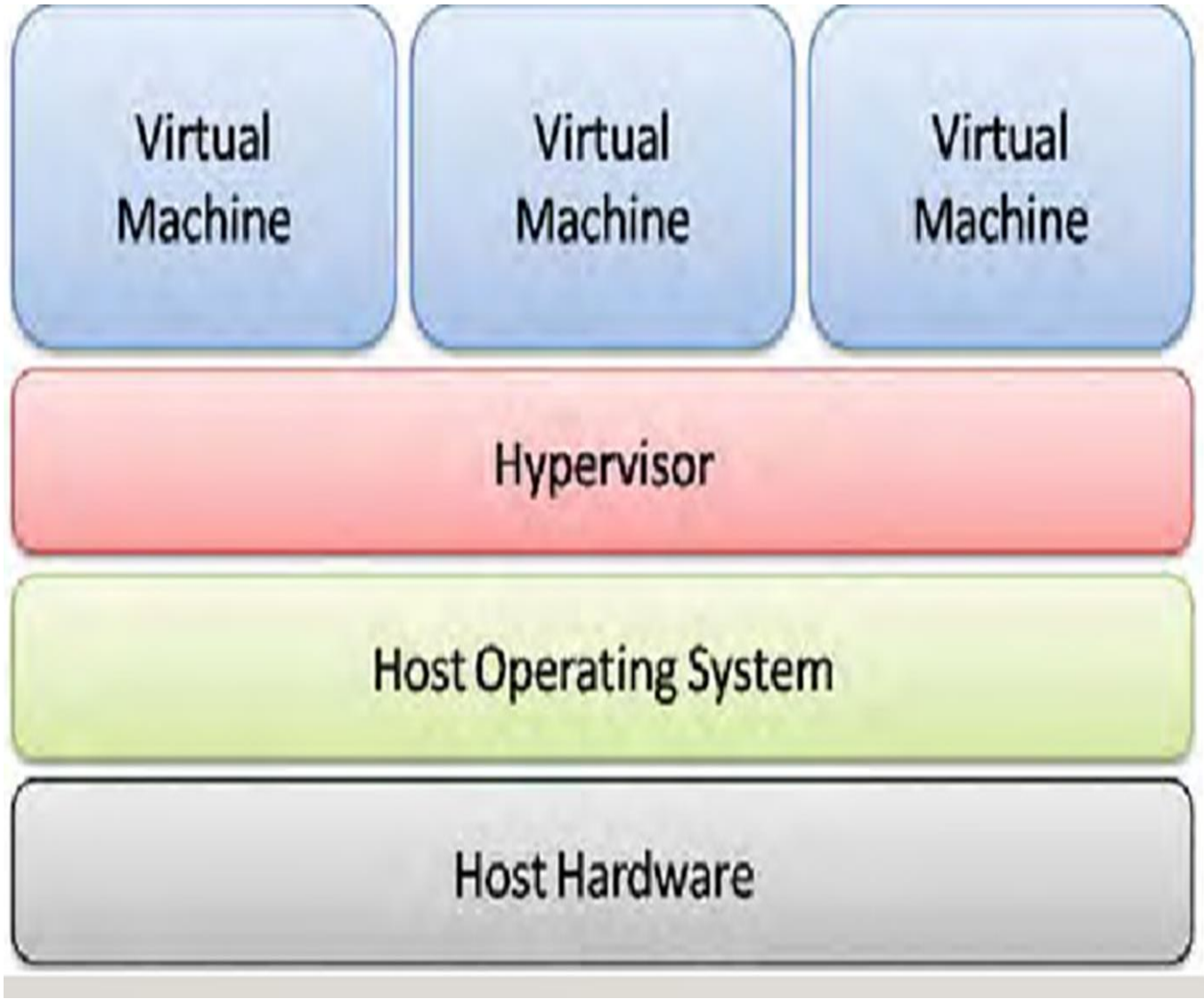
What is Virtualization?

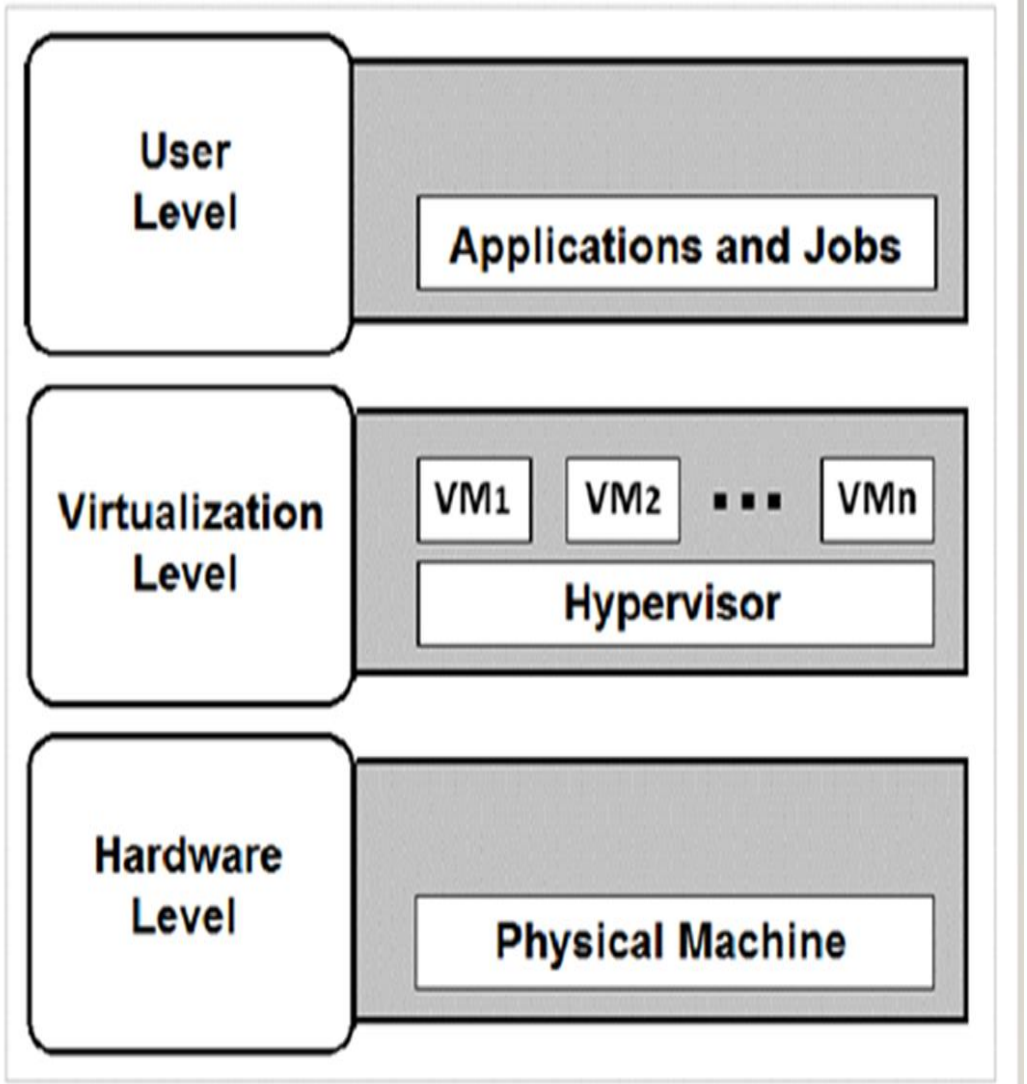
- Virtualization is technology that lets you create useful IT services using resources that are traditionally bound to hardware.
- It allows you to use a physical machine's full capacity by distributing its capabilities among many users or environments.
- Virtualization is a technique of importing a Guest OS on the top of the host OS.
- This technique was a revelation at the beginning as it allowed developers to run multiple OS in different VMs on the same physical host.



How does Virtualization work?

- Software called hypervisors separate the physical resources from the virtual environments—the things that need those resources.
- Hypervisors can sit on top of an operating system (like on a laptop) or be installed directly onto hardware (like a server) which is how most enterprises virtualize.
- Hypervisors take your physical resources and divide them up so that virtual environments can use them.
- A hypervisor, also known as a virtual machine monitor or VMM, is a software that creates and runs virtual machines (VMs).


Virtual Machine
Virtual Machine
Virtual Machine
Hypervisor
Host Operating System
Host Hardware


User Level
Applications and Jobs
Virtualization Level
VM1
VM2
VMn
Hypervisor
Hardware Level
Physical Machine

Types of Hypervisors

- There are two main hypervisor types, referred to as “Type I” (or “bare metal”) and “Type II” (or “hosted”).

- A **type I** hypervisor acts like a lightweight operating system and runs directly on the host’s hardware, while a **type II** hypervisor runs as a software layer on an operating system like other computer programs.

- A **type II** hypervisor is called a hosted hypervisor because it is installed on an existing operating system, and they are not more capable of running more complex virtual tasks. People use it for basic development, testing and simulation.

Types of Virtualization

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

Types of Virtualization

1) 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.

Usage:

Hardware virtualization is mainly done because controlling virtual machines is much easier than controlling a physical server.

2) 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.

Usage:

Operating System Virtualization is mainly used for testing the applications on different platforms of OS.

3) Server Virtualization:

When the virtual machine software or virtual machine manager *(VMM) is directly installed on the Server system* is known as server virtualization.

Usage:

Server virtualization is done because a single physical server can be divided into multiple servers on the demand basis and for balancing the load.

4) Storage Virtualization:

Storage virtualization is the *process of grouping the physical storage from multiple network storage devices so that it looks like a single storage device*.

Usage:

Storage virtualization is mainly done for back-up and recovery purposes.

Virtualization Terminologies

Virtualization – Refers to the technology that allows for the creation of software-based virtual machines that can run multiple operating systems from a single physical machine. This allows a single physical machine to run multiple virtual machines which in turn may run their own operating systems to serve individual dedicated purposes.

Host Machine – The physical machine that hosts one or more virtual machines. To accomplish this, virtualization software such as a hypervisor is installed on the Host Machine.

Hypervisor – The software or firmware that manages virtual machines, allowing them to interact directly with the underlying hardware. The hypervisor is an operating platform which manages and executes the Guest VM operating systems.

Virtualization Terminologies

Virtual Machine (Guest VM) – A self-contained software emulation of a machine, which does not physically exist, but shares resources of an underlying physical machine. It runs its own operating system, applications, processes , etc.

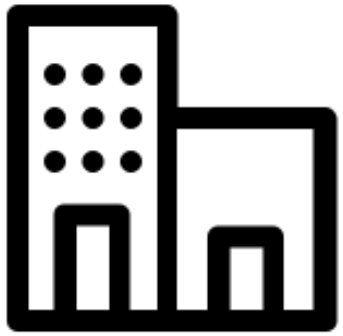
VM Cluster – A collection of VM Hosts that act as a single large host. If one of the hosts is removed, all of the VMs that the host was running seamlessly continue running on the other hosts. A true VM cluster requires shared storage such as a SAN/NAS device.

TOP 10 Virtualization Vendors

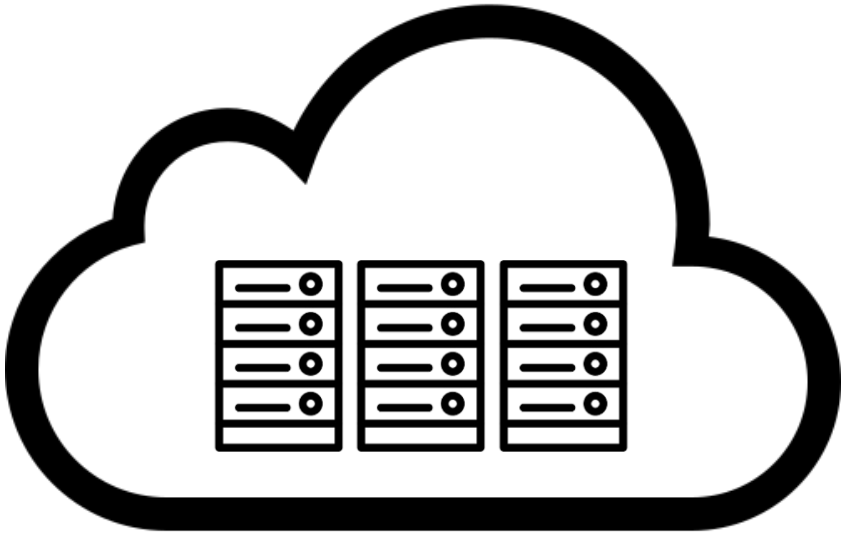
1. VMware
2. Microsoft
3. Citrix
4. Red Hat
5. Oracle
6. Amazon
7. Google
8. Parallel / Odin
9. Huawei
10. Verde VDI

What is Cloud Computing?

- Cloud computing is the on-demand delivery of compute power, database storage, applications, and other IT resources
- Through a cloud services platform with pay-as-you-go pricing
- You can provision exactly the right type and size of computing resources you need
- You can access as many resources as you need, almost instantly
- Simple way to access servers, storage, databases and a set of application services.



Office



The Cloud

Why Cloud Computing?

Cost

Cloud computing eliminates the capital expense of buying hardware and software and setting up and running on-site datacenters—the racks of servers, the round-the-clock electricity for power and cooling, and the IT experts for managing the infrastructure. It adds up fast.

Speed

Most cloud computing services are provided self service and on demand, so even vast amounts of computing resources can be provisioned in minutes, typically with just a few mouse clicks, giving businesses a lot of flexibility and taking the pressure off capacity planning.

Global scale

The benefits of cloud computing services include the ability to scale elastically. In cloud speak, that means delivering the right amount of IT resources—for example, more or less computing power, storage, bandwidth—right when they're needed, and from the right geographic location.

Productivity

On-site datacenters typically require a lot of "racking and stacking"—hardware setup, software patching, and other time-consuming IT management chores. Cloud computing removes the need for many of these tasks, so IT teams can spend time on achieving more important business goals.

Performance

The biggest cloud computing services run on a worldwide network of secure datacenters, which are regularly upgraded to the latest generation of fast and efficient computing hardware. This offers several benefits over a single corporate datacenter, including reduced network latency for applications and greater economies of scale.

Reliability

Cloud computing makes data backup, disaster recovery, and business continuity easier and less expensive because data can be mirrored at multiple redundant sites on the cloud provider's network.

Security

Many cloud providers offer a broad set of policies, technologies, and controls that strengthen your security posture overall, helping protect your data, apps, and infrastructure from potential threats.

You've been using some Cloud services

Gmail

- E-mail cloud service
- Pay for ONLY your emails stored (no infrastructure, etc.)

Dropbox

- Cloud Storage Service
- Originally built on AWS

Netflix

- Built on AWS
- Video on Demand

The Deployment Models of the Cloud

Private Cloud:

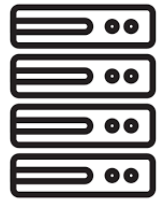
- Cloud services used by asingle organization, not exposed to the public.
- Complete control
- Security for sensitive applications
- Meet specific business needs

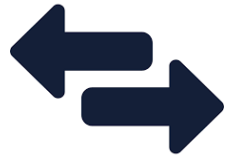
Public Cloud:

- Cloud resources owned and operated by a third-party cloud service provider delivered over the Internet.
- Six Advantages of Cloud Computing

Hybrid Cloud:

- Keep some servers on premises and extend some capabilities to the Cloud
- Control over sensitive assets in your private infrastructure
- Flexibility and cost-effectiveness of the public cloud





Features of Cloud Computing

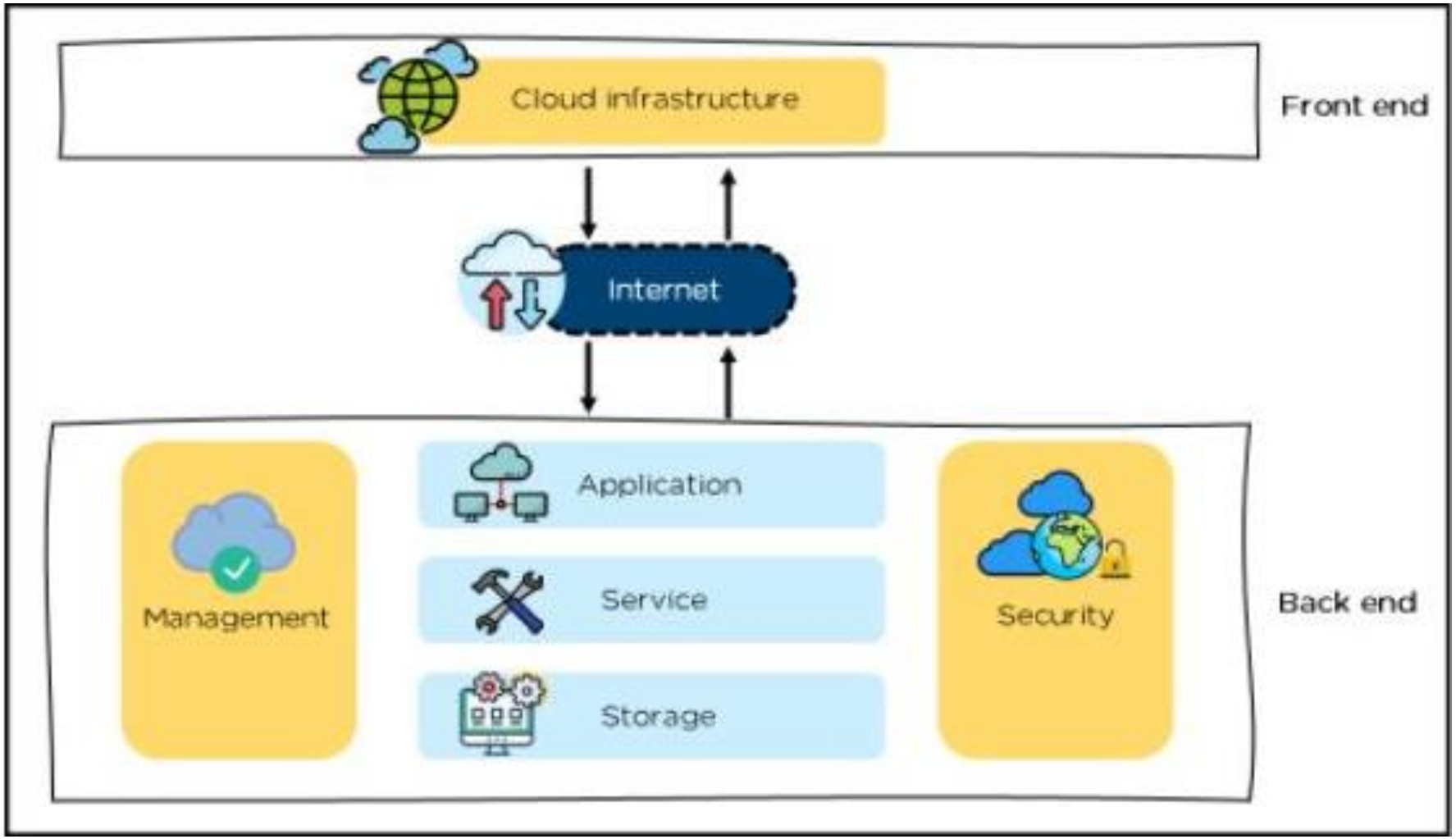
- On-demand self service:
  - Users can provision resources and use them without human interaction from the service provider
- Broad network access:
  - Resources available over the network, and can be accessed by diverse client platforms
- Multi-tenancy and resource pooling:
  - Multiple customers can share the same infrastructure and applications with security and privacy
  - Multiple customers are serviced from the same physical resources
- Rapid elasticity and scalability:
  - Automatically and quickly acquire and dispose resources when needed
  - Quickly and easily scale based on demand
- Measured service:
  - Usage is measured, users pay correctly for what they have used

Problems solved by Cloud

- Flexibility: change resource types when needed
- Cost-Effectiveness: pay as you go, for what you use
- Scalability: accommodate larger loads by making hardware stronger or adding additional nodes
- Elasticity: ability to scale out and scale-in when needed
- High-availability and fault-tolerance: build across data centers
- Agility: rapidly develop, test and launch software applications

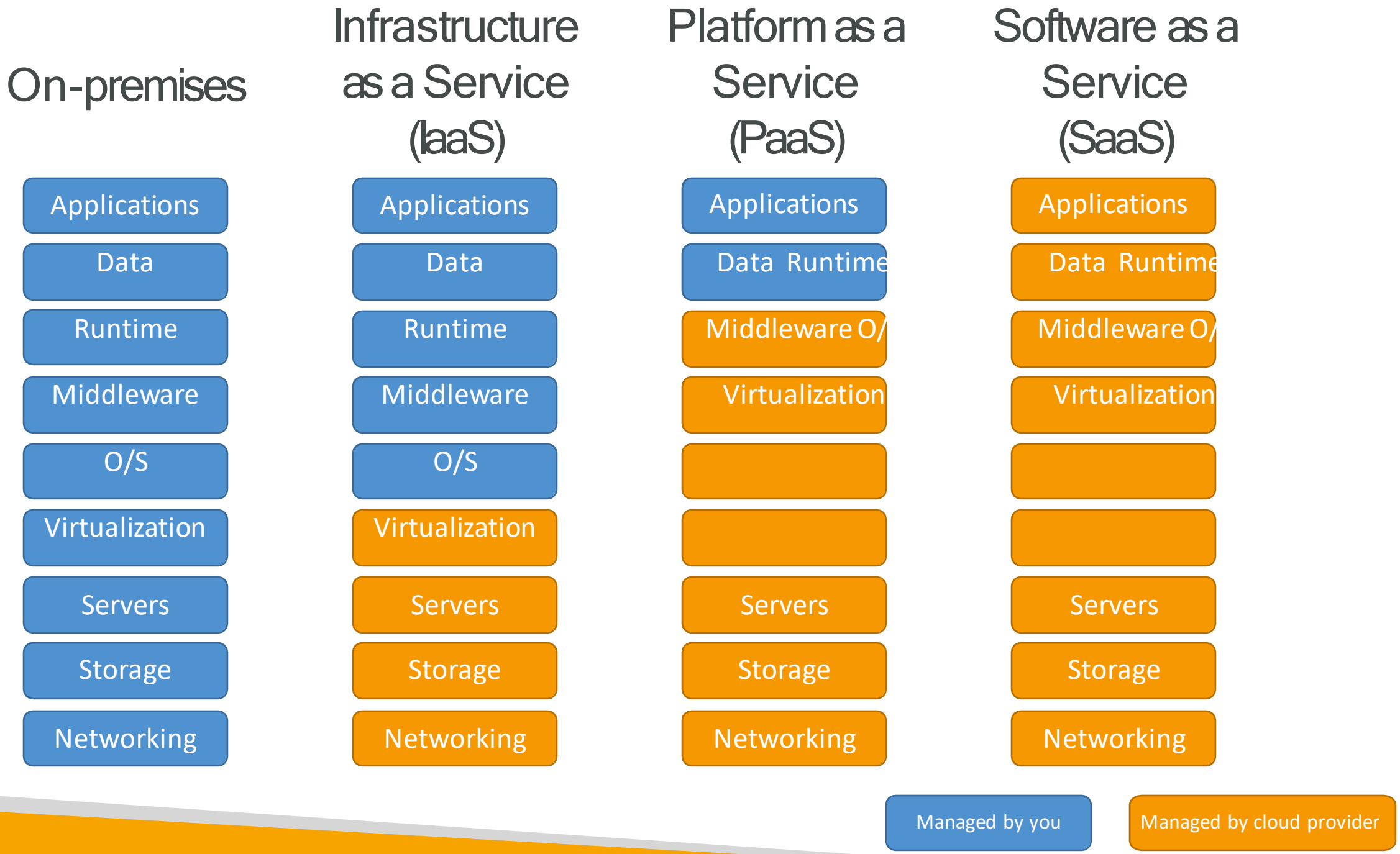
Cloud Computing Architecture

Cloud Computing Architecture is divided into two parts, i.e., front-end and back-end. Front-end and back-end communicate via internet.



Cloud Computing Service Models

- Infrastructure as a Service (IaaS)
  - Provide building blocks for cloud IT
  - Provides networking, computers, data storage space
  - Highest level of flexibility
  - Easy parallel with traditional on-premises IT
- Platform as a Service (PaaS)
  - Removes the need for your organization to manage the underlying infrastructure
  - Focus on the deployment and management of your applications
- Software as a Service (SaaS)
  - Completed product that is run and managed by the service provider


On-premises
Infrastructure as a Service (IaaS)
Platform as a Service (PaaS)
Software as a Service (SaaS)
Applications
Data
Runtime
Middleware
O/S
Virtualization
Servers
Storage
Networking
Applications
Data
Runtime
Middleware
O/S
Virtualization
Servers
Storage
Networking
Applications
Data Runtime
Middleware O/
Virtualization
Servers
Storage
Networking
Applications
Data Runtime
Middleware O/
Virtualization
Servers
Storage
Networking
Managed by you
Managed by cloud provider

Examples of cloud service models

- Infrastructure as a Service:
  - Amazon EC2 (on AWS)
  - GCP, Azure, Rackspace, Digital Ocean, Linode
- Platform as a Service:
  - Elastic Beanstalk (on AWS)
  - Google App Engine (GCP), Windows Azure (Microsoft)
- Software as a Service:
  - Many AWS services (ex: Rekognition for Machine Learning)
  - Google Apps (Gmail), Dropbox, Zoom

Cloud APIs

Cloud Application Programming Interface is a type of API that permits for the development of services and applications used for the provisioning of cloud platforms, software and hardware.

A cloud application programming interface works as an interface or gateways that provides direct and indirect cloud software and infrastructure services to users.

Popular Cloud APIs and different cloud implementers:

AWS APIs

- AWS Cloud Control - an integrated set of APIs designed to make it easy for developers to manage these five operations: create, read, update, delete and list (CRUD-L).
- Amazon API Gateway - a cloud service for service creating REST, HTTP, and WebSocket APIs at scale and maintaining them throughout their lifecycle.

Azure APIs

- Azure Communication Services - provides APIs for voice, video, chat, SMS and email. Requires applications to use the same infrastructure as Microsoft Teams.
- Azure API Management - provides developers with an API gateway, management plane and developer portal so they can expose services hosted on Azure as APIs.

Google Cloud APIs

Google Compute Engine API - used to create and run virtual machines (VMs) on Google Cloud.

Google Storage Transfer API - used to transfer data from an external source to Google Cloud.

Cloud computing benefits

- Trade capital expense (CAPEX) for operational expense (OPEX)
  - Pay On-Demand: don't own hardware
  - Reduced Total Cost of Ownership (TCO) & Operational Expense (OPEX)
- Benefit from massive economies of scale
  - Prices are reduced as AWS is more efficient due to large scale
- Stop guessing capacity
  - Scale based on actual measured usage
- Increase speed and agility
- Stop spending money running and maintaining data centers
- Go global in minutes: leverage the AWS global infrastructure

Cloud computing challenges

- Data Breach:

Data Breach is the process in which the confidential data is viewed, accessed, or stolen by the third party without any authorization, so organization's data can be hacked by the hackers.

- Vendor lock-In:

Organizations may face problems when transferring their services from one vendor to another. As different vendors provide different platforms, that can cause difficulty moving one cloud to another.

- Increased complexity strains IT staff:

Migrating, integrating, and operating the cloud services is complex for the IT staff. IT staff must require the extra capability and skills to manage, integrate, and maintain the data to the cloud.

Cloud Computing challenges

- Account Hijacking:

It is the process in which individual user's or organization's cloud account (bank account, e-mail account, and social media account) is stolen by hackers. The hackers use the stolen account to perform unauthorized activities.

- Internet dependency:

Internet is the unsung hero of the entire cloud setup. An unfailing internet service provider is what an organization needs to achieve 100% success in the cloud.

Costing Models: 5 Cloud Cost Models

Pay-As-You-Go

In this model, cloud services are billed per actual usage. Cloud services may bill for utilization of computing power, storage, networking, or other resources. The advantage is that you only pay for actual usage, and can scale down resources when needed. The downside is that as you add more resources to your cloud deployment, ongoing costs can quickly skyrocket.

Prepaid/Fixed Subscriptions

In a subscription-based model, cloud customers pay for services upfront. Subscription prices deliver a predetermined package of services for a specified time. The longer the period, the lower the price. Subscription pricing is common for cloud services that combine multiple hardware and software elements, like platform as a service (PaaS) and software as a service (SaaS).

Reserved Instances

Reserved instances allow companies to commit to cloud resources for a long period of time, typically 1 or 3 years. The longer the period, the greater the discount. A three-year term is usually the most cost effective. Cloud providers typically offer discounts of 50-75% compared to pay-as-you-go rates for reserved instances with the same capabilities.

Costing Models: 5 Cloud Cost Models

AWS Savings Plan

Similar to reserved instances, Savings Plans are a flexible pricing model that allows organizations to enjoy lower than on-demand pricing, in exchange for a one-year or three-year specific usage commitment.

AWS offers saving plans such as:

- **Compute Savings Plans** - apply to all usage of Amazon compute services usage, including EC2, AWS Lambda and Fargate.
- **EC2 Savings Plans** - applies only to usage of Amazon EC2 instances.

Spot Instances

Spot instances are usually the lowest-cost computing option, offering discounts of up to 90% compared to pay-as-you-go rates. Spot instances are used by cloud providers to sell off spare capacity. The discount comes with a catch—spot instances can be interrupted at very short notice.

Cloud Computing Security

Security in cloud computing is a major concern. Data in cloud should be stored in encrypted form. To restrict client from accessing the shared data directly, proxy and brokerage services should be employed.

Security Planning

Before deploying a particular resource to cloud, one should need to analyze several aspects of resource such as:

- Select resource that needs to move to the cloud and analyze its sensitivity to risk.
- Consider cloud service models such as **IaaS, PaaS,** and **SaaS.** These models require customer to be responsible for security at different levels of service.
- Consider the cloud type to be used such as **public, private,** or **hybrid.**
- Understand the cloud service provider's system about data storage and its transfer into and out of the cloud.

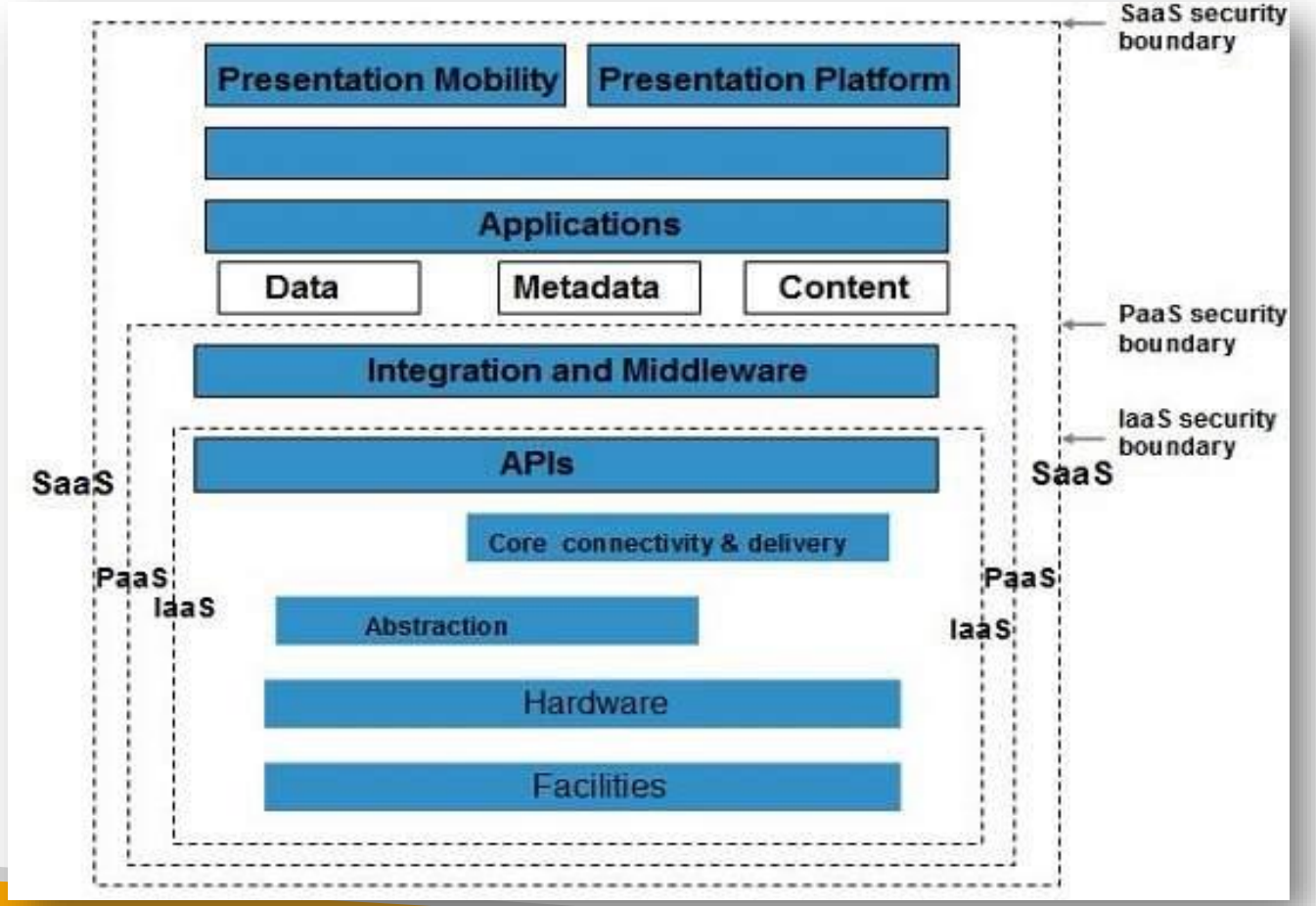
The risk in cloud deployment mainly depends upon the service models and cloud types.

Understanding Security of Cloud

Security Boundaries

A particular service model defines the boundary between the responsibilities of service provider and customer.

Cloud Security Alliance (CSA) stack model defines the boundaries between each service model and shows how different functional units relate to each other.



Top 10 Cloud Service Providers

The top 10 cloud service providers globally in 2022 are ranked in the following table, which includes the number of regions and availability zones that each vendor possesses:

#	Cloud Service Provider	Regions	Availability Zones
1	Amazon Web Services (AWS)	26	84
2	Microsoft Azure	60	116
3	Google Cloud Platform (GCP)	34	103
4	Alibaba Cloud	27	84
5	Oracle Cloud	38	46
6	IBM Cloud (Kyndryl)	11	29
7	Tencent Cloud	21	65
8	OVHcloud	13	33
9	DigitalOcean	8	14
10	Linode (Akamai)	11	11

Your AWS Certification journey

FOUNDATIONAL

Six months of fundamental AWS Cloud and industry knowledge

PROFESSIONAL

Two years of experience designing, operating, and troubleshooting solutions using the AWS Cloud

ASSOCIATE

One year of experience solving problems and implementing solutions using the AWS Cloud

SPECIALTY

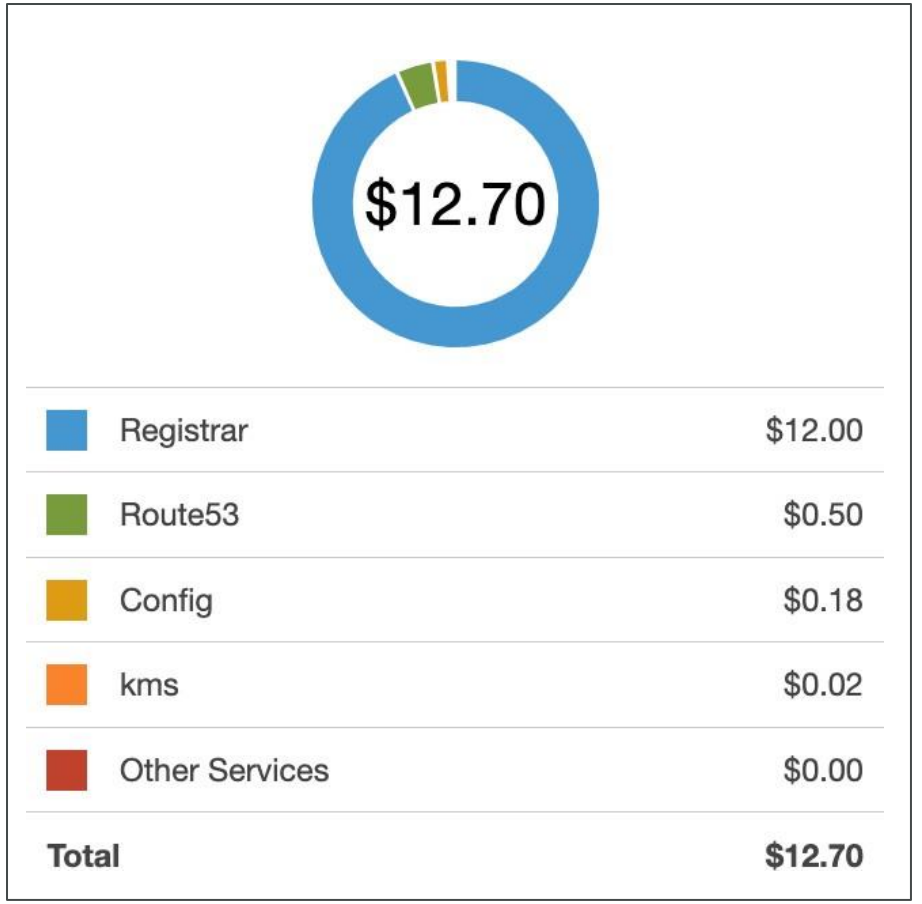
Technical AWS Cloud experience in the Specialty domain as specified in the exam guide

Estimated Cost for this Course

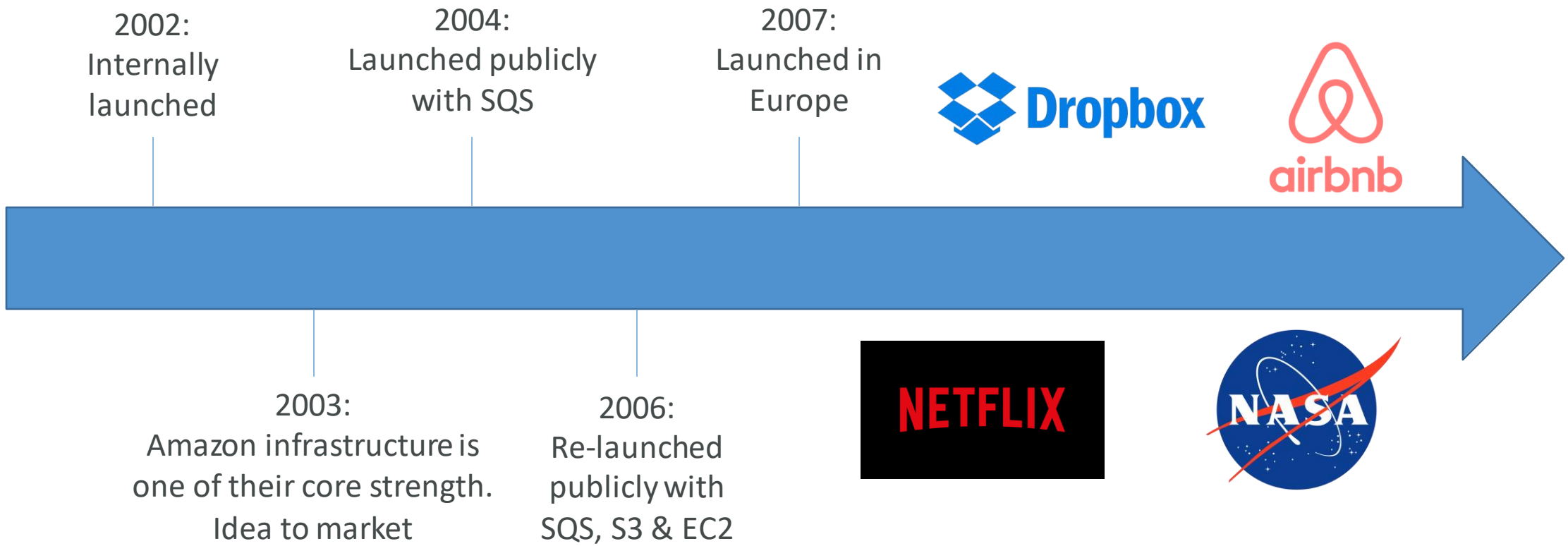
- Most of the services we'll use will be within the AWS Free Tier = $0

- If I use a service which will cost you money, I will mention it

- You can read more about the AWS Free Tier at: https://aws.amazon.com/free/



AWS Cloud History



AWS Cloud Number Facts

- In 2019, AWS had $35.02 billion in annual revenue
- AWS accounts for 47% of the market in 2019 (Microsoft is 2nd with 22%)
- Pioneer and Leader of the AWS Cloud Market for the 9th consecutive year
- Over 1,000,000 active users

Gartner Magic Quadrant

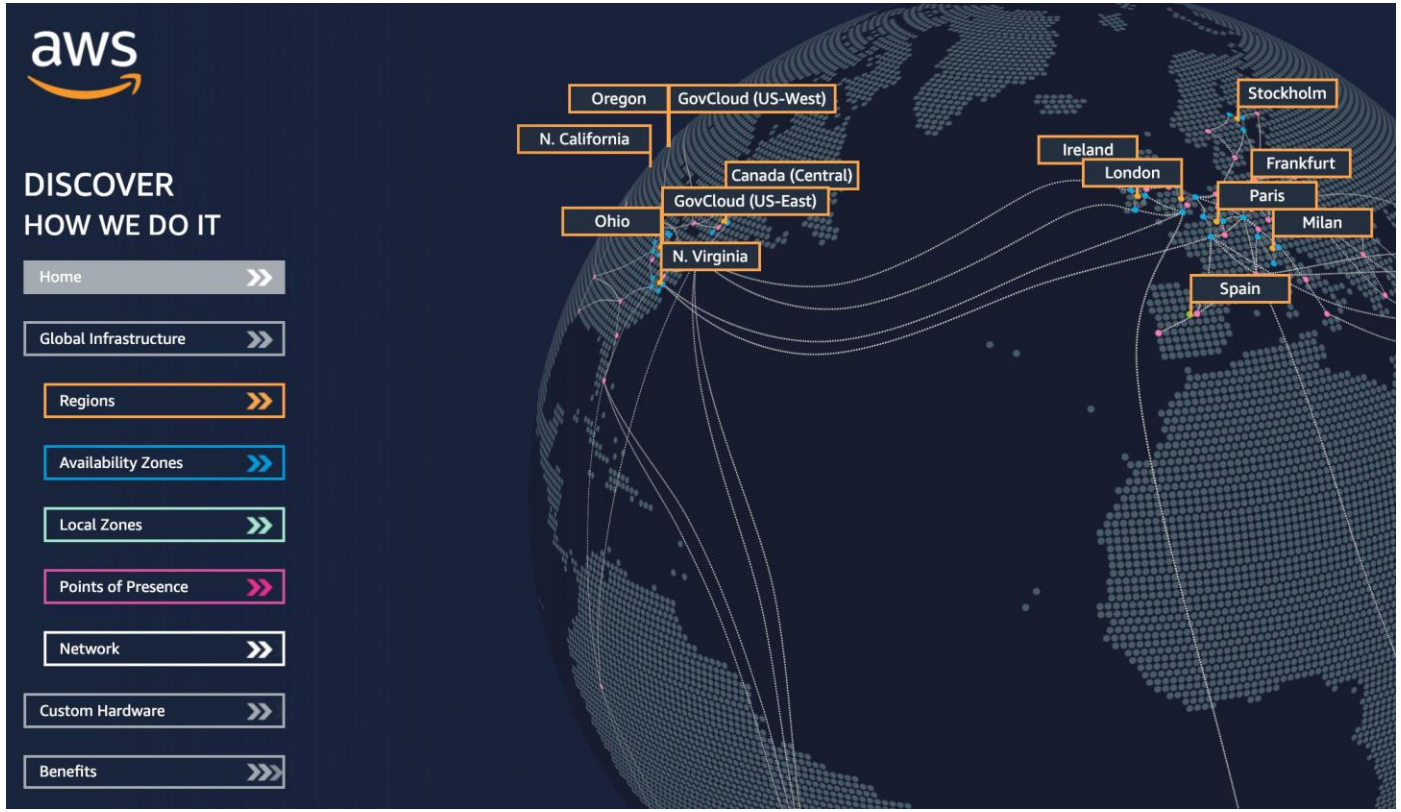
AWS Cloud Use Cases

- AWS enables you to build sophisticated, scalable applications
- Applicable to a diverse set of industries
- Use cases include
  - Enterprise IT, Backup & Storage, Big Data analytics
  - Website hosting, Mobile & Social Apps
  - Gaming

AWS Global Infrastructure

- AWS Regions
- AWS Availability Zones
- AWS Data Centers
- AWS Edge Locations / Points of Presence

- https://infrastructure.aws/



AWS Regions

- AWS has Regions all around the world
- Names can be us-east-1, eu-west-3 ...
- A region is a cluster of data centers
- Most AWS services are region-scoped

https://aws.amazon.com/about-aws/global-infrastructure/

US East (N. Virginia)	us-east-1
US East (Ohio)	us-east-2
US West (N. California)	us-west-1
US West (Oregon)	us-west-2
Africa (Cape Town)	af-south-1
Asia Pacific (Hong Kong)	ap-east-1
Asia Pacific (Mumbai)	ap-south-1
Asia Pacific (Seoul)	ap-northeast-2
Asia Pacific (Singapore)	ap-southeast-1
Asia Pacific (Sydney)	ap-southeast-2
Asia Pacific (Tokyo)	ap-northeast-1
Canada (Central)	ca-central-1
Europe (Frankfurt)	eu-central-1
Europe (Ireland)	eu-west-1
Europe (London)	eu-west-2
Europe (Paris)	eu-west-3
Europe (Stockholm)	eu-north-1
Middle East (Bahrain)	me-south-1
South America (São Paulo)	sa-east-1

How to choose an AWS Region?

If you need to launch a new application, where should you do it?

- Compliance with data governance and legal requirements: data never leaves a region without your explicit permission
- Proximity to customers: reduced latency
- Available services within a Region: new services and new features aren't available in every Region
- Pricing: pricing varies region to region and is transparent in the service pricing page

AWS Availability Zones

- Each region has many availability zones (usually 3, min is 2, max is 6). Example:
  - ap-southeast-2a
  - ap-southeast-2b
  - ap-southeast-2c
- Each availability zone (AZ) is one or more discrete data centers with redundant power, networking, and connectivity
- They're separate from each other, so that they're isolated from disasters
- They're connected with high bandwidth, ultra-low latency networking