Disclaimer: **These slides are copyrighted and strictly for personal use only**

- This document is reserved for people enrolled into the Ultimate Big Data Masters Program (Cloud Focused) by Sumit Sir

- **Please do not share this document,** it is intended for personal use and exam preparation only, thank you.

- If you've obtained these slides for free on a website that is not the course's website, please reach out to legal@trendytech.in . Thanks!

- All the Best and Happy Learning!

legal@trendytech.in
www.trendytech.in

Challenges of On-Premise Systems

- Involves huge expenditure on infrastructure and resources like servers, cooling equipment, physical premises to accommodate the servers.
- Involves a lot of backend and maintenance activities like Installation of required software and software updations.
- Would require a dedicated IT team to maintain the servers. This adds to the overall business expenses.

After all these challenges, if the requirement increases, scaling up the resources is a tedious task. Also, if after upscaling, the requirement goes down, then the resources are wasted.

A better approach is using Cloud Services

Cloud - Is on-demand availability of computer services like servers, data storage, networking, databases, etc. Cloud is a service delivery model over the internet.

Major Cloud Providers - AWS, Azure, GCP

CapEx Vs OpEx

CapEx / Capital Expenditure

Is a significant initial investment required to setup the infrastructure and resources.

OpEx / Operational Expenditure

Is a nominal regular investment to use or maintain the services without any huge initial investment.

	On-Premise	Cloud
Upfront Cost	Significant	None
Ongoing Cost	Low	Based on Usage
Early Termination	No	Anytime
Maintainence	Significant	Low
Value over Time	Lower	No Change

Cloud Characteristics

- **Agility** : Time required to procure and deploy the resources to serve a request is significantly low in case of Cloud, therefore Cloud is considered to be very Agile.
- **Scalability** : Scaling up/in & down/out (adding or removing) the resources in case of cloud is effortless and straightforward. Two types scaling include
    a) **Vertical Scaling** - Adding more resources to the same machine
    b) **Horizontal Scaling** - Adding more machines or nodes to the cluster which eventually adds more resources.
- **Elasticity** : Ability of the system to automatically scale by dynamically allocating the resources as per the requirement.
- **Fault tolerance** : Ability of the system to function even on failure, that the services are still up and running which can be used seamlessly by the users. Ability to maintain the system up-time while physical and service component failures occur. This can be achieved by replication and rack mechanism.
- **Disaster recovery** : Is a design principle that allows a system to recover from natural or human induced disasters. This can be achieved

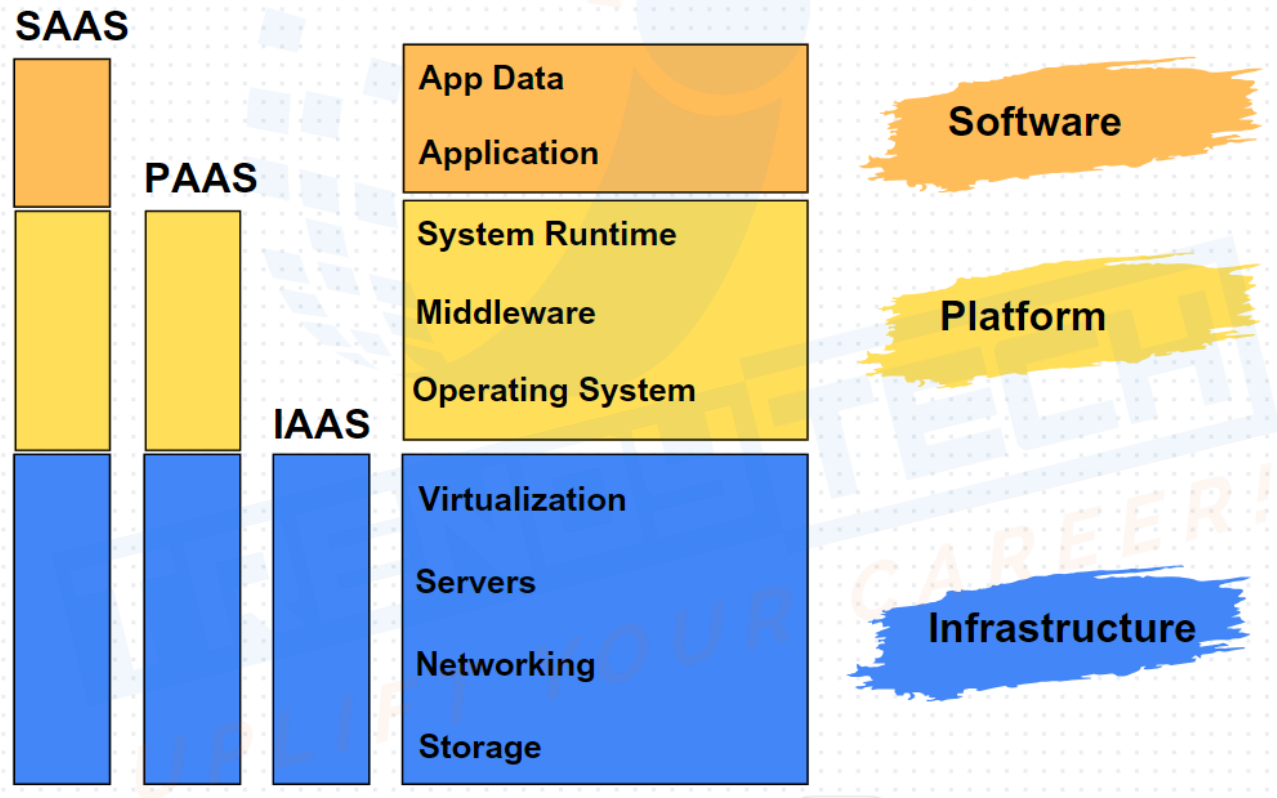
by having a backup datacenter that would become active and serve the request when its counterpart goes down due to a disaster.

- **High Availability** : Implies that the services should be up and running with little or no downtime.

Availability = Total Up-time / (Up-time + Down-time)

Categories of Cloud Services

1. **IAAS** / Infrastructure as a Service - Ex : Virtual Machines
2. **PAAS** / Platform as a Service - Ex : SQL Database
3. **SAAS** / Software as a Service - Ex : Google Drive, Office 365, Outlook, Slack, etc.



Cloud Deployment Models

1. **Public Cloud** - Cloud services that are open to the public over the internet. Data and other information delivered over the internet that is shared among public and different organisations. Ex - AWS, Azure, GCP

Can be a best fit when the data that is being handled is not sensitive or confidential.

Pro - Ease of use, Little or No Upfront Cost, Not much Expertise needed.

Con - Compliance and Security Issues.

2. **Private Cloud** - Cloud services that are open and accessible to users within an organisation.

   Can be a best fit when the data that is being handled is very sensitive or confidential.

   Pro - Highly Compliant and Secure.

   Con - Need Expert In-house Team, Huge Upfront Cost.

3. **Hybrid Cloud** - This environment is a combination that uses both Private and Public Cloud.

Can be a best fit when the data that is being handled is a blend of confidential and non-confidential information.

Pro - Leverages the benefits of both Public and Private Cloud.

Con - Complicated, Requires In-house Expertise, High Upfront Cost.

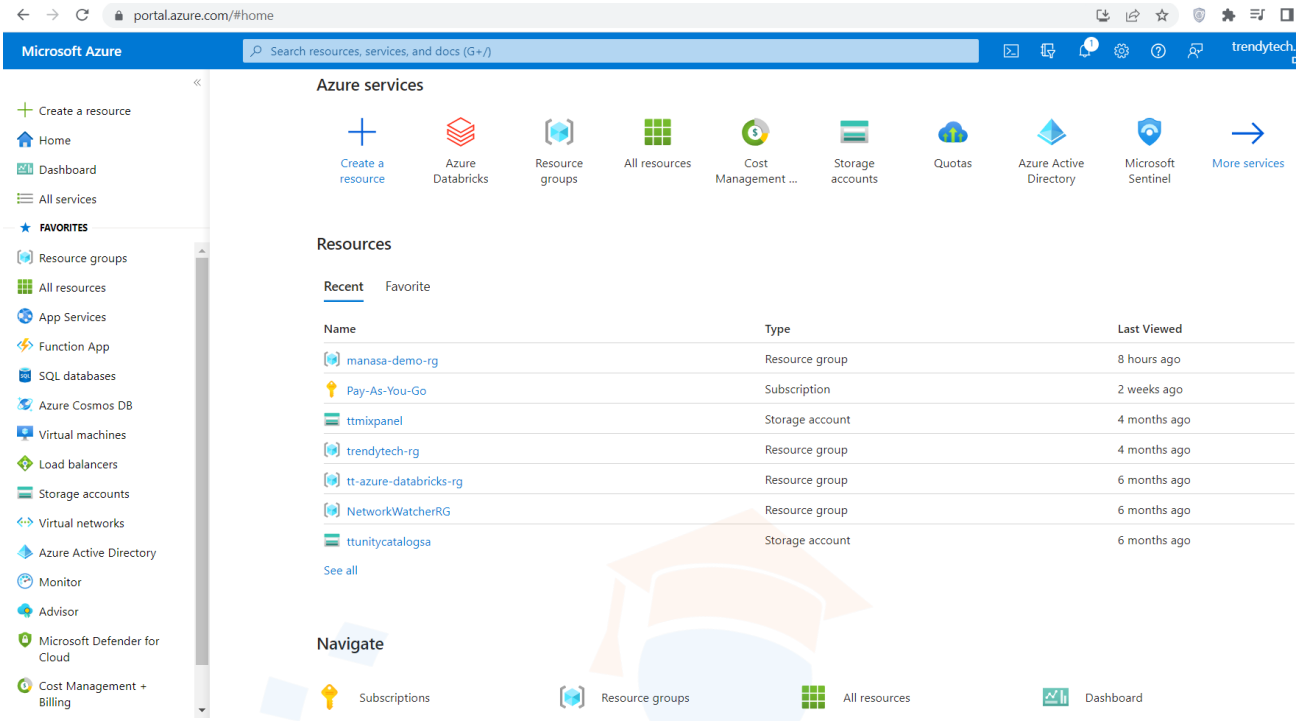
Steps to Create a Free Azure Account

Step 1 - Create a Microsoft Outlook Account.

Step 2 - Create a Microsoft Azure Profile by filling in the basic information.

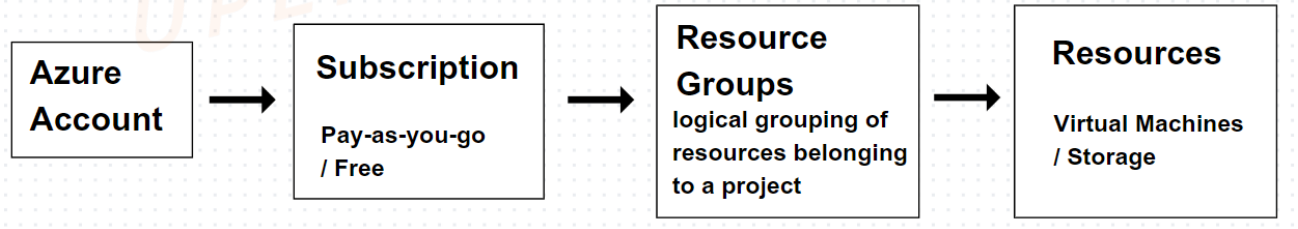
Step 3 - You would need bank card details(ideally credit card) for the “identity verification by card” step.

Step 4 - Once the account is verified, you will have to login and enter “portal.azure.com” in the URL bar to access the azure cloud account and services.



Key Points

- On creation of an account, a subscription is associated.
- Every Subscription has a billing unit.
- Each subscription has a logical unit, Resource Group (logical collection of resources belonging together as part of a project).
- Every Service on Azure cloud is treated as a Resource(An object used to manage the Service)
- Each Resource has to belong to a Resource Group.



Different means through which Resources are created

-API/SDK

-Portal

-Powershell

-CLI / BASH

(Cloud Shell - Provides powershell / bash on the azure cloud)

Good to Know -

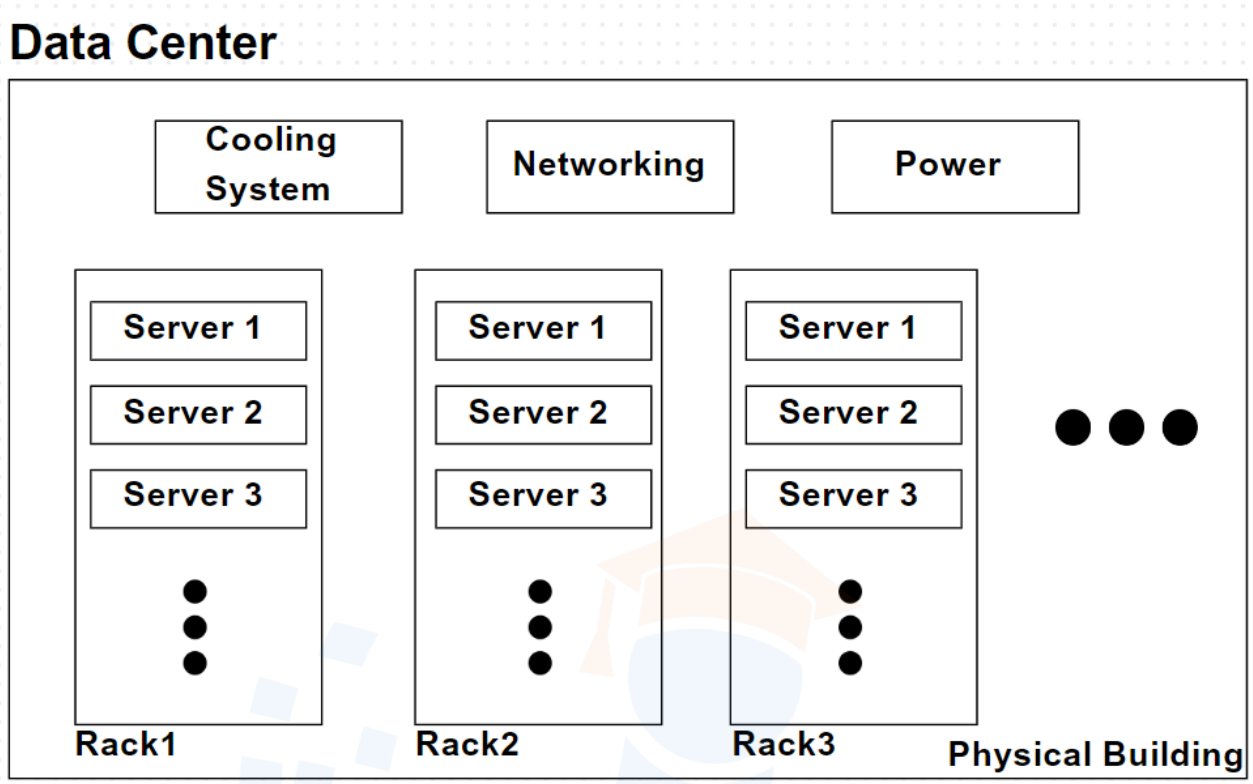
- Each Resource can be a part of only one Resource Group
- Resource Groups cannot be nested.
- Resources can be shuffled among the different Resource Groups.
- Resource Groups have their own location where the metadata is stored.
- Resources belonging to a Resource Group can reside in different locations which are not necessarily in the same location as the Resource Group.

Azure Global Infrastructure

All the services are hosted on a Server present in a physical location.

Data Center

- A physical facility where the servers are present is known as a Data Center.
- It is a unique physical building with many interconnected servers and its own power, cooling, network infrastructure



Region

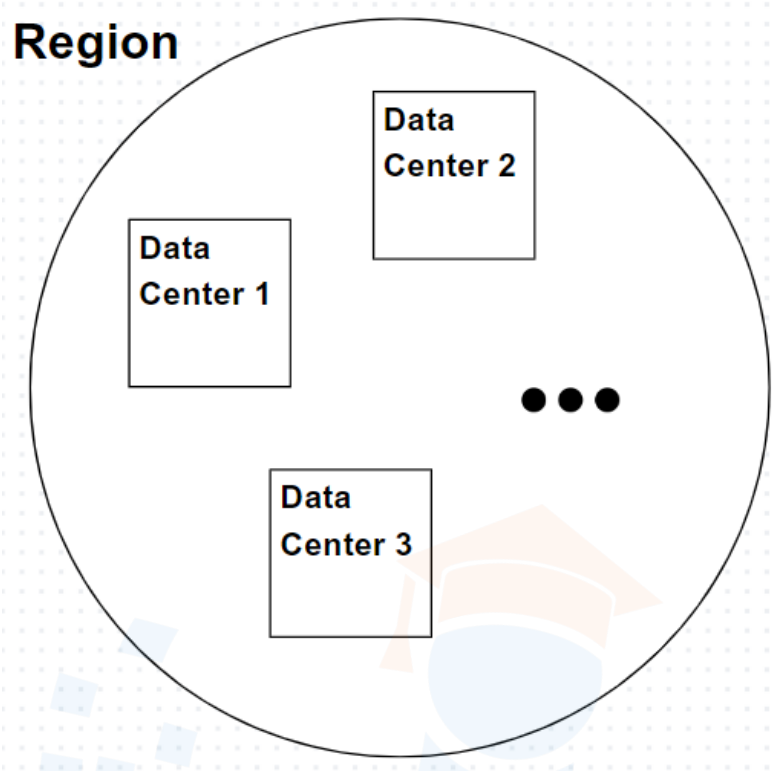
- Geographical area with a Combination of one or more Data Centers.

  Ex - East US, West India, …

- All the Data Centres within a region are interconnected by low latency network (< 2 milliseconds)
- Every Service requires a Region to be selected while resource creation. The services will be procured in the specified Region. Certain Services are not Region bound which doesn't require a Region to be selected.

Criterias to be considered for Region Selection

- The Region should be near to the end-users for maintaining lower latency.
- Not all services are available in all regions. Therefore, it is important to select the Regions which provide the required services. (Best practice is to select the nearest region that provides the maximum services.)
- Prices of services vary across the Regions. Select Regions that provide cost-effective and maximum services.

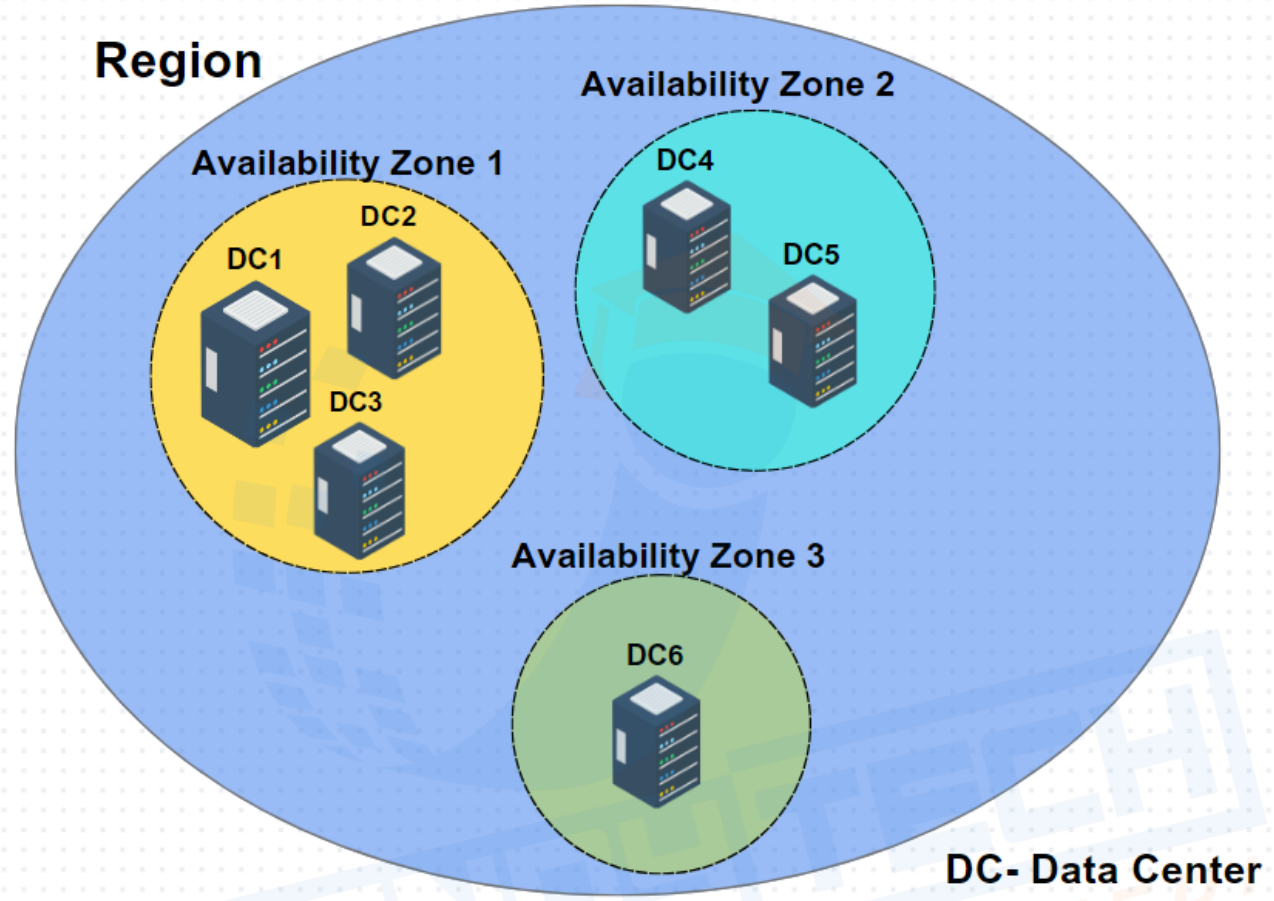


Availability Zone

- While creating the Resources, it was required to select the Region where the services were to be deployed. However, there was no option of choosing the Data Centers within the region to deploy the services.
- Availability Zones are required when we need the control of specifying the Data Centers where the services have to be deployed.
- This ensures high availability of services even on Data Center failures.
- A Region is Zone Enabled if it contains 3 or more Data Centers. Not all regions are Zone Enabled.
- Availability Zones helps in avoiding the services from going down completely in case of Data Center level failures.
- Faults and Updates are tolerated in case of Availability Zones.

Category of Services in Availability Zone

1. Zonal Services - Services in which an option for choosing the Zones is provided. Ex : VMs, Managed Disks, etc.
2. Zone-Redundant Services - Such services will be deployed on all Zones of the Region. Ex : SQL Database, Storage Service, etc.

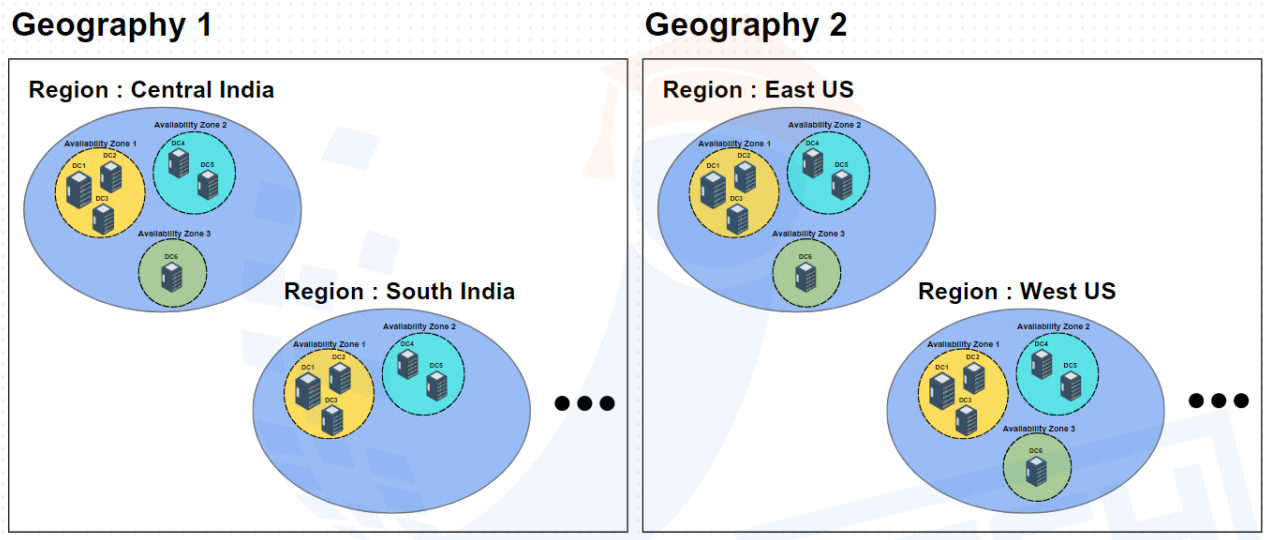


Region Pairs

- Are required to handle Region Level Failures caused due to natural or human induced disasters.
- Each Region in Azure is paired with exactly one other Region. There is no flexibility of choosing the Region Pairs and each pair resides in the same geography.
- Azure Traffic Manager redirects the requests to the corresponding counter-pair of the failed region.

Geography

- Geography contains one or more Regions.
- It ensures Data Residency and Compliance requirements are.
- Highly confidential data belonging to a specific geography cannot be leaked to other geographies. Such critical data has to be securely stored in the respective geography only.
- Each Region can belong to only one Geography.
- Azure geography is available to only the US Government.



Geography -> Regions -> Region Pairs -> Availability Zone -> Data Center -> Servers

Azure Virtual Machine

Problem Statement

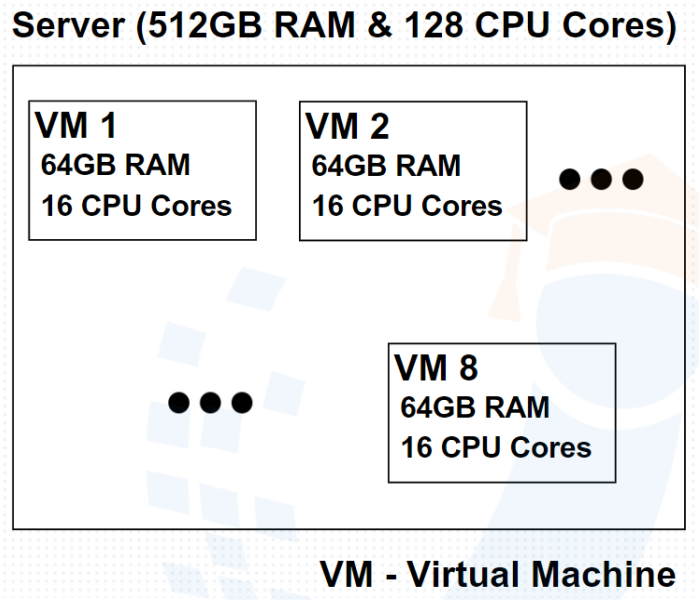
Challenges with running multiple applications on a single powerful server with high-end configurations (Ex : 512GB RAM and 128 CPU Cores)

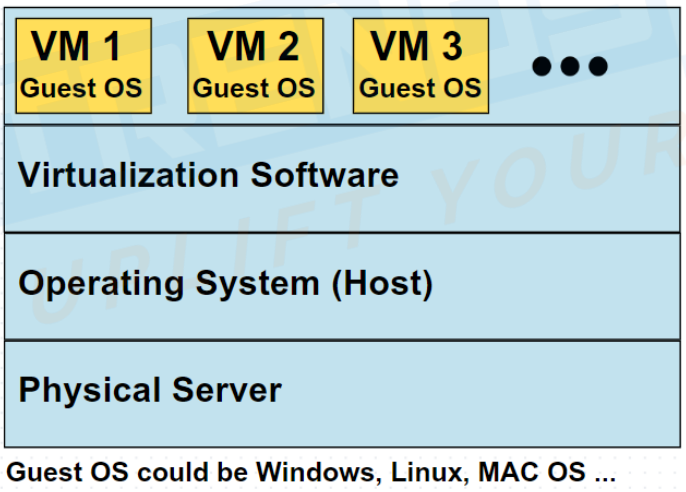
- Applications can interfere and collide with each other as all the applications are using the same cores and file systems.

Running a single application on such a powerful machine would lead to underutilization of the computing power.

Solution

- To overcome the above challenges, Virtual Machines are used.
- Consider the above example of Server with 512GB RAM and 128 CPU Cores, there could be 8 VMs created with 64GB RAM and 16 CPU Cores on one large server.





When to use VM ?

VMs are used when there is a need for complete control over the resources. IAAS deployment model caters to this need of requiring only the Virtual Machine without any softwares installed.

Example Use-case of VM - Migration projects on legacy systems which require its own customized softwares.

Virtual Machine Scale Set VMSS

- Helps us to easily create and manage multiple VMs on a click of a button.
- All the VMs in the Scale Set have identical configurations.
- Easier to scale up or scale down the VMs as per the requirement.

Availability Set

With Availability Set, we are deploying multiple VMs across different Fault and Update domains. This is to ensure that any power failures(Fault) or system software updates(Update) do not impact all the VMs.

Fault Domain - It is a physical grouping that is internally divided into racks.

Update Domain - It is a logical grouping of servers on which patching and software updates are carried out. Any update errors on one logical group of servers will not impact the other groups.

Availability Set Vs Availability Zone

Availability Set protects from failures within the same Data Center.

Availability Zone protects from Data Center Level failures.