



Cloud Computing

Computing Model

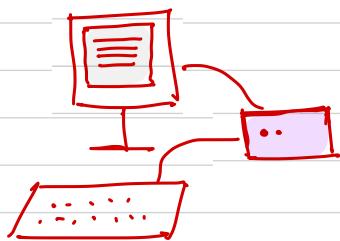
- ① Desktop computing
 - ② Client-Server computing
 - Cluster computing
 - Grid computing
 - Cloud Computing
-
- Hand-drawn red arrows point from the first four items in the list to the fifth item, indicating a relationship or flow between them.



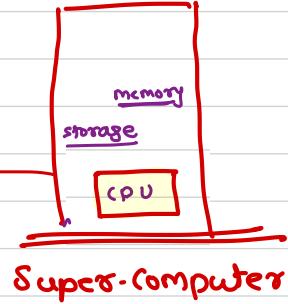
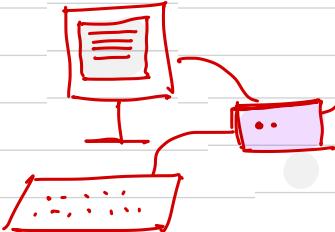
⑥

No competing

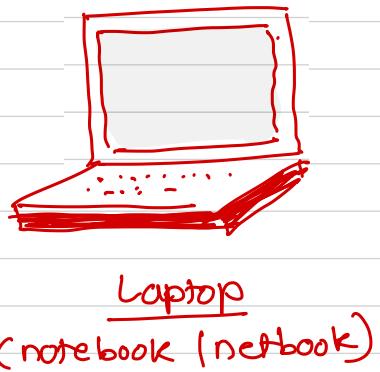
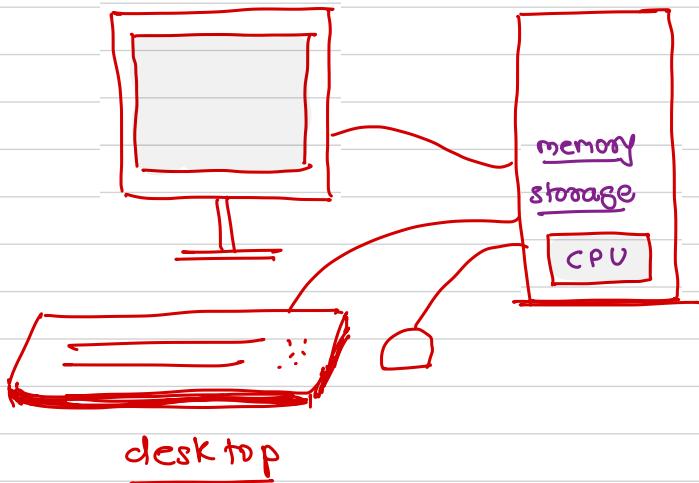
terminal ①



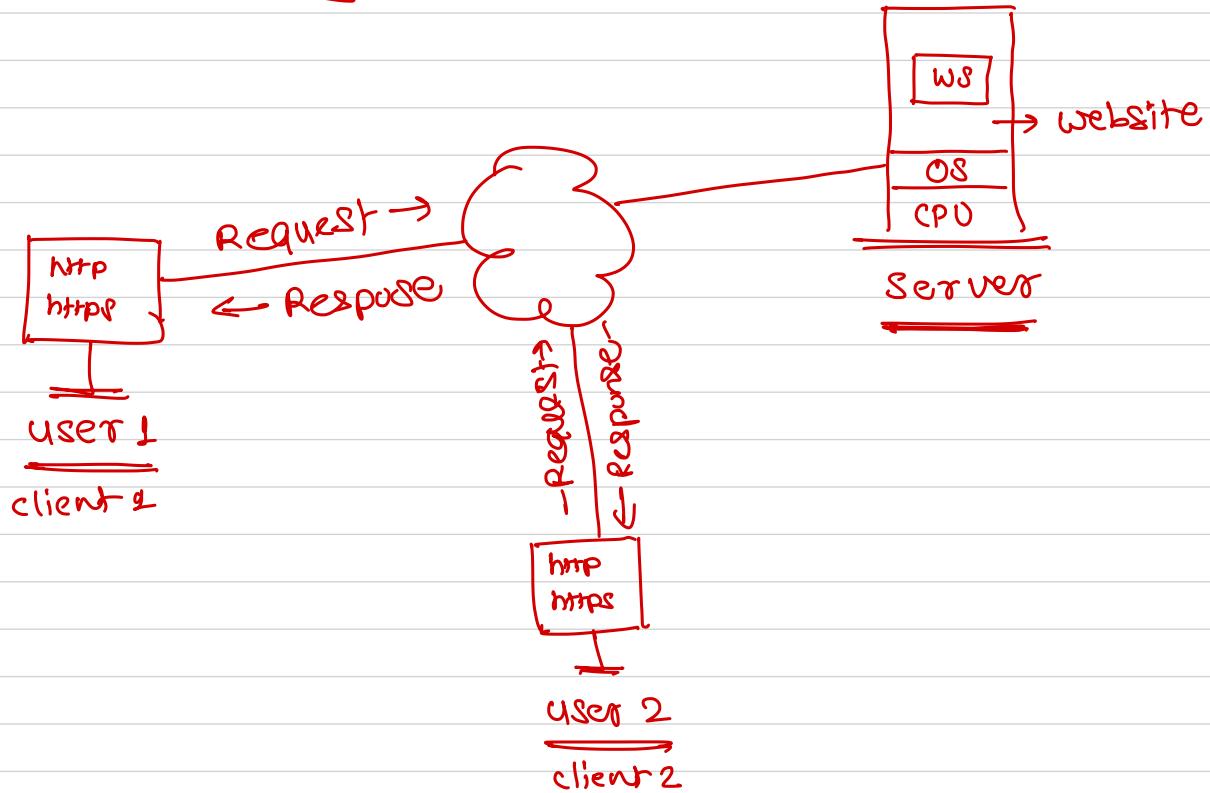
terminal ②



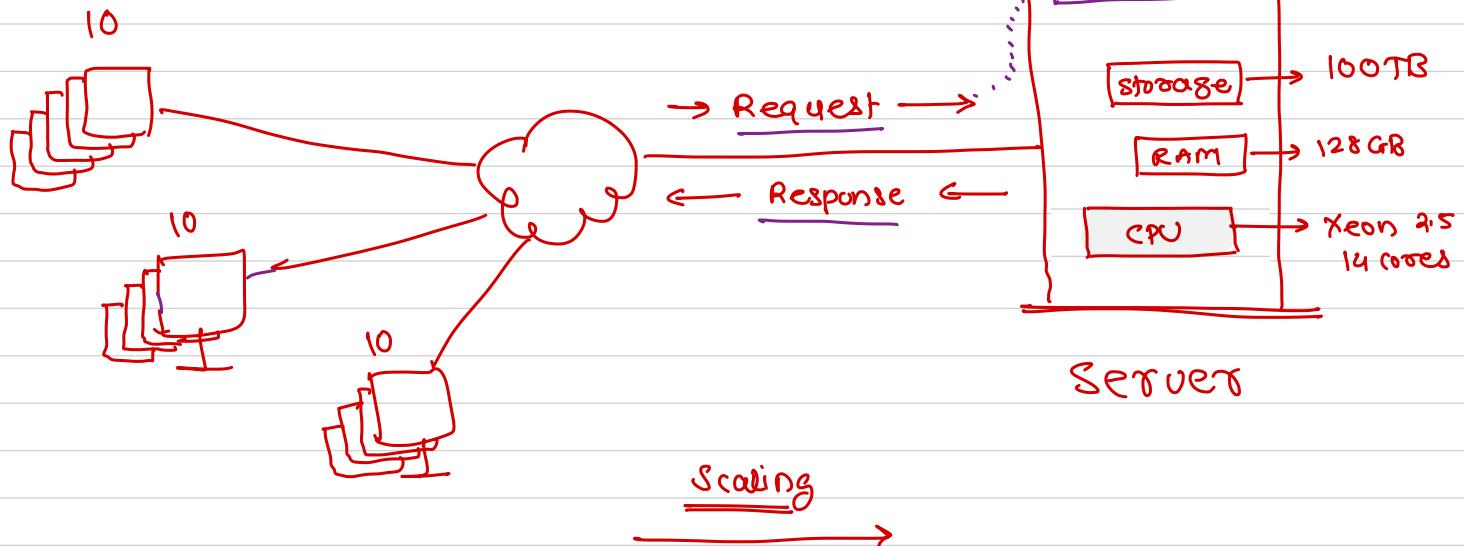
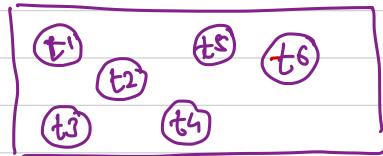
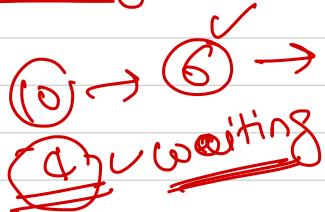
① Desktop computing



2.1.) Client-Server Computing



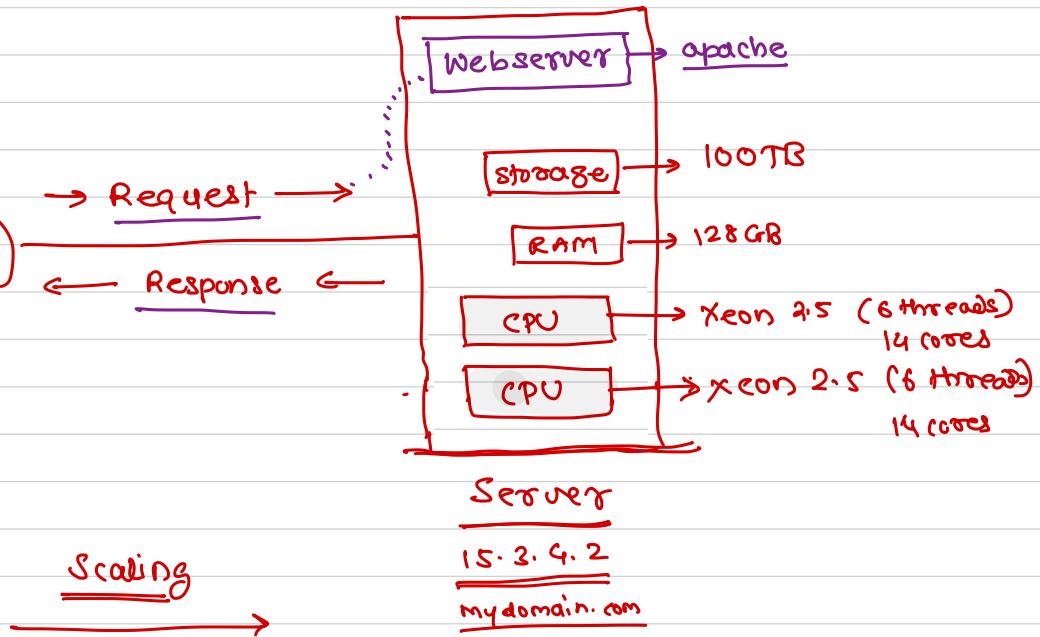
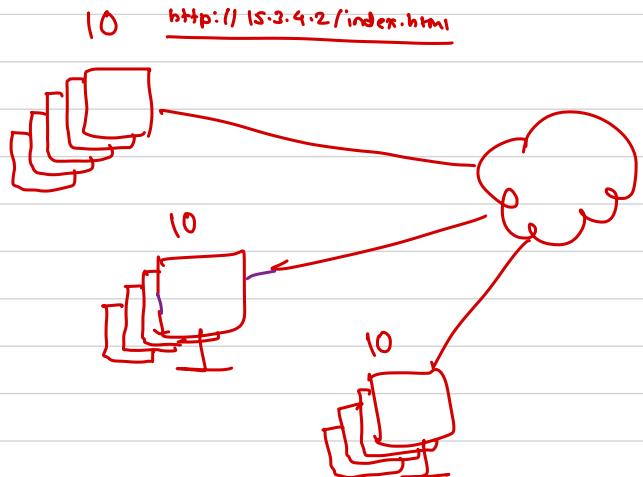
2.2) Client-Server computing



Scaling

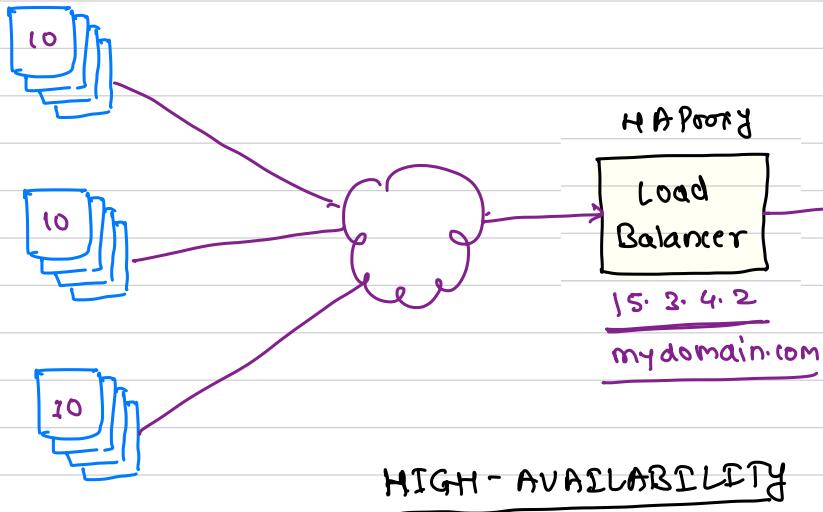
Vertical
Horizontal

- ① dependency on one machine
- ② limit on the configuration



③ cluster computing distributed computing

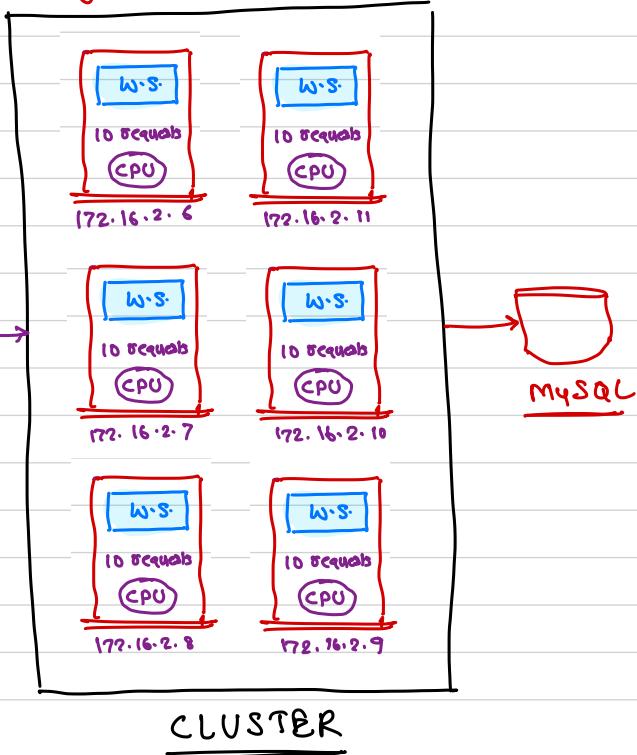
- ① no dependency on one machine
- ② increased cost



Horizontal scaling

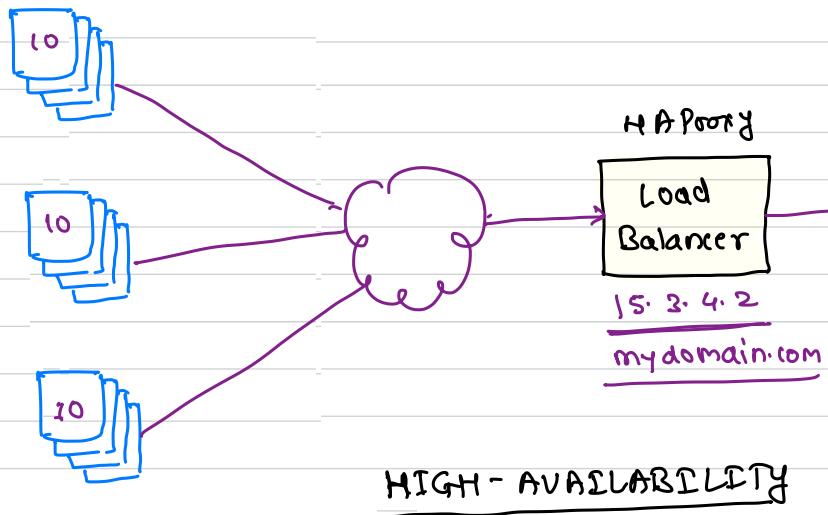
- out → adding more machines
- in → remove machines

physical machines

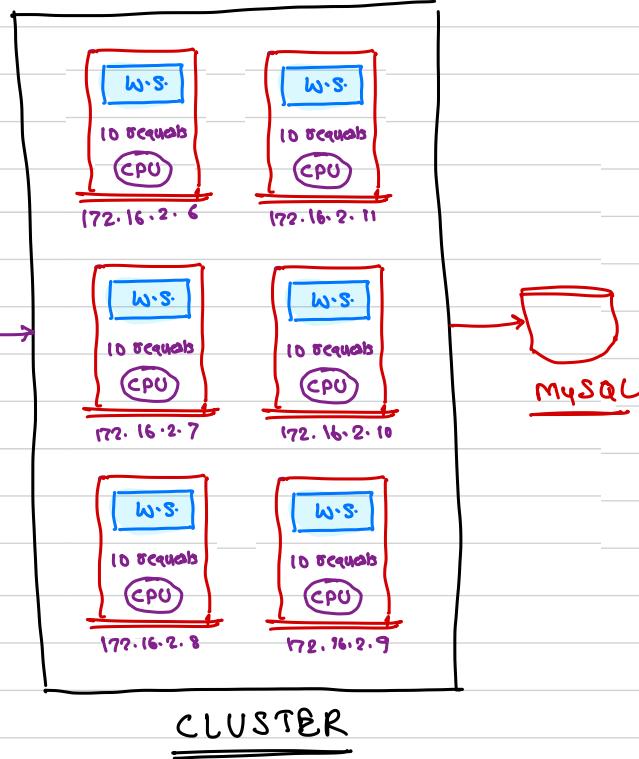


④ Cloud computing

- ① no dependency on one machine
- ② increased cost



virtual machines



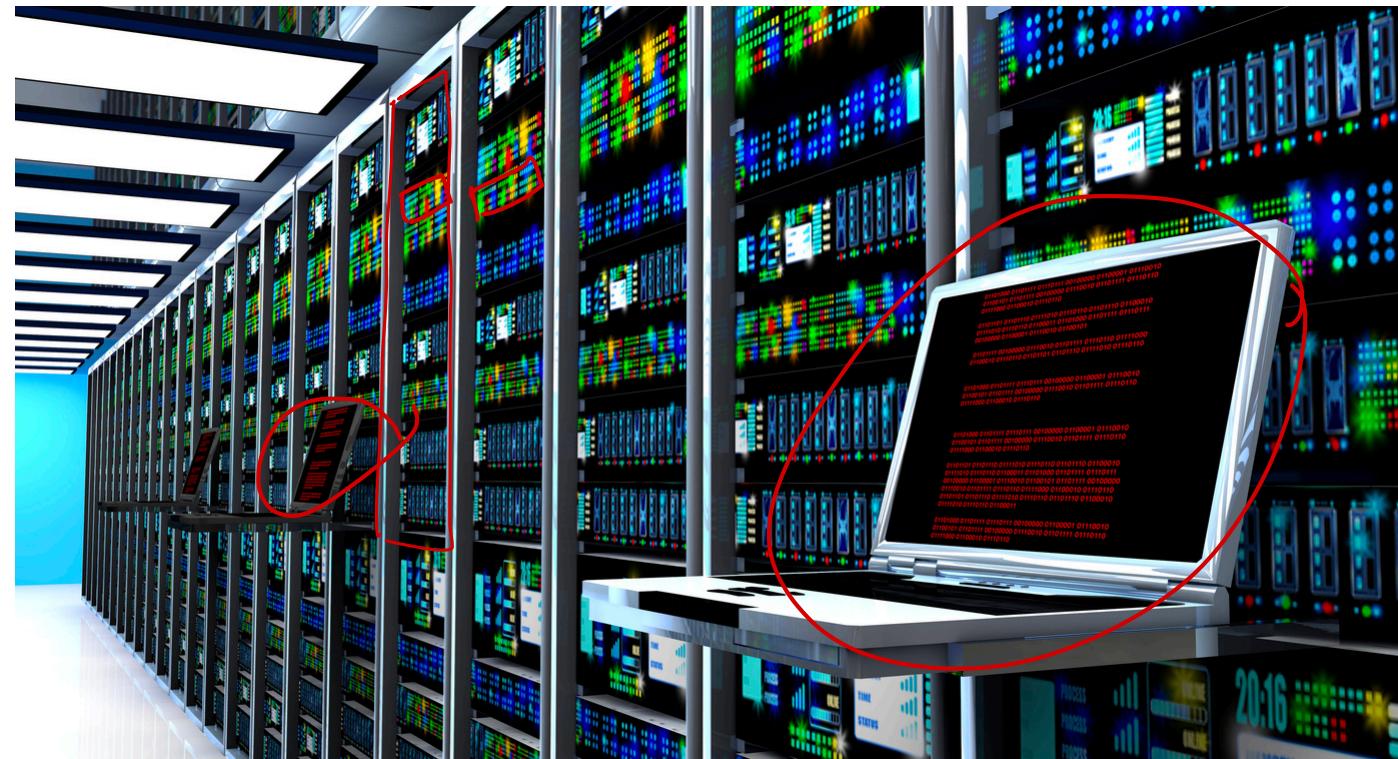
What is cloud computing ?

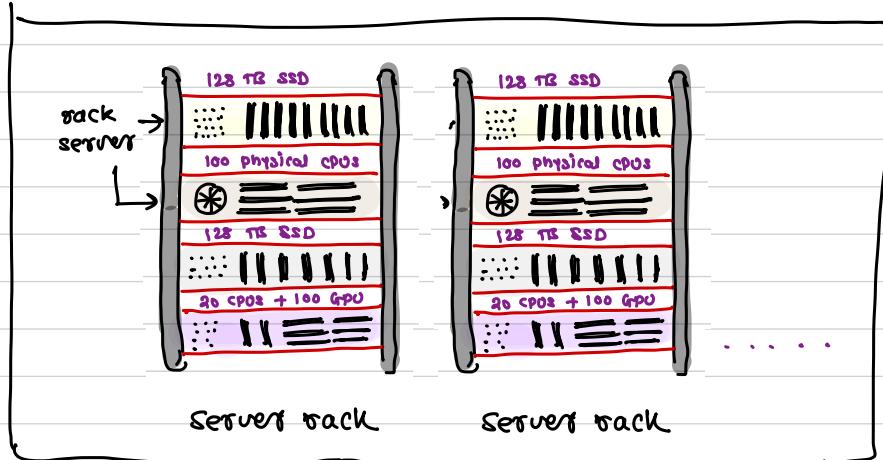
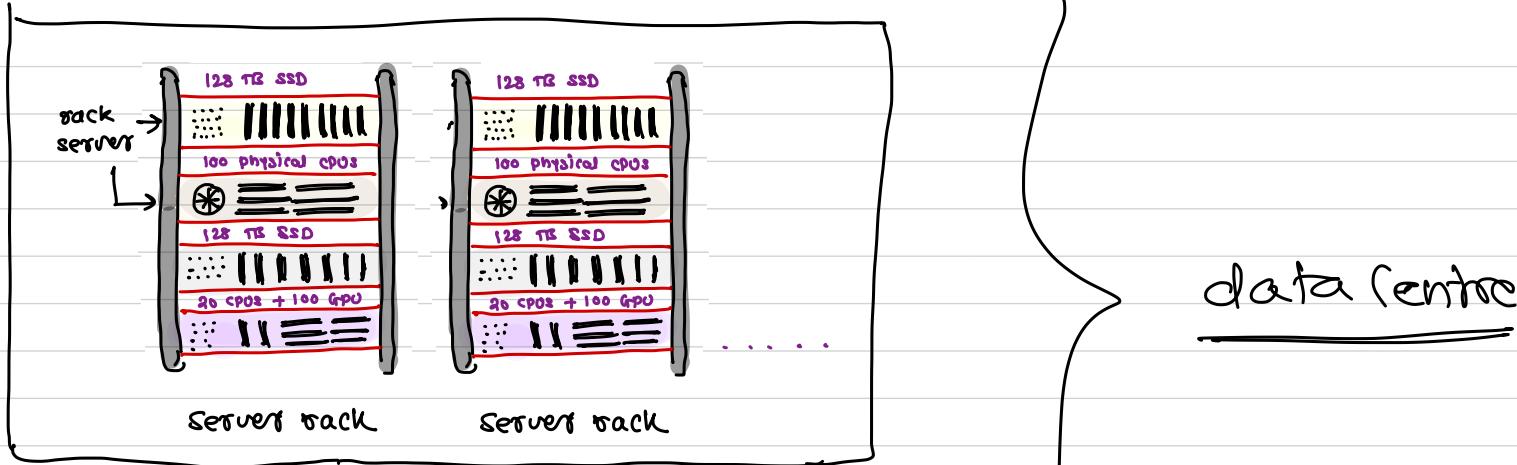
- The practice of using a network of remote servers hosted on the Internet to store, manage, and process data, rather than a local server or a personal computer.
- Is the delivery of on-demand computing resources – everything from data centers over the internet on a pay for use basis
- Cloud computing is an umbrella term used to refer to Internet based development and services



What is Data Center ?

- Where your IT devices and applications are located
- For a non-technical person it is the cloud where the user's files/data is stored
- Components
 - ✓ Servers
 - ✓ Security
 - ✓ WAN
 - ✓ Storage
 - ✓ File Sharing



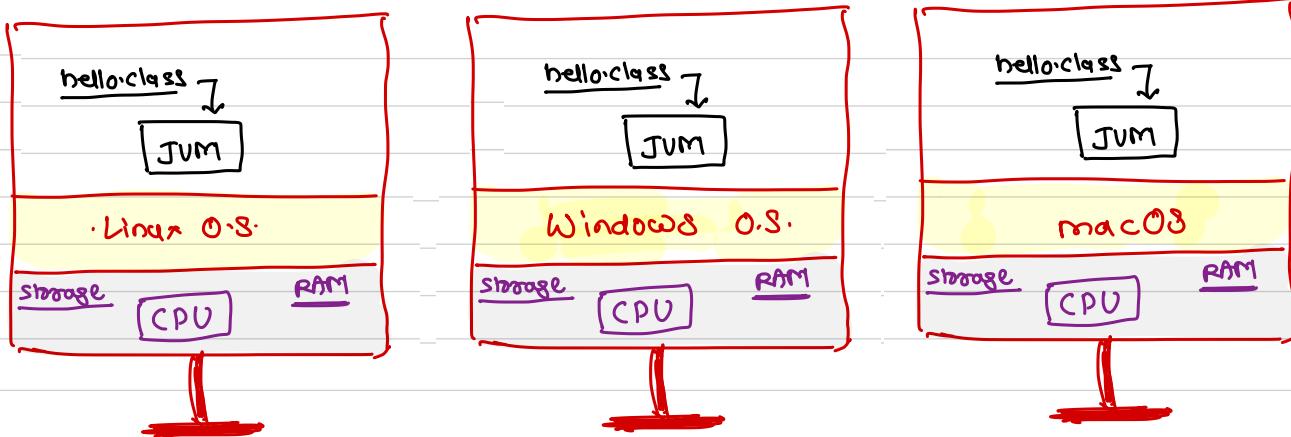


What is Virtualization ?

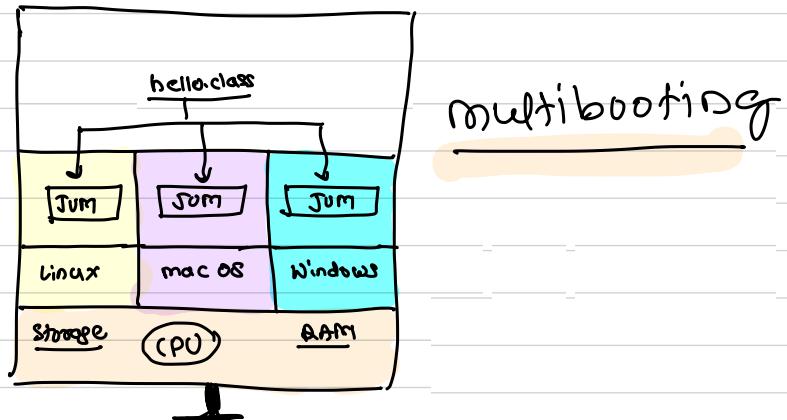
- Refers to the act of creating a virtual (rather than actual) version of something, including virtual computer hardware platforms, storage devices, and computer network resources
- Types
 - ✓ ▪ Type I
 - ✓ ▪ Type II
 - Containerization



①



②

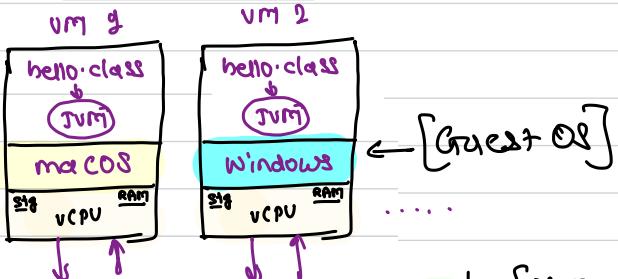


cloud providers

Virtualization

developers &
testers, end
users

Type I



[Guest OS]

Xen
VMware ESXi

Linux

Hypervisor

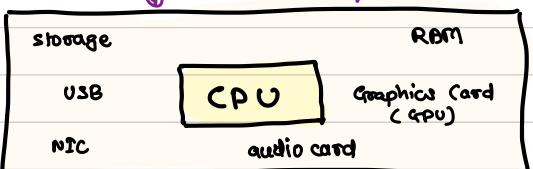
storage

vCPU

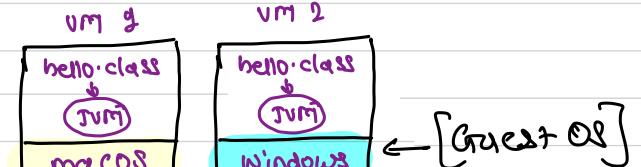
RAM
NIC

Hardware Abstraction Layer

HAL
collection of
device drivers



Type II



[Guest OS]

VMware
virtualbox
virtualPC
parallels

Hypervisor

storage

vCPU

RAM
NIC

Operating System

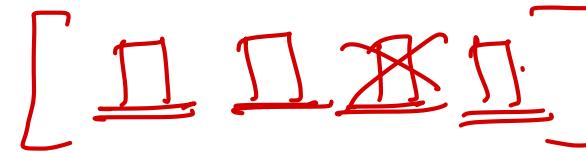
Linux [Host OS]

Hardware Abstraction Layer

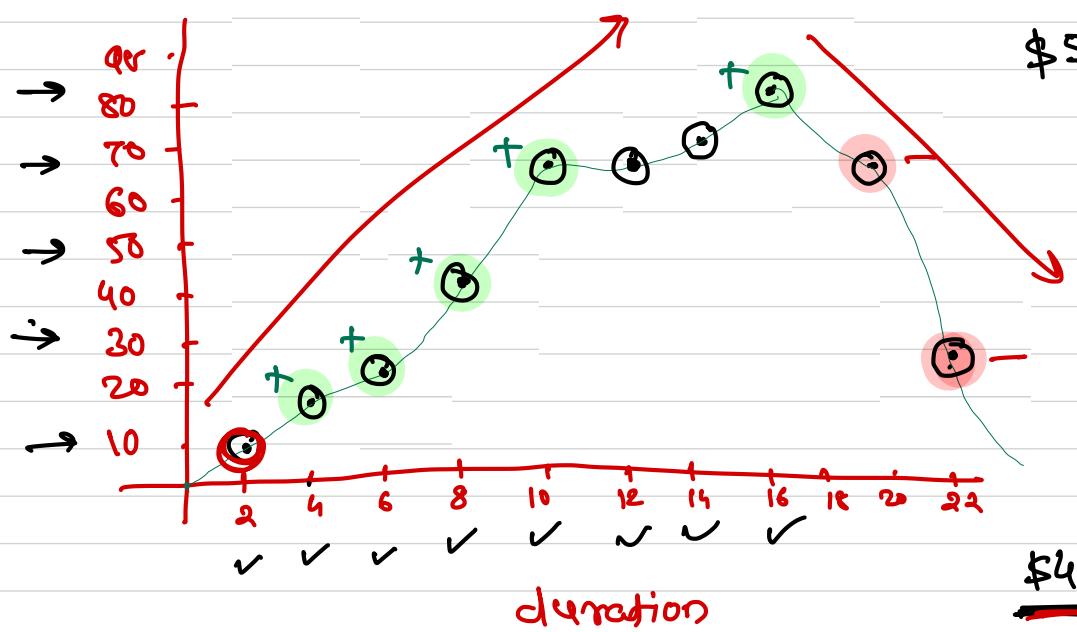
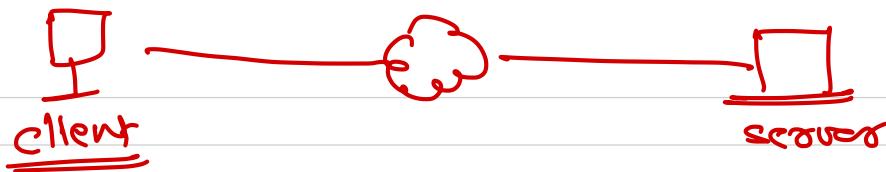
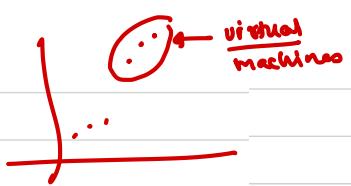
HAL
collection of
device drivers



Terminologies



- Scalability
 - refers to the idea of a system in which every application or piece of infrastructure can be expanded to handle increased load
- Elasticity
 - the degree to which a system is able to adapt to workload changes by provisioning and de-provisioning resources in an autonomic manner, such that at each point in time the available resources match the current demand as closely as possible
- Availability
 - refers to the ability of a user to access information or resources in a specified location and in the correct format
- Information Assurance
 - availability, integrity, authentication, confidentiality and nonrepudiation
- On-demand service
 - A model by which a customer can purchase cloud services as needed



\$5K Infra $\Rightarrow 10K$

$$\begin{array}{r}
 + \\
 + \\
 \hline
 2x \\
 + \\
 \hline
 3x \\
 + \\
 \hline
 10K = 30K
 \end{array}$$

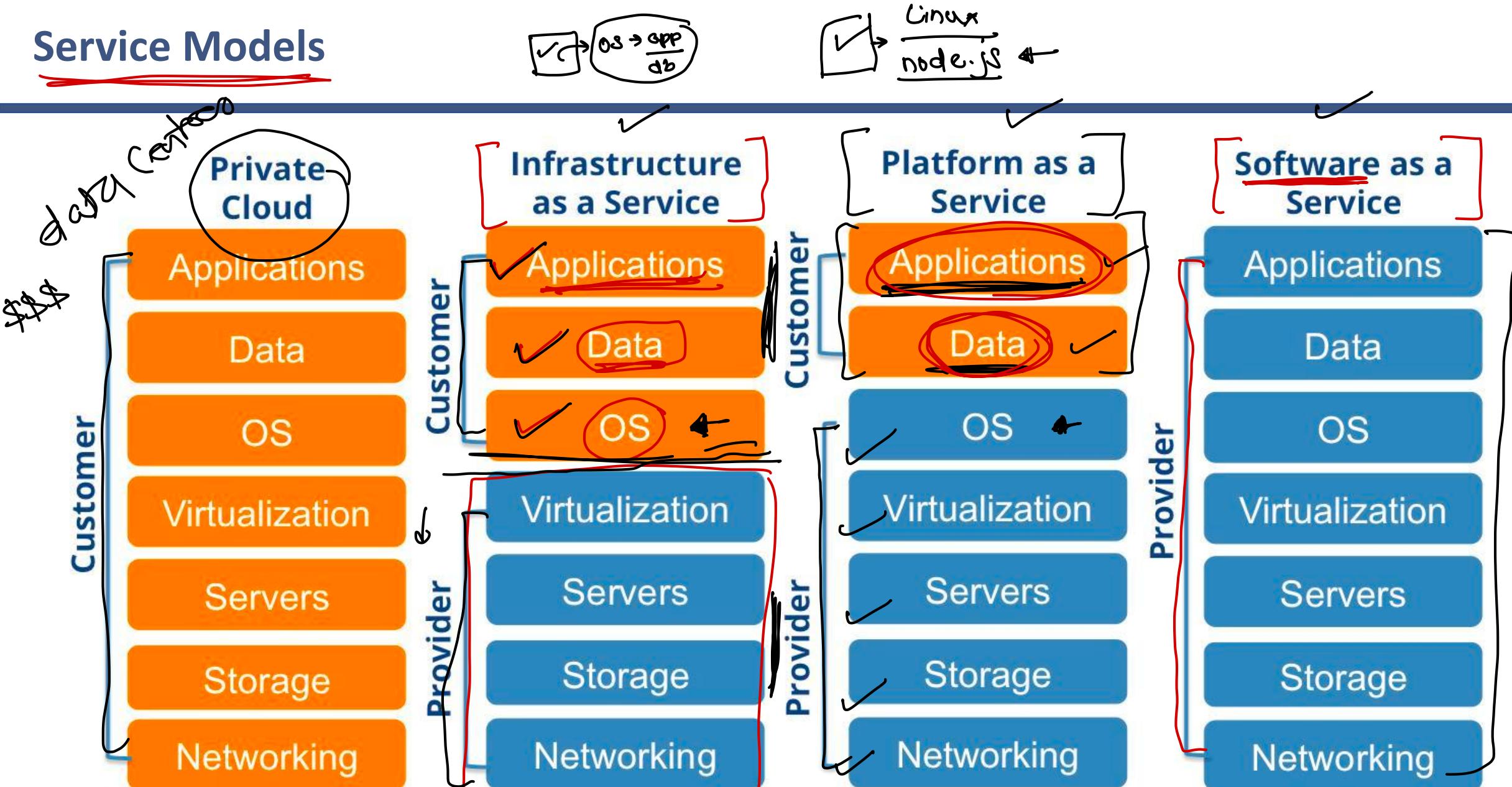
\$45K

$$\begin{array}{r}
 + \\
 + \\
 \hline
 5x \\
 + \\
 \hline
 8x \\
 + \\
 \hline
 10K = 80K
 \end{array}$$

$\frac{1}{2}$

$$\begin{array}{r}
 + \\
 + \\
 \hline
 9x \\
 + \\
 \hline
 10K = 90K
 \end{array}$$

Service Models



Service Models

Software
as a Service (SaaS)

End user

Business apps, web services, multimedia

Applications / Software

Platform
as a Service (PaaS)

developer

Frameworks (Java/Python)

Platforms

Infrastructure
as a Service (IaaS)

System admin
ops team

Computation, Storage

Infrastructure

CPU, Memory, Disk, NIC

Hardware

**Google Apps,
Facebook, YouTube,
Dropbox, Google Photos**

**Google App Engine,
Amazon Simple DB, S3,
Microsoft Azure**

**Amazon EC2,
Google Compute VM,
Azure VM**

Data Center



Service Models: IaaS

- Infrastructure as a Service
- Allocates virtualized computing resources to the user through the internet
- IaaS is completely provisioned and managed over the internet
- helps the users to avoid the cost and complexity of purchasing and managing their own physical servers
- Every resource of IaaS is offered as an individual service component and the users only have to use the particular one they need
- The cloud service provider manages the IaaS infrastructure while the users can concentrate on installing, configuring and managing their software
- Generally meant for operations team to setup the required infrastructure
- Benefits
 - Time and cost savings: more installation and maintenance of IT hardware in-house,
 - Better flexibility: On-demand hardware resources that can be tailored to your needs,
 - Remote access and resource management.



Service Models: PaaS

- Provides a platform allowing customers to develop, run, and manage applications without the complexity of building and maintaining the infrastructure typically associated with developing and launching an app
- Generally meant for developers
- Benefits
 - Mastering the installation and development of software applications
 - Time saving and flexibility for development projects: no need to manage the implementation of the platform, instant production
 - Data security: You control the distribution, protection, and backup of your business data



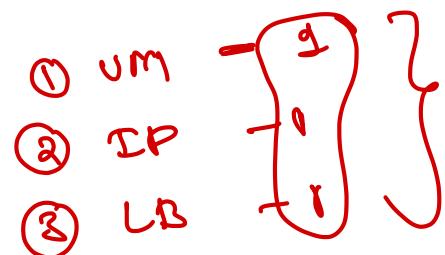
Service Models: SaaS

- Software as a Service
- Software distribution model in which a third-party provider hosts applications and makes them available to customers over the Internet
- User wont know which computer or operating system or infrastructure is used to host the software
- Generally meant for end user
- Benefits
 - You are entirely free from the infrastructure management and aligning software environment: no installation or software maintenance
 - You benefit from automatic updates with the guarantee that all users have the same software version
 - It enables easy and quicker testing of new software solutions.



Cloud Computing Characteristics

- Rapid Elasticity
- On Demand Self Service (Automatic)
- Broad Network Access
- Location Independent Resource Sharing
- Measured Services (\$\$)



Cloud Deployment Models: Public

- Supports all users who want to make use of a computing resource, such as hardware (OS, CPU, memory, storage) or software (application server, database) on a subscription basis
- Most common uses of public clouds are for application development and testing, tasks such as file-sharing, and e-mail service
- Requires internet to access the resources



Cloud Deployment Models: Private

- Typically infrastructure used by a single organization
- Such infrastructure may be managed by the organization itself to support various user groups, or it could be managed by a service provider that takes care of it either on-site or off-site
- Private clouds are more expensive than public clouds due to the capital expenditure involved in acquiring and maintaining them
- However, private clouds are better able to address the security and privacy concerns of organizations



Cloud Deployment Models: Hybrid

- Organization makes use of interconnected private and public cloud infrastructure
- Many organizations make use of this model when they need to scale up their IT infrastructure rapidly, such as when leveraging public clouds to supplement the capacity available within a private cloud
- For example, if an online retailer needs more computing resources to run its Web applications during the holiday season it may attain those resources via public clouds.



Cloud Services

- Compute: used to create the Virtual Machine (CPU) ←
- Storage: used to provide the storage ↗
- Database: RDBMS + No SQL ↗
- Security and Identity Management
- Media Services
- Machine Learning
- Cost Management
- Application Integration



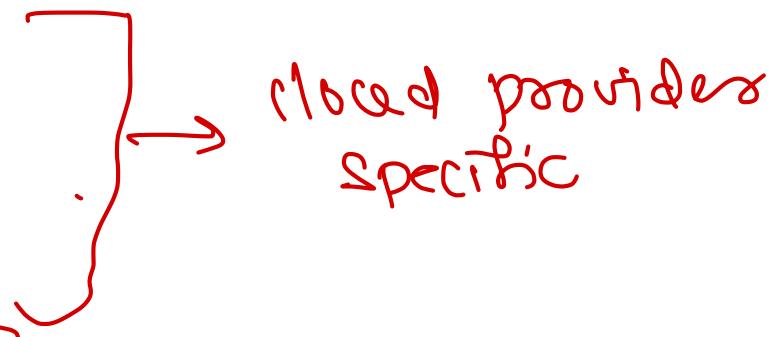
Advantages

- Lower computer costs
- Improved performance
- Reduced software costs
- Instant software updates
- Improved document format compatibility
- Unlimited storage capacity
- Increased data reliability
- Universal document access
- Latest version availability



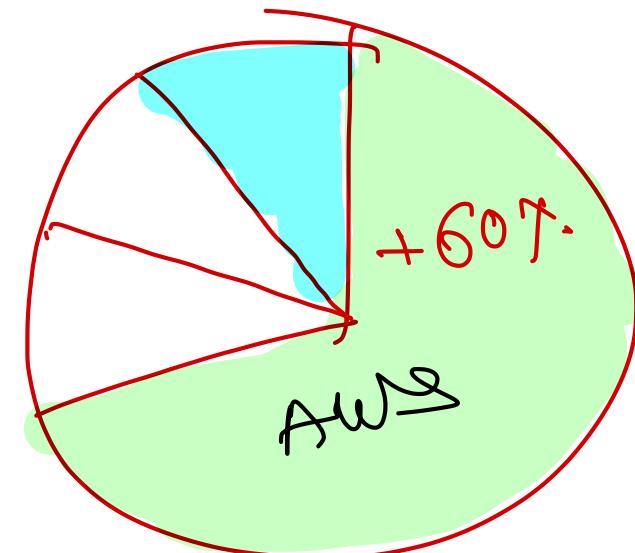
Disadvantages

- Requires a constant Internet connection
- Does not work well with low-speed connections
- Features might be limited
- Stored data might not be secure
- Stored data can be lost
- Each cloud system uses different protocols and different APIs



Cloud Providers

- Amazon Web Services
- Google Cloud Platform
- Microsoft Azure
- Rackspace
- DigitalOcean
- Alibaba Cloud
- Oracle Cloud
- IBM Cloud

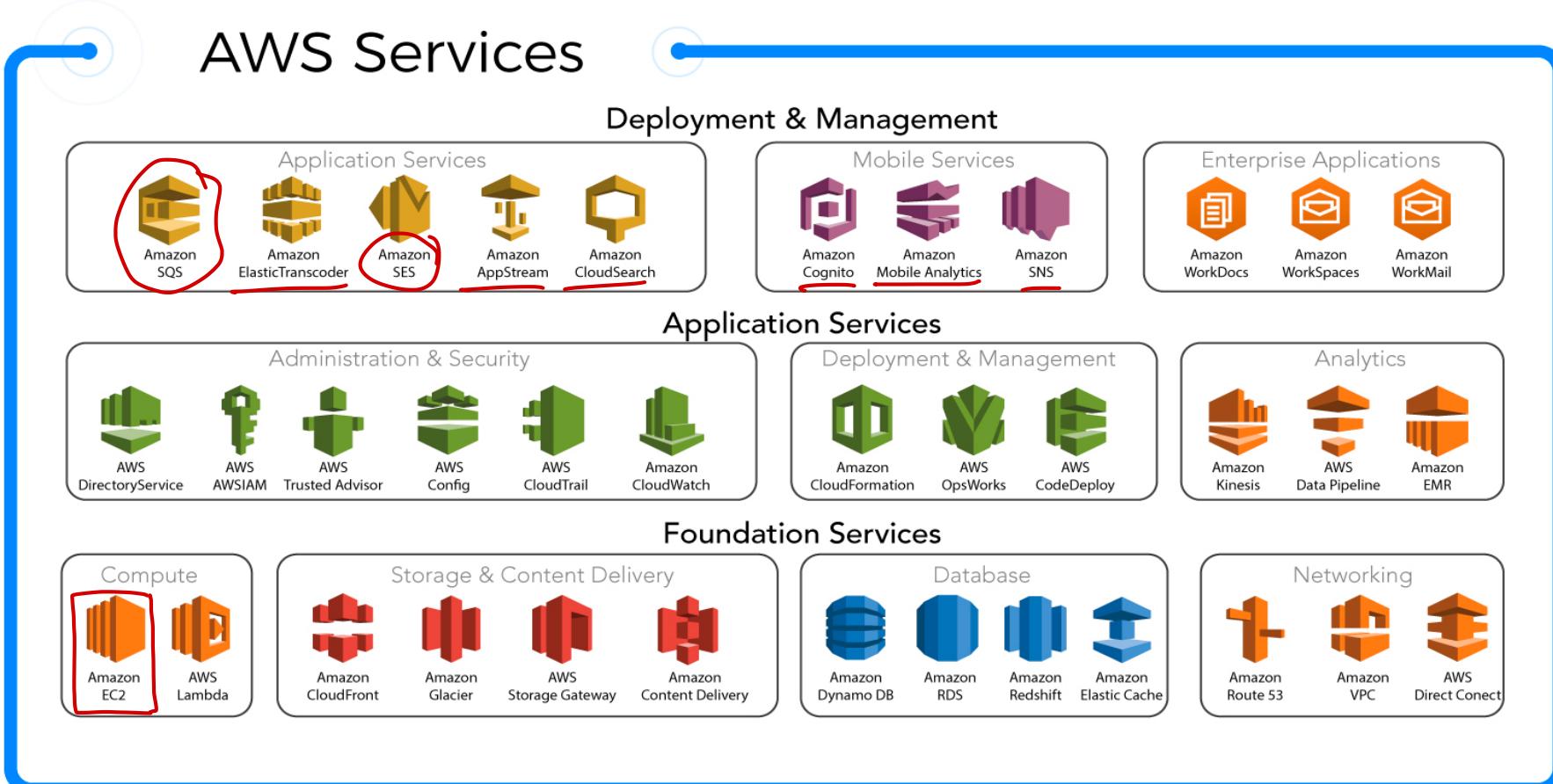


What is AWS ?

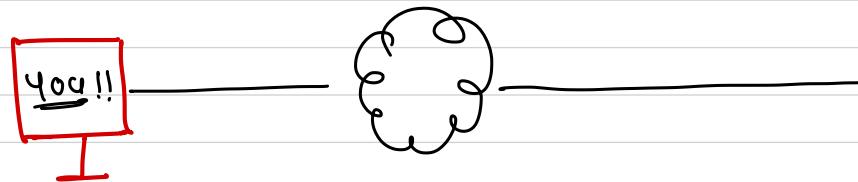
- AWS stands for Amazon Web Services
- Platform that offers flexible, reliable, scalable, easy-to-use and cost-effective cloud computing solutions
- Amazon's cloud implementation
- It's a combination of IaaS, PaaS and SaaS offerings



AWS Services

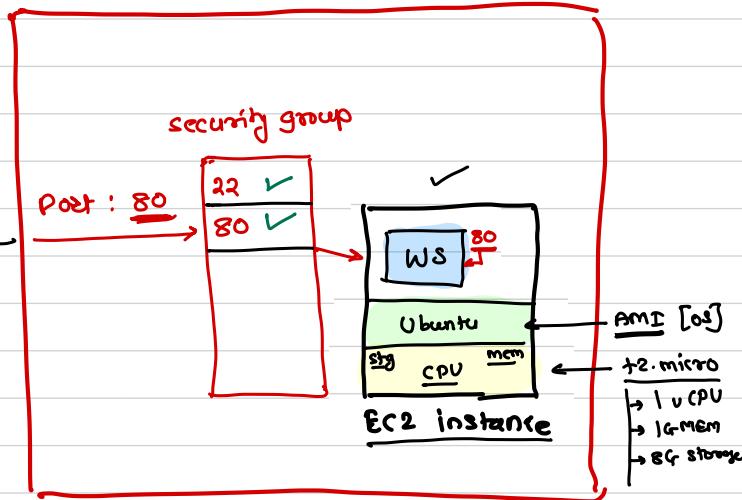


EC2 instance



> ssh -i <priv> ubuntu@<ip>

VPC: virtual Private cloud



Global Infrastructure: Region

- Geographic area having availability zone(s)
- Collection of availability zones that are geographically located close to one other
- Every Region will act independently of the others, and each will contain at least two Availability Zones
- E.g.
 - US East: N. Virginia, Ohio
 - US West: N. California, Oregon
 - Asia Pacific: Mumbai, Seoul, Singapore, Sydney, Tokyo



Global Infrastructure: Availability Zone

- Essentially the physical data centers of AWS
- This is where the actual compute, storage, network, and database resources are hosted that we as consumers provision within our Virtual Private Clouds (VPCs)
- Availability Zones are always referenced by their Code Name, which is defined by the AZs Region Code Name that the AZ belongs to, followed by a letter
- E.g.
 - the AZs within the eu-west-1 region (EU Ireland), are
 - eu-west-1a
 - eu-west-1b
 - eu-west-1c



Global Infrastructure: Edge Locations

- Edge Locations are AWS sites deployed in major cities and highly populated areas across the globe
- Generally used to cache data and reduce latency for end-user access by using the Edge Locations as a global Content Delivery Network (CDN)
- Edge Locations are primarily used by end users who are accessing and using your services
- E.g.
 - Route 53: DNS Lookup
 - CloudFront
 - Content Delivery Network (CDN)
 - Cached contents, streaming distribution, acceleration

