Azure Intro, VMs, VM Scale sets - Week 1

Cloud Computing is delivery of on-demand computing resources over the internet on a pay-per-use basis .

Cloud Models

Cloud Models are defined according to where the infrastructure for the deployment resides and who has control over that infrastructure .

- Azure supports three approaches to deploying cloud resources - public, private, and the hybrid cloud.
- Selecting between them will change several factors of the services you move into Azure including cost, maintenance requirements, and security.

Public cloud

- Public clouds are the most common way of deploying cloud computing. Services are offered over the public internet and available to anyone who wants to purchase them.
- The cloud resources such as servers and storage are owned and operated by a third-party cloud service provider and delivered over the internet.
- Services may be free or sold on demand, allowing customers to pay only per usage for the CPU cycles, storage, or bandwidth they consume. Microsoft Azure is an example of a public cloud.

Private cloud

- A private cloud consists of computing resources used exclusively by users from one business or organization. It can be physically located at your organization's on-site datacenter, or it can be hosted by a third-party service provider.
- The term private cloud should not be considered a rebranding of traditional on-premises data centers.
- A private cloud uses on-premises infrastructure and services to provide similar benefits of the public cloud. It uses an abstraction platform to provide *cloud-like* services such as Kubernetes clusters or a complete cloud environment like Azure

Stack. The organization is responsible for the purchase, configuration, and maintenance of the hardware.

- Communication between the systems is usually on the network infrastructure that the business owns and maintains. For example, a private internal network or a dedicated fiber optic connection between buildings.

Hybrid cloud

- A hybrid cloud is a computing environment that combines a public cloud and a private cloud by allowing data and applications to be shared between them.
- When computing and processing demand fluctuates, hybrid cloud computing gives businesses the ability to seamlessly scale their on-premises infrastructure up to the public cloud to handle any overflow - without giving third-party data centers access to the entirety of their data.
- Organizations gain the flexibility and computing power of the public cloud for basic and non-sensitive computing tasks, while keeping business-critical applications and data on-premises, safely behind a company firewall.

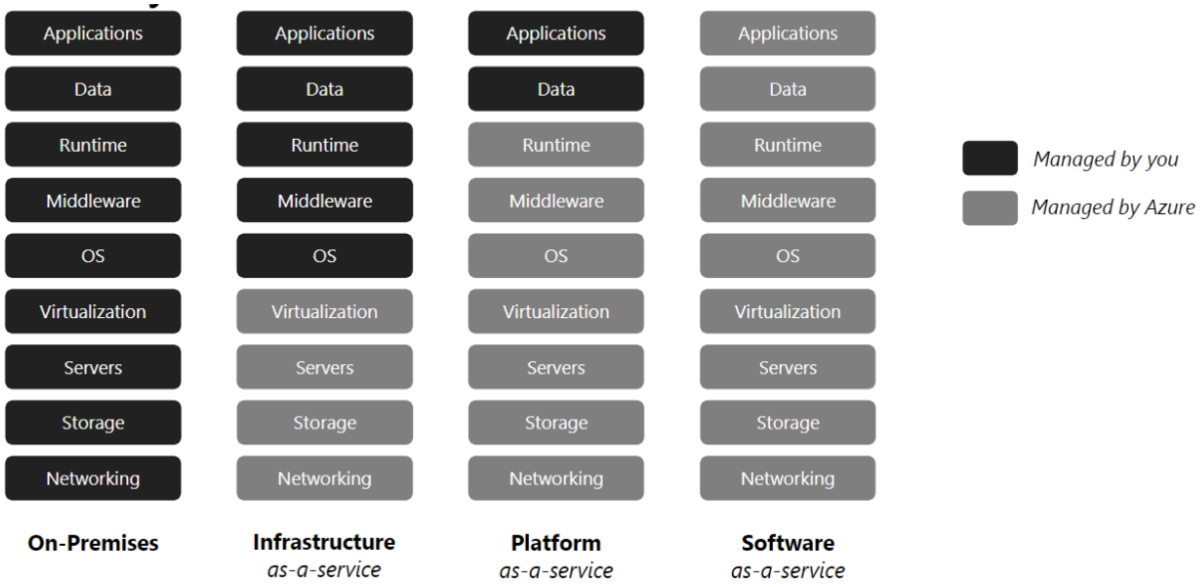
Please refer to the given link for more information - Azure solutions for public, private, and hybrid cloud .

Delivery Models

- Cloud computing types are service deployment models which let you choose the level of control over your information and types of services you need to provide. There are three main types of cloud computing services, sometimes called the cloud computing stack because they build on top of one another.

- The first cloud computing type is infrastructure-as-a-service (IaaS), which is used for Internet-based access to storage and computing power.
    - ❖ The most basic category of cloud computing types, IaaS lets you rent IT infrastructure - servers and virtual machines, storage,

networks and operating systems - from a cloud provider on a pay-as-you-go basis.



- The second cloud computing type is platform-as-a-service (PaaS) which gives developers the tools to build and host web applications.
    - ❖ PaaS is designed to give users access to the components they require to quickly develop and operate web or mobile applications over the Internet, without worrying about setting up or managing the underlying infrastructure of servers, storage, networks and databases.

- The third cloud computing type is software-as-a-service (SaaS) which is used for web-based applications.
    - ❖ SaaS is a method for delivering software applications over the Internet where cloud providers host and manage the software applications making it easier to have the same application on all of your devices at once by accessing it in the cloud.

Regions and availability zones

Source : https://azure.microsoft.com/en-in/global-infrastructure/geographies/

- Azure regions and availability zones are designed to help you achieve resiliency and reliability for your business-critical workloads. Azure maintains multiple geographies.
- These discrete demarcations define disaster recovery and data residency boundaries across one or multiple Azure regions. Maintaining many regions ensures customers are supported across the world.

Please refer to the given link for more information -
https://docs.microsoft.com/en-us/azure/availability-zones/az-overview
https://docs.microsoft.com/en-us/learn/modules/azure-architecture-fundamentals/regions-availability-zones

Types of Services

- Compute
- Networking
- Storage
- Databases
- Web
- IoT / Event
- Big Data / Analytics
- Identity
- AI
- Monitoring
- DevOps

Please refer to the links given below for more information
https://docs.microsoft.com/en-us/learn/modules/azure-compute-fundamentals/overview
https://docs.microsoft.com/en-us/azure/networking/fundamentals/networking-overview
https://docs.microsoft.com/en-us/azure/storage/common/storage-introduction
https://docs.microsoft.com/en-us/azure/app-service/overview
https://docs.microsoft.com/en-us/azure/iot-fundamentals/iot-introduction
https://docs.microsoft.com/en-us/azure/architecture/guide/architecture-styles/big-data
https://docs.microsoft.com/en-us/azure/active-directory/fundamentals/active-directory-ops-guide-iam
https://docs.microsoft.com/en-us/learn/modules/introduction-to-ai-technology/3-introduction-to-ai-technology
https://docs.microsoft.com/en-us/azure/azure-monitor/overview
https://docs.microsoft.com/en-us/azure/devops/user-guide/what-is-azure-devops?view=azure-devops

Azure Components

- **Azure tenant** - A dedicated and trusted instance of Azure AD. The tenant is automatically created when your organization signs up for a Microsoft cloud service subscription.These subscriptions include Microsoft Azure, Microsoft Intune, or Microsoft 365. An Azure tenant represents a single organization.

- **Azure Active Directory** - Azure Active Directory (Azure AD) is a cloud-based identity and access management service. This service helps your employees access external resources, such as Microsoft 365, the Azure portal, and thousands of other SaaS applications.

    - ❖ Azure AD also helps them access internal resources. These are resources like apps on your corporate network and intranet, along with any cloud apps developed by your own organization.

For more information, please refer to the given link -
https://docs.microsoft.com/en-us/azure/active-directory/fundamentals/active-directory-whatis

- **Azure management groups** - If your organization has many Azure subscriptions, you may need a way to efficiently manage access, policies, and compliance for those subscriptions. *Management groups* provide a governance scope above subscriptions.
    - ❖ You organize subscriptions into management groups and the governance conditions you apply cascade by inheritance to all associated subscriptions.

For more information, please refer to the given link -
https://docs.microsoft.com/en-us/azure/governance/management-groups/overview

- **Subscriptions -** A subscription is an agreement with Microsoft to use one or more Microsoft cloud platforms or services, for which charges accrue based on either a per-user license fee or on cloud-based resource consumption.

Please refer to the link given below for more information -
https://docs.microsoft.com/en-us/microsoft-365/enterprise/subscriptions-licenses-accounts-and-tenants-for-microsoft-cloud-offerings?view=o365-worldwide

- **Azure resource groups** - A resource group is a container that holds related resources for an Azure solution.

    - ❖ The resource group can include all the resources for the solution, or only those resources that you want to manage as a group. You decide how you want to allocate resources to resource groups based on what makes the most sense for your organization.

    - ❖ The resource group stores metadata about the resources. Therefore, when you specify a location for the resource group, you are specifying where that metadata is stored.

Please refer to the link given below for managing Azure resource groups using the Azure portal
https://docs.microsoft.com/en-us/azure/azure-resource-manager/management/manage-resource-groups-portal

- **Regions and availability zones**
  Please refer to the link given below for more information-
  https://docs.microsoft.com/en-us/azure/availability-zones/az-overview#azure-regions-with-availability-zones

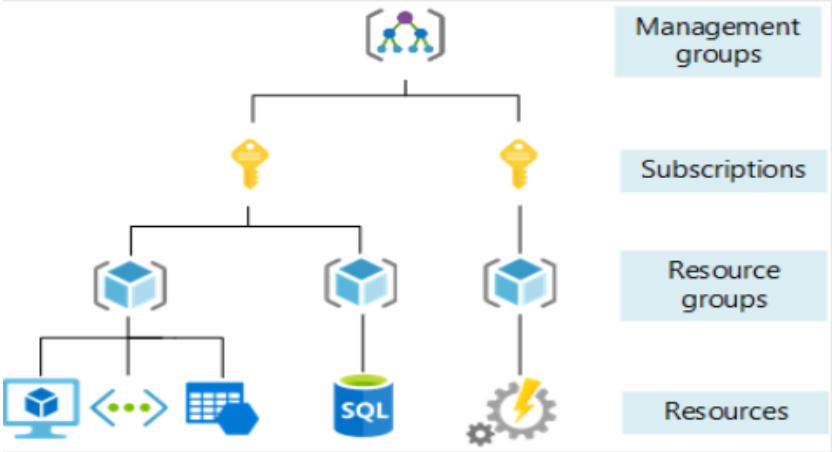
- **Hierarchy**

Azure provides four levels of management scope:

    - ❖ Management groups

- Subscriptions
- Resource groups
- Resources

The following image shows the relationship between these levels.



Source - https://docs.microsoft.com/en-us/azure/cloud-adoption-framework/ready/azure-setup-guide/organize-resources?tabs=AzureManagementGroupsAndHierarchy

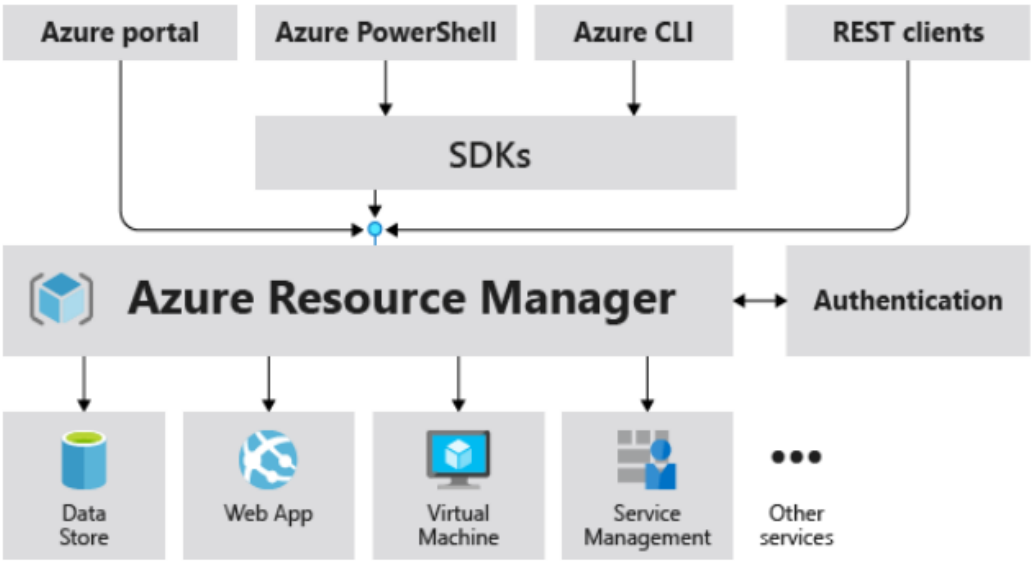
- **Management groups** help you manage access, policy, and compliance for multiple subscriptions. All subscriptions in a management group automatically inherit the conditions applied to the management group.

- **Subscriptions** logically associate user accounts with the resources they create. Each subscription has limits or quotas on the amount of resources it can create and use. Organizations can use subscriptions to manage costs and the resources created by users, teams, or projects.

- **Resource groups** are logical containers where you can deploy and manage Azure resources like web apps, databases, and storage accounts.

- **Resources** are instances of services that you can create, like virtual machines, storage, or SQL databases.

Azure Resource Manager



- Azure Resource Manager is the deployment and management service for Azure. It provides a management layer that enables you to create, update, and delete resources in your Azure account.

- You use management features, like access control, locks, and tags, to secure and organize your resources after deployment.

Please refer to the given link for more information -
https://docs.microsoft.com/en-us/azure/azure-resource-manager/management/overview

Networking Components

- **Azure Virtual Network (VNet)** is the fundamental building block for your private network in Azure.
  - ❖ VNet enables many types of Azure resources, such as Azure Virtual Machines (VM), to securely communicate with each other, the internet, and on-premises networks.
  - ❖ VNet is similar to a traditional network that you'd operate in your own data center, but brings with it additional benefits of Azure's infrastructure such as scale, availability, and isolation.
  - ❖ It Defines a range of private IP addresses.

Please refer to the given link for more information-
-https://docs.microsoft.com/en-us/azure/virtual-network/virtual-networks-overview

- **Network interface card** enables an Azure Virtual Machine to communicate with internet, Azure, and on-premises resources(Outside network). When creating a virtual machine using the Azure portal, the portal creates one network interface with default settings for you.
    - ❖ Multiple NICs can be assigned to a VM

Please refer to the given link for more information-
https://docs.microsoft.com/en-us/azure/virtual-network/virtual-network-network-interface

- **Public IP addresses** allow Internet resources to communicate inbound to Azure resources and they are Optionally assigned to NIC(associated with VM).
    - ❖ Public IP addresses enable Azure resources to communicate to Internet and public-facing Azure services.
    - ❖ A resource without a public IP assigned can communicate outbound.
    - ❖ Azure dynamically assigns an available IP address that isn't dedicated to the resource.

Please refer to the given link for more information
https://docs.microsoft.com/en-us/azure/virtual-network/ip-services/public-ip-addresses

- **Network security group** contains security rules that allow or deny inbound network traffic to, or outbound network traffic from, several types of Azure resources.
    - ❖ For each rule, you can specify source and destination, port, and protocol.
    - ❖ Can be applied at subnet or NIC(associated with VM)level.

Please refer to the given link for more information
https://docs.microsoft.com/en-us/azure/virtual-network/network-security-groups-overview

- **Azure Firewall** is a cloud-native and intelligent network firewall security service that provides the best of breed threat protection for your cloud workloads running in Azure.

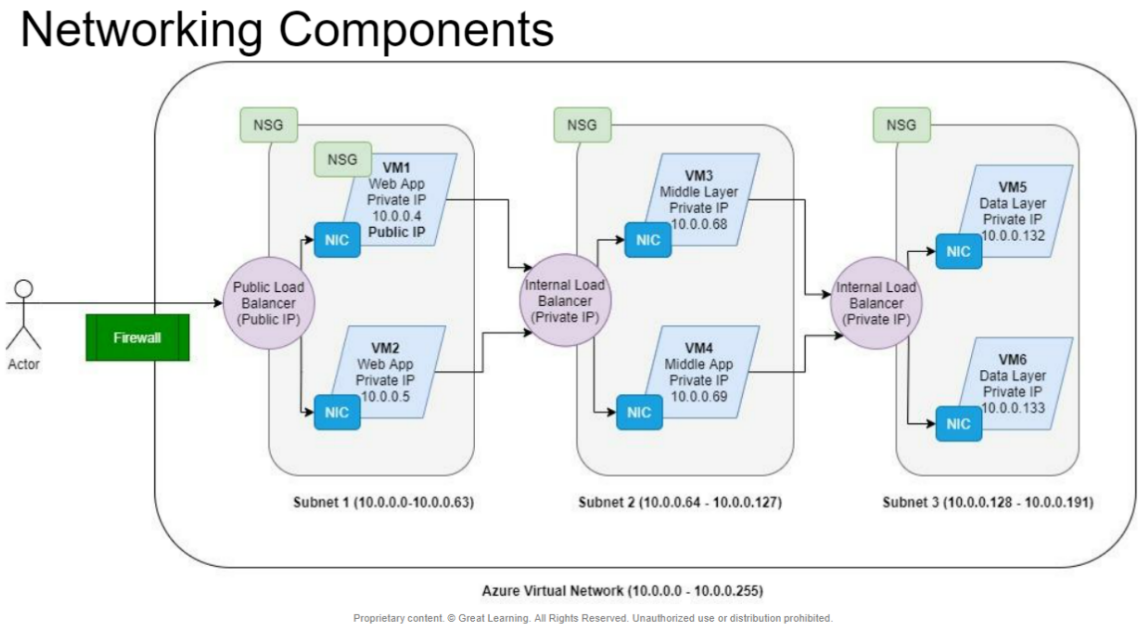
- ❖ It's a fully stateful, firewall as a service with built-in high availability and unrestricted cloud scalability.

Please refer to the given link for more information
https://docs.microsoft.com/en-us/azure/firewall/overview

- **Azure Load Balancer** operates at layer 4 of the Open Systems Interconnection (OSI) model.
  - ❖ *Load balancing* refers to evenly distributing load (incoming network traffic) across a group of backend resources or servers.
  - ❖ It's the single point of contact for clients. Load balancer distributes inbound flows that arrive at the load balancer's front end to backend pool instances.
  - ❖ These flows are according to configured load-balancing rules and health probes. The backend pool instances can be Azure Virtual Machines or instances in a virtual machine scale set.
  - ❖ **Azure Application Gateway** is a web traffic load balancer that enables you to manage traffic to your web applications.
    - ➢ Traditional load balancers operate at the transport layer (OSI layer 4 - TCP and UDP) and route traffic based on source IP address and port, to a destination IP address and port.
  - ❖ **Azure Traffic Manager** is a DNS-based traffic load balancer.
    - ➢ This service allows you to distribute traffic to your public facing applications across the global Azure regions.
    - ➢ Traffic Manager also provides your public endpoints with high availability and quick responsiveness.
  - ❖ **Azure Front Door** is a global, scalable ent-ry-point that uses the Microsoft global edge network to create fast, secure, and widely scalable web applications.
    - ➢ With Front Door, you can transform your global consumer and enterprise applications into robust, high-performing personalized modern applications with contents that reach a global audience through Azure.

- **Subnet** is a network inside a network . A virtual network can be segmented into one or more subnets up to the limits.
    - ❖ Each subnet must have a unique address range, specified in CIDR format, within the address space of the virtual network.
    - ❖ The address range cannot overlap with other subnets in the virtual network.

Please refer to the given links for more information -
https://docs.microsoft.com/en-us/azure/load-balancer/load-balancer-overview
https://docs.microsoft.com/en-us/azure/application-gateway/overview
https://docs.microsoft.com/en-us/azure/traffic-manager/traffic-manager-overview
https://docs.microsoft.com/en-us/azure/frontdoor/front-door-overview
https://docs.microsoft.com/en-us/azure/virtual-network/virtual-network-vnet-plan-design-arm



- **Azure Virtual Machines (VM)** is one of several types of on-demand, scalable computing resources that Azure offers. Typically, you choose a VM when you need more control over the computing environment than the other choices offer.

- To create a Virtual Machine, deploy:

- ❖ Virtual Network
- ❖ Subnets
- ❖ Network Security Group
- ❖ Public IP Address
- ❖ Network Interface Card
- ❖ Disks
- ❖ Azure Virtual Machine

These articles give you information about what you should consider before you create a VM, how you create it, and how you manage it.

https://docs.microsoft.com/en-us/azure/virtual-machines/linux/overview
https://docs.microsoft.com/en-us/azure/virtual-machines/windows/overview

- **VM Series**(Different series of machines)
  - ❖ A (entry level)
  - ❖ Bs (economic)
  - ❖ D (general purpose)
  - ❖ E (in-memory, hyper-threaded)
  - ❖ F (compute optimized)
  - ❖ H (HPC computing)
  - ❖ M (memory optimized)
  - ❖ N (GPU-enabled)

➢ Based on configuration of physical servers

➢ Uses different series of Intel and AMD processors

➢ Select VM series based on the type of workload

Please refer to the given link for more information
https://docs.microsoft.com/en-us/azure/virtual-machines/sizes

- **VM Sizes**
  - ❖ Each VM series has different sizes of VMs
  - ❖ Size include
    - ➢ vCPUs

- Memory
- Max number of disks
- Max IOPS

- **Pricing Options**

Pay-as-you-go	• Pay-per-usage model • Billed per second
Reserved VMs	• Upfront purchase in a region • Provide cost savings
Spot VMs	• Use unused capacity in Azure. Can be taken back by Azure • Highly discounted
Azure Hybrid Benefit	• Use existing Windows, SQL Server, RedHat Linux, SUSE Linux licenses to save cost

Please refer to the given link for more information -
https://spot.io/resources/azure-pricing-the-complete-guide/

- **VM Images**

- VM image has OS and other software preloaded
- Variety of images are available in the marketplace
- Build and use your own custom image

- **VM Storage**

- OS Hard Disk
- Temporary Hard Disk
- Attach multiple data disks (optional)

- **High Availability for VMs**

- ❖ An availability set is a logical grouping of VMs that allows Azure to understand how your application is built to provide for redundancy and availability.
  - ➢ Provides high availability for VMs within one datacenter
  - ➢ Fault Domains
    - ➔ Logical grouping of hardware that share common power source and network switch
    - ➔ Helps during unplanned hardware events
  - ➢ Update Domains
    - ➔ Logical grouping of VMs and underlying hardware that are updated (and rebooted) at the same time
    - ➔ Helps during planned maintenance events
  - ➢ VM can be assigned to only one Availability Set, and during creation only

- ❖ Factors affecting availability
  - ➢ Application failures
  - ➢ Within Datacenter
    - ➔ Hardware Failure
      - ☐ Leads to unexpected application downtime
    - ➔ Unplanned Hardware Maintenance
      - ☐ When hardware might fail or is about to fail
      - ☐ Leads to poor performance or downtime
    - ➔ Planned Maintenance

Please refer to the given link for more information
https://docs.microsoft.com/en-us/azure/virtual-machines/availability-set-overview

- **Azure availability zones** are physically separate locations within each Azure region that are tolerant to local failures.
  - ❖ Failures can range from software and hardware failures to events such as earthquakes, floods, and fires.
  - ❖ Provides high availability for VMs within one region
  - ❖ VM can be assigned to Availability Zone during creation only

- ❖ Availability Zones and Sets options cannot be used together for a VM

Please refer to the given link for more information
https://docs.microsoft.com/en-us/azure/availability-zones/az-overview

- **Azure Virtual Machine Scale Sets** let you create and manage a group of identical and load-balanced VMs.The number of VM instances can automatically increase or decrease in response to demand or a defined schedule
  - ❖ Scale sets provide high availability to your applications, and allow you to centrally manage, configure, and update a large number of VMs.
  - ❖ With virtual machine scale sets, you can build large-scale services for areas such as compute, big data, and container workloads.
  - ❖ Maintain consistent configuration across all VMs
  - ❖ Provides high availability to the applications

  - ❖ **Azure virtual machine (VM) extensions** are small applications that provide post-deployment configuration and automation tasks on Azure VMs
    - ➢ For example, if a virtual machine requires software installation, antivirus protection, or the ability to run a script inside it, you can use a VM extension.
    - ➢ You can run Azure VM extensions by using the Azure CLI, PowerShell, Azure Resource Manager templates (ARM templates), and the Azure portal.
    - ➢ You can bundle extensions with a new VM deployment or run them against any existing system.
    - ➢ Available for VMs and VM Scale Sets

Please refer to the given link for more information
https://docs.microsoft.com/en-us/azure/virtual-machine-scale-sets/overview

- **Compare Scale Sets with VMs in Availability Sets/Zones**

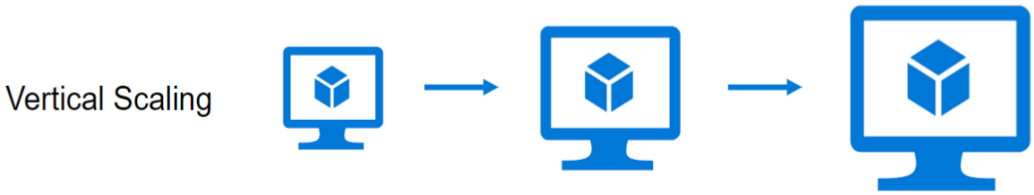
VMs in Availability Set / Zone

- Configure Fault & Update Domains
- Each VM to be created separately
- Size/config of VMs may be different
- OS image may be different
- Apps needs to be installed separately
- Manual scaling
- Add VM to load balancer manually
- Same datacenter in Availability Set and different datacenters for Availability Zones

Virtual Machine Scale Sets

- FDs & UDs are auto managed
- VMs can be created as a group
- Size/config of VMs is same
- OS image is same on each VM
- Apps can be installed using Extensions
- Manual and auto scaling
- VM can be auto added to load balancer
- Can deploy in different datacenters (AZ)

- **Scaling Virtual Machine Scale Sets**
  Autoscaling is the process of dynamically allocating resources to match performance requirements. As the volume of work grows, an application may need additional resources to maintain the desired performance levels and satisfy service-level agreements (SLAs).
  - ❖ There are two main ways that an application can scale:
    - ➢ **Vertical scaling**, also called scaling up and down, means changing the capacity of a resource.
      - ➔ For example, you could move an application to a larger VM size.
      - ➔ Vertical scaling often requires making the system temporarily unavailable while it is being redeployed. Therefore, it's less common to automate vertical scaling.
    - ➢ **Horizontal scaling**, also called scaling out and in, means adding or removing instances of a resource.
      - ➔ The application continues running without interruption as new resources are provisioned.
      - ➔ When the provisioning process is complete, the solution is deployed on these additional resources. If demand drops, the additional resources can be shut down cleanly and deallocated.



- **Autoscale with Azure virtual machine scale sets**
  - ❖ An Azure virtual machine scale set can automatically increase or decrease the number of VM instances that run your application
  - ❖ This automated and elastic behavior reduces the management overhead to monitor and optimize the performance of your application.
  - ❖ You create rules that define the acceptable performance for a positive customer experience.
  - ❖ When those defined thresholds are met, autoscale rules take action to adjust the capacity of your scale set.

Please refer to the given link for more information
https://docs.microsoft.com/en-us/azure/virtual-machine-scale-sets/virtual-machine-scale-sets-autoscale-overview

- **Scaling Profiles**
  There are three types of Profiles:
  - ❖ **Regular profile** is the most common profile. If you don't need to scale your resource based on the day of the week, or on a particular day, you can use a regular profile.
    - ➢ This profile can then be configured with metric rules that dictate when to scale out and when to scale in. You should only have one regular profile defined.
  - ❖ **Fixed date profile** is for special cases.
    - ➢ The profile is configured to run only on the event's day. For any other day, Autoscale uses the regular profile.

- ❖ **Recurrence profile** enables you to ensure that this profile is always used on a particular day of the week.
  - ➢ Recurrence profiles only have a start time. They run until the next recurrence profile or fixed date profile is set to start.
  - ➢ An Autoscale setting with only one recurrence profile runs that profile, even if there is a regular profile defined in the same setting.

Please refer to the the given link for more information
https://docs.microsoft.com/en-us/azure/azure-monitor/autoscale/autoscale-understanding-settings#autoscale-profiles

- **Protection policy for VMSS**
  Scale sets provide two types of instance protection capabilities
  - ❖ **Protect from scale-in**
    - ➢ Enabled through **protectFromScaleIn** property on the scale set instance
    - ➢ Protects instance from Autoscale initiated scale-in
    - ➢ User-initiated instance operations (including instance delete) are **not blocked**
    - ➢ Operations initiated on the scale set (upgrade, reimage, deallocate, etc.) are **not blocked**

  - ❖ **Protect from scale set actions**
    - ➢ Enabled through **protectFromScaleSetActions** property on the scale set instance
    - ➢ Protects instance from Autoscale initiated scale-in
    - ➢ Protects instance from operations initiated on the scale set (such as upgrade, reimage, deallocate, etc.)
    - ➢ User-initiated instance operations (including instance delete) are **not blocked**
    - ➢ Delete of the full scale set is **not blocked**

Please refer to the given link for more details
https://docs.microsoft.com/en-us/azure/virtual-machine-scale-sets/virtual-machine-scale-sets-instance-protection

- **Scale-in policies with Azure virtual machine scale sets**

  A virtual machine scale set deployment can be scaled-out or scaled-in based on an array of metrics, including platform and user-defined custom metrics.

  - ❖ The scale-in policy feature provides users a way to configure the order in which virtual machines are scaled-in, by way of three scale-in configurations:
    - ➢ Default scale-in policy
      - ➔ By default, virtual machine scale sets apply this policy to determine which instance(s) will be scaled in.
    - ➢ NewestVM scale-in policy
      - ➔ This policy will delete the newest created virtual machine in the scale set, after balancing VMs across availability zones (for zonal deployments).
      - ➔ Enabling this policy requires a configuration change on the virtual machine scale set model.

Event	Instance IDs in Zone1	Instance IDs in Zone2	Instance IDs in Zone3
Initial	3, 4, 5, 10	2, 6, 9, 11	1, 7, 8
Scale-in	3, 4, 5, 10	2, 6, 9, *11*	1, 7, 8
Scale-in	3, 4, 5, *10*	2, 6, 9	1, 7, 8
Scale-in	3, 4, 5	2, 6, *9*	1, 7, 8
Scale-in	3, 4, 5	2, 6	1, 7, *8*
Scale-in	3, 4, *5*	2, 6	1, 7
Scale-in	3, 4	2, 6	1, *7*

Source : https://docs.microsoft.com/en-us/azure/virtual-machine-scale-sets/virtual-machine-scale-sets-instance-protection

- ➢ OldestVM scale-in policy
  - ➔ This policy will delete the oldest created virtual machine in the scale set, after balancing VMs across availability zones (for zonal deployments).

➔ Enabling this policy requires a configuration change on the virtual machine scale set model.

Event	Instance IDs in Zone1	Instance IDs in Zone2	Instance IDs in Zone3
Initial	3, 4, 5, 10	2, 6, 9, 11	1, 7, 8
Scale-in	3, 4, 5, 10	2, 6, 9, 11	1, 7, 8
Scale-in	3, 4, 5, 10	6, 9, 11	1, 7, 8
Scale-in	4, 5, 10	6, 9, 11	1, 7, 8
Scale-in	4, 5, 10	6, 9, 11	7, 8
Scale-in	5, 10	6, 9, 11	7, 8
Scale-in	5, 10	9, 11	7, 8

Source :https://docs.microsoft.com/en-us/azure/virtual-machine-scale-sets/virtual-machine-scale-sets-scale-in-policy