What is Datacenter ?

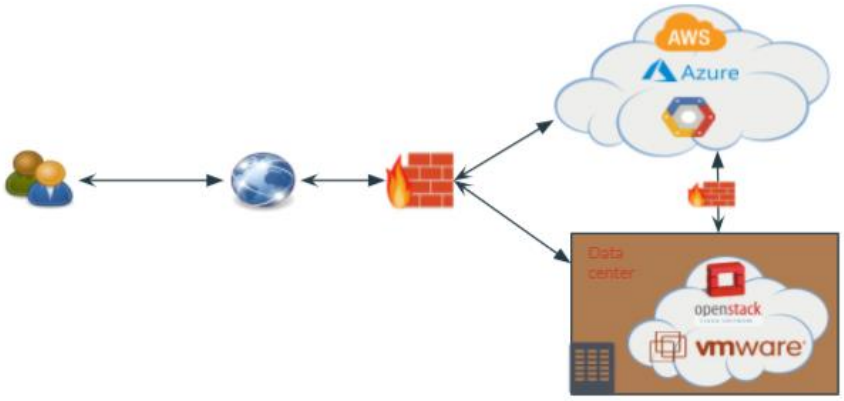
- Before understanding what is cloud computing let's go back to the traditional IT Infrastructure or On-premises Data center and understand what and how it was before
- An IT Infrastructure or On-premises Data center is a physical usable facility/place where organizations has deployed their core IT components/equipment of the data center which includes the core shared resources which are accessible round-the-clock.
- Core IT Resources
    - Network (routers, switches)
    - Compute (servers & databases)
    - Storage (SAN, backup/tape storage)
    - Software and Applications
- Core Non-IT Resources
    - Power & cooling devices
    - Server racks
    - Cables
    - Domain specific engineers

What is Cloud Computing ?

- To understand it in an easiest way let's break it into two parts
    - **Cloud** = Internet
    - **Computing** = IT infrastructure (allows to build, deploy, run and access your application/data)
- So, it means cloud computing is a virtual data center which is accessible remotely from anywhere over internet.
- Another simple way to understand is hire a data center and other IT services rather build your own one.
- It allows you to build, deploy, run and access your infrastructure or any available IT services on demand basis and you pay only for what you use.
- It also allows you to build your infrastructure in minutes when you need and destroy them in minutes when you do not need them.

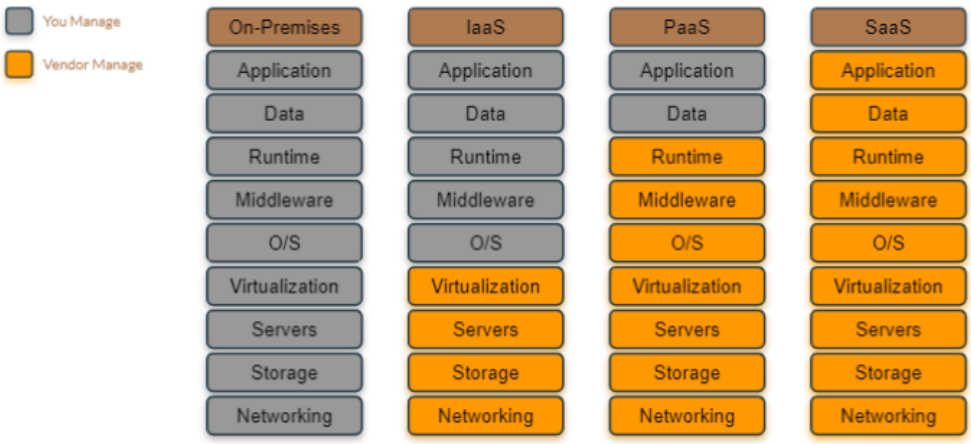
Deployment Model :

- Basically cloud computing comes with 3-types of service models
    - Public Cloud : Cloud infrastructures are managed by cloud providers and available publicly which is accessible over Internet, such as AWS, Azure and Google cloud
    - Private Cloud : Cloud infrastructures are managed by organizations or a third party vendor and only accessible by the organization or their partners
    - Hybrid Cloud : When organizations started using combining both public and private cloud then it becomes a hybrid cloud environment



Cloud Service Offering Model :

- Basically cloud computing has 3-types of deployment models
  - IaaS : Infrastructure as a Service, provides networking, storage, servers and OSs as a service which allows you to design, control and manage your virtual infrastructure on cloud the way you want.
  - PaaS : Platform as a Service, provides OSs, middleware, databases and web servers as a service where you just need to focus on your application and data rather underlying infrastructure
  - SaaS : Software as a Service, allows you to access the applications online rather installing and managing the on your own system such as Google apps
- Let's understand the models in details



Why Cloud Computing ?

- To move away from a heavy maintenance and expensive IT resources which an organization use in their daily operational work and acquire well maintained inexpensive IT resources on demand.
- Cloud computing offers many other benefits like
  - Easy Accessibility : Access from anywhere with any devices over internet
  - Cost Effective : You pay only for what you use, what they called "pay-as-you-go"
  - Reduce Time & Effort : Cloud helps in provisioning resources in minutes and automated way to minimize your infrastructure plan and effort, which allows you to focus on your other development work.

- Scalability : It allows you to scale up and down with your workload depending on the demand
- Agility & Flexibility : Based on the business requirement you can select the best deployment model and enhance your workload by integrating with other available services
- Security : Cloud providers gives special attention on the security area to meet the most security standards and procedures to keep the data secure at rest and in transit
- Reliability : They simplified a lot to plan your disaster recovery which allows you to backup and restore your data in a faster and reliable way based on the RPO and RTO
- Global Infrastructure : You can expand and deploy your infrastructure globally in minutes because of its extensive global infrastructure

Top Cloud Service Providers :

- Amazon Web Services (AWS)
- Microsoft Azure
- Google Cloud Platform (GCP)
- Alibaba Cloud
- IBM Cloud
- Oracle Cloud

Amazon Web Services (AWS) :

- Provides a highly reliable, scalable, low-cost infrastructure platform in the cloud around the world
- They first started introducing with SQS service in the 2004
- AWS was the first cloud service provider and they started providing IaaS in 2006 and they started with S3, EC2 and SQS
- In 2009 they started introducing AWS Management Console for better and easier management of the AWS resources
- In 2009 they started introducing Virtual Private Cloud (VPC), which allows customers to connect their own existing network to an isolated private network on AWS.
- AWS is the leading cloud provider in today's date on the marketplace
- It offers more than 200 services
- It has 25 geographic regions around the world with 80 availability zones
- Announced plans for 15 more Availability Zones and 5 more AWS Regions in Australia, India, Indonesia, Spain, and Switzerland
- 97 Direct Connect locations
- 230+ Points of Presence (218+ Edge Locations and 12 Regional Edge Caches)

- For latest information browse the link : https://aws.amazon.com/about-aws/global-infrastructure/regions_az/

Important AWS Terms and Concepts of AWS Global Infrastructure:

- AWS Console (GUI)
- What are geographic regions
- What are availability zones
- What is Points of Presence (PoP)

- What are Edge Locations
- What is Regional Edge Caches