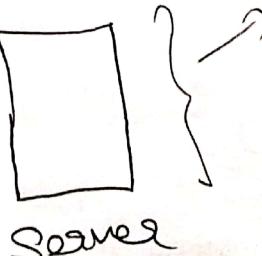
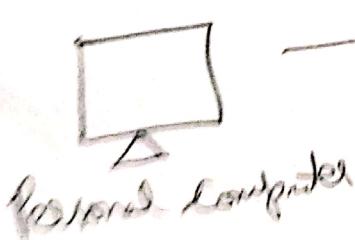


Chapter 1: Understanding Cloud Concepts, Vocabulary and Terminology

Cloud

Today we use different applications like YouTube, Instagram, WhatsApp etc. These applications are developed in the developer's laptop but these applications has to be accessible to everyone. So, these applications have to be deployed on the servers.

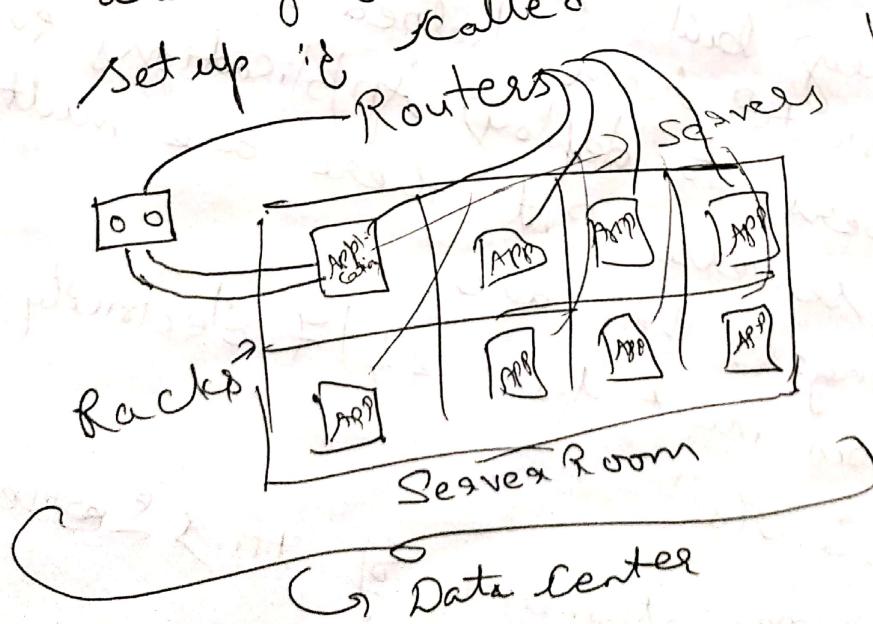
Server → Server is also a computer but server has some local storage + CPU + RAM. And the purpose of these servers is to run applications. Your laptop is also a server which has a very good display due to Graphic card. It is advanced / rich server. But servers don't have this rich features but they have the important ones that is a local storage, CPU & RAM that are required to run the applications as well as processes.



Computer
↓
local storage
+
CPU
+
RAM

→ 15 Years back, YouTube, Google they use to be a dedicated System Admin, whose responsibility is to procure the servers. Like we have vendors like IBM, HP who sell these servers. System Administrators are to procure multiple servers. Deploying multiple applications onto one server doesn't work. So, they use to purchase multiple servers.

→ System Administrator teams use to procure servers & connect these servers to high data transfer cables, high data transfer switches & from there all of these servers use to get connected to the routers. System Administrator teams build this entire system where these servers use to be placed in some racks (Server Room). Applications are deployed on these servers. This entire set up is called as Data center.



Typically, most companies use to have their own Data centers.

Ex:
Google Data Center

Now Google's System Administrators team has built this servers.

Now let's say a developer called "X". This developer needs a server. When this developer requests for a server, the System Administrator team looks for all these servers and they use to say this particular server is empty/free. You can use part of server or entire server. This is this server or private cloud.

called as Private Cloud.

Now Amazon came into picture, they said they were starting something called as

AWS. Google buy lot of servers

for Google ex, they buy lot of servers for System Administrator team.

& they set up Data center. But System Administrator team helped Application developers to deploy Applications in multiple servers. But there are multiple challenges. They are:

1. Server Room must have 24/7 Electricity supply

2. Overhead cost

3. Continuous maintenance of this servers.

→ Now Amazon came & they said they built AWS. And we will set up our collection of servers → Datacenter Data Centers across the world.

→ Google can come to Amazon & request for (ex: 10) Servers. And we will give allocate 10 servers. And you don't have to worry about those challenges. Just tell us which region you want. And this particular thing is called as Public Cloud.

→ There are some banking, financial companies who needs their own setup due to security purpose. So, they were following private cloud!

→ We have public clouds like AWS, Azure, GCP.

Hybrid Cloud → Ex: Google said that they loved Azure. So, Microsoft has done similar job like AWS & Microsoft set up data centers across the world & Google requested for 100 servers & Microsoft gave 100 servers. Google deployed their applications on these servers. But after a while, Google has some sensitive data & they want to store this sensitive data.

data in database. But these databases have to be in our private cloud only. We cannot give this sensitive information to Microsoft. [We will keep it with us only. Now Google setup a simple data center. This is a private cloud. Because this is in their premises i.e., on-premises. It's a mix of private cloud ~~on-premises~~ & public cloud. So, it is called as hybrid-cloud.

→ Hybrid Cloud & Multicloud is not same. Multicloud is like using Ex:

AWS + Azure - - - - -

→ On-premises & private cloud is not same.

On-Premises: Refers to IT infrastructure (like servers & storage) that is physically located within a company's own data center. The organization owns, manages & maintains the hardware and software.

Private cloud: Also involves dedicated infrastructure, but it offers cloud like features such as scalability, virtualization & on-demand resources. It can be hosted either on-premises or by a third-party provider, but it's exclusively used by one organization only.

Cloud Computing :-

→ Whatever we discussed till now (~~private~~)
public & hybrid cloud) is basically
basically a cloud computing.

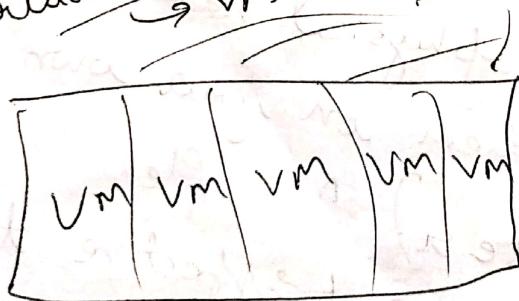
→ If you're computing on cloud i.e,
if you're running your applications on
cloud & if the applications are accessed
by the users, so system Administrators,
DevOps Engineers, cloud engineers do
whatever the tasks that these engineers
perform & whatever the tasks that AWS,
Azure -- perform which the cloud
platform that you're using. Together
this entire computing is called as
Cloud computing.

→ Cloud Computing is the delivery of
computing services like storage, servers &
software over the internet, allowing you to
access & use them without ~~owning~~
owning the physical hardware. (Instead
of having them on your own computer
or datacenter) It's flexible, Scalable
& often more cost effective than
managing physical hardware.

→ Ability to easily increase or
decrease the size of a system

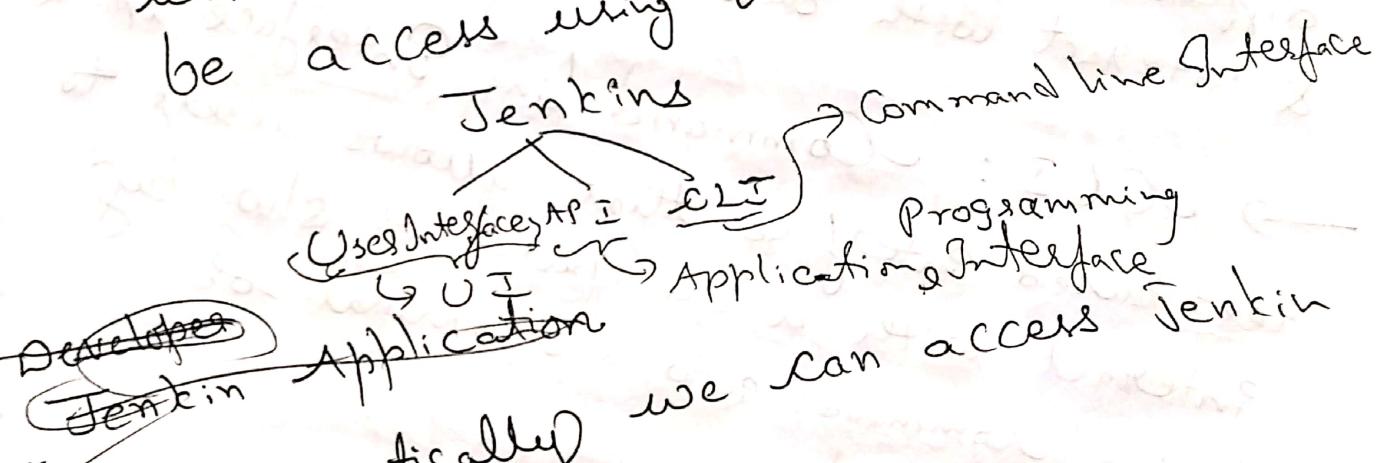
Virtualization →

whether it is private cloud or public cloud,
lets say system administrators or devops
Engineers they bought a server from IBM.
When IBM sells this server, this server
comes with huge configuration like 512 GB
RAM, 1000 CPUs... let's say a developer
(Dev) requests for a server. But sys Admin,
Devops engineer won't dedicate this entire
server to Dev. Then come the concept called Virtualization, where
~~Virtualization~~ System Administrators use
hypervisor to install a particular SW called Hypervisor
Using this hypervisor, it will break this
server logically (won't break physically)
into multiple servers. (Ex: 512 GB RAM,
using hypervisor, you can break it into 512
logical servers, these servers are called
as Virtual Machines)



API :-

Ex: let's take an example of Facebook. You can go to your browser 'www.facebook.com'. (User Interface) You can access this application (facebook). Typically you're accessing this application from the User Interface. Now when we talk about DevOps Applications like Kubernetes, Jenkins. These Applications can also access using the User Interface. Along with this, these Applications can also be accessed using API (or) CLI.



Programmatically

→ Application. Being User of Facebook, we can only use the Facebook to view. We don't have to access Facebook through some scripting to access Facebook through some programming. But application developers, DevOps Engineers, QA who want to develop test Facebook, who want to perform some scripting on the Facebook

For the Application developers of Facebook, they allow in 2 different ways i.e. API & CLI

API: Application Interface

In your organization, applications that are developed by developer expose the application using the API so that someone can access these applications programmatically.

API is a set of rules & protocols that allows different software applications to communicate & interact with each other.

→ CLI: Command Line Interface is a text based interface that allows user to interact with a computer or SW by typing commands into a terminal or command prompt.

Regions & Availability Zones (AZ):

We can consider an Availability Zone as a 'Data Center'.

Azure has their Data Centers across multiple parts of the world like US, UK, India.

Let's say Azure has their Data Centers across multiple parts of the world like US, UK, India.

People who are requesting the resources from Azure, there has to be a geographical identification → Region.

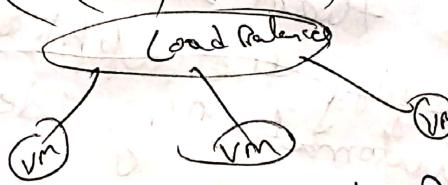
Load Balancer: → flow of data & information across a network that

A load balancer is a device or software that distributes incoming network traffic across multiple servers to ensure no single server is overloaded or burdened.

→ This helps improve the performance, reliability & availability of applications by balancing the workload & preventing server overload.

→ If an application gets multiple requests at a time then it will die so the

load balancer will send few requests to one VM & few to another like that



Features Of Cloud Computing →

(1) Scalability

(2) Elasticity

(3) Agility

(4) High Availability

(5) Fault Tolerance

(6) Disaster Recovery

Scalability →

Let's say we're watching cricket in Hotstar. Sometimes 5 cr members are using the match, drastically it may increase to 25 cr. When this number of users increased, then there will be a slowness in the Application.

So, there is a concept: there

→ Auto Scaling means i.e., Auto Scaling. DevOps engineer can tell the Azure platform you can tell the you can tell the (you can tell the application where you have deployed) they have

(where Hotstar is deployed) where you the Auto Scaling services where you can say to the Auto scale service

can say to the Auto scale service that if one VM is occupied, you can automatically scale one more VM & deploy the application. If you need

more requirement then scale one more VM & ~~deploy the application~~ identifies

more VM & load balancer automatically identifies & load balancer add the servers &

& automatically add the servers & automatically manage the load if called as AutoScaling.

This mechanism is called as Manual Scaling.

→ If you do it manually, then it is called as Manual Scaling.

→ This particular feature (Auto Scaling & Manual Scaling) of Azure is called as ~~as~~ Scalability.

Elasticity →

The other name of elasticity is:

Dynamic Scaling

(or) Auto Scaling.
Automatically scaling your Infrastructure

High Availability →

High Availability (HA) means making your application available most of the times.

Ex: Instagram. Rarely we can see application going down. And Banking

Instagram going down frequently.

Servers used to get going down frequently.

Because Instagram has

high availability setup.

Infrastructure talks about high availability.

When someone talks about infrastructure

then they are talking about applications to be

available most of the times.

highly available

Disaster Recovery (DR) →

It is a technique (or) mechanism where you need to have a plan (or) action, if

something goes wrong.

Scanned with OKEN Scanner

- DR talks about 'how do you Recover something from the Disaster'.
- DR plan is like having a Backup, dealing with cloud platforms like AWS, Azure, although they have very good SLA's, i.e., they give a lot of promises to you, but if you have data in one of the databases in Azure platform, In MS Azure, you requested for a database (similar thing like you requested for a VM) & in storing particular db, if you started storing all of your users information. And for some reason this db goes down, You have to make sure you have a backup. (Probably you can store it in another Availability Zone) Such plans are Disaster Recovery Plan.