Cloud Computing

Caltech | Center for Technology & Management Education

Designing Infrastructure Solutions on Azure

Powerd by simplilearn

Design for High Availability

Learning Objectives

By the end of this lesson, you will be able to:

- Classify the architectural best practices for reliability into categories
- Recommend a solution for recovery in different regions
- Recommend a solution for Azure backup management
- Design a solution for data archiving and retention

A Day in the Life of an Azure Architect

You are working as a cloud architect for a company that offers an OTT platform. You need to design a solution that could help the company with the following use cases:

- It can serve millions of customers at the same time across the globe.
- It can automatically scale your application when resource demand varies.
- You need to have multiple deployments of an application in different regions.
- You need to ensure that customers get the content served depending on the location.

To achieve all of the above, along with some additional features, we would be learning a few concepts in this lesson that will help you find a solution for the above scenario.

Identify the Availability Requirements for Azure Resources

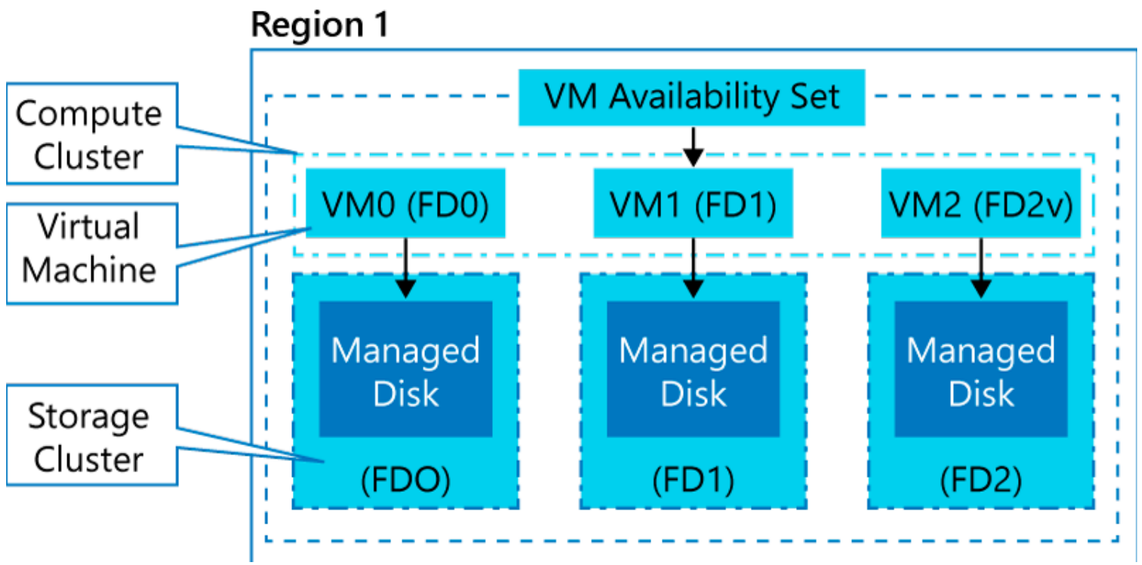
Delivering Reliability in Azure

The improvement of reliability requires creating solutions that continue to function even when they fail.

Resilient foundation	Foundation is the Microsoft investment in the platform, including availability zones.
Resilient services	Services support high availability, such as zone-redundant storage (ZRS).
Resilient applications	Applications should be architected to support resiliency.

Availability Sets

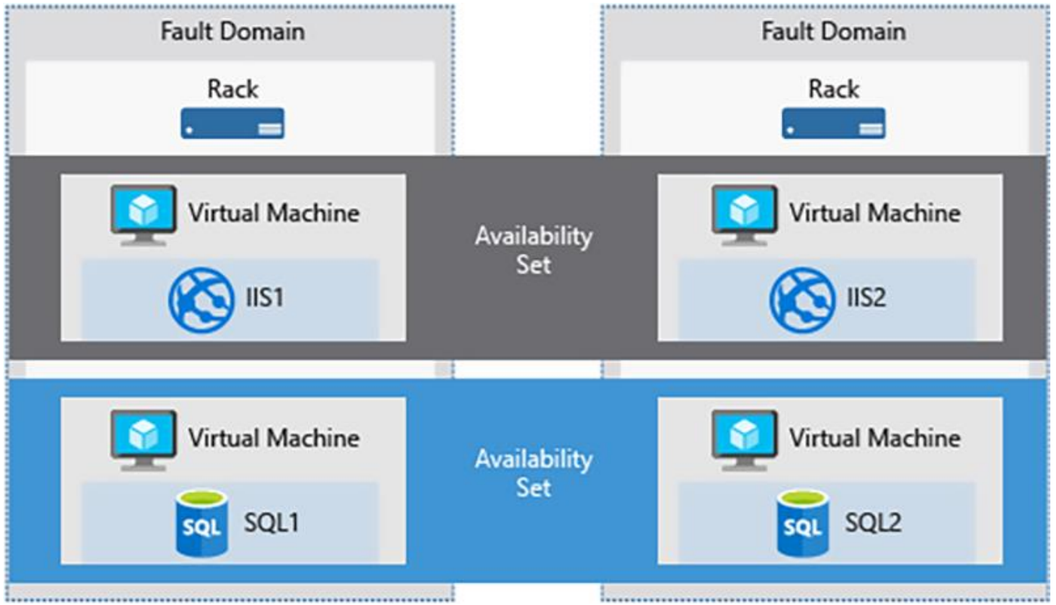
An availability set is a logical grouping of VMs within a data center.



- On compute clusters, VMs are automatically distributed across fault and update domains.
- On storage clusters, it provides equal resiliency with managed disks.

It provides a service level agreement (SLA) of 99.95% availability (for 2 or more VMs).

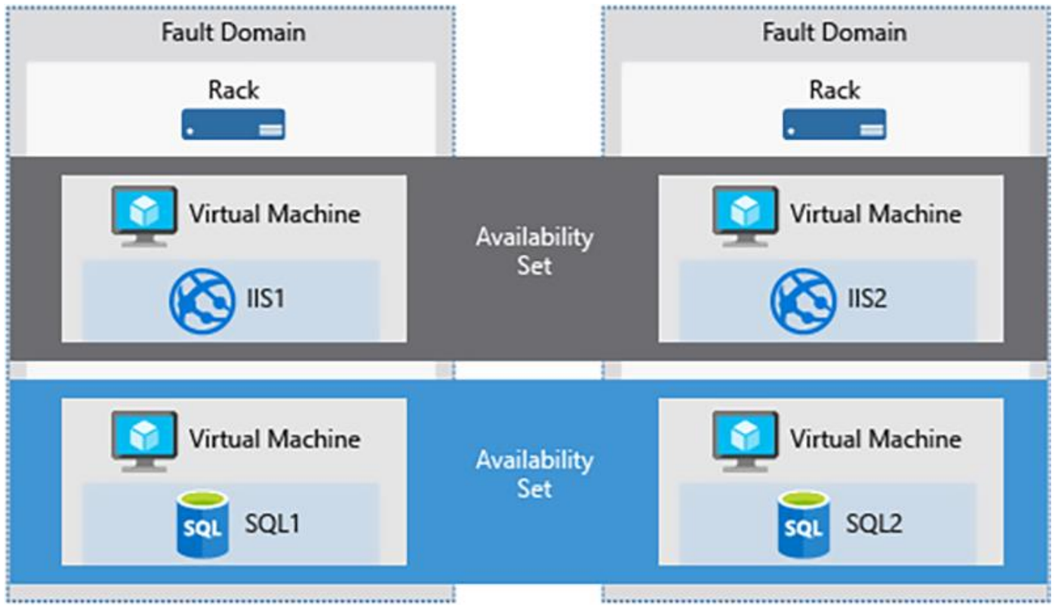
Update and Fault Domains



Update domains

- Azure can make incremental or rolling updates across a deployment using update domains.
- Only one update domain is rebooted at a time during scheduled maintenance.
- There are five update domains by default, but you can add up to 20 more.

Update and Fault Domains



Fault domains

- These are a set of virtual machines that share a single point of failure and a similar set of hardware and switches.
- At least two fault domains are allocated to each VM in an availability set.
- Up to three fault domains can be configured.

Service Level Agreements

SLA 01	SLA 02	SLA 03
99.99% Uptime guarantee For all virtual machines with two or more instances deployed in two or more Availability Zones	**99.95%** Uptime guarantee For all virtual machines with two or more instances deployed in the same Availability Set	**99.90%** Uptime guarantee For any single instance virtual machine that uses premium storage for all operating system disks and data disks

Note

A VM can only be added to an availability set when it is created.

High Availability and Azure SQL Database

There are two types of high availability models used in the Azure SQL database:

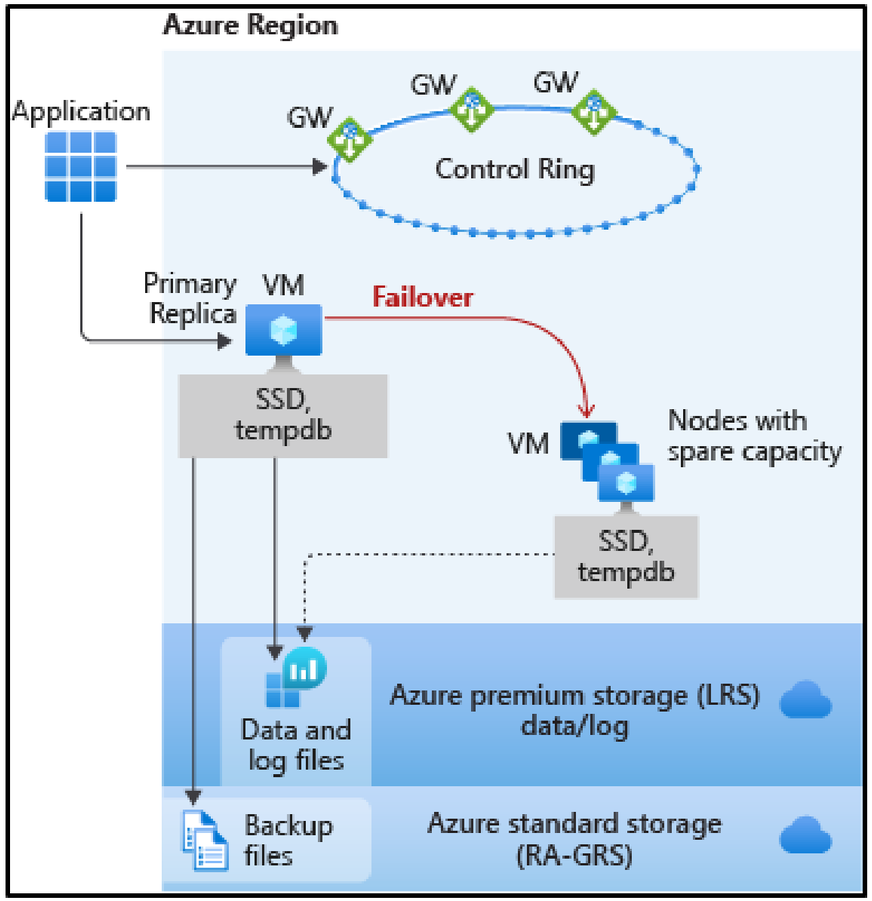
Standard Availability Model

Premium Availability Model

The standard availability model is based on a separation of compute and storage, whereas the premium availability model is based on a cluster of database engine processes.

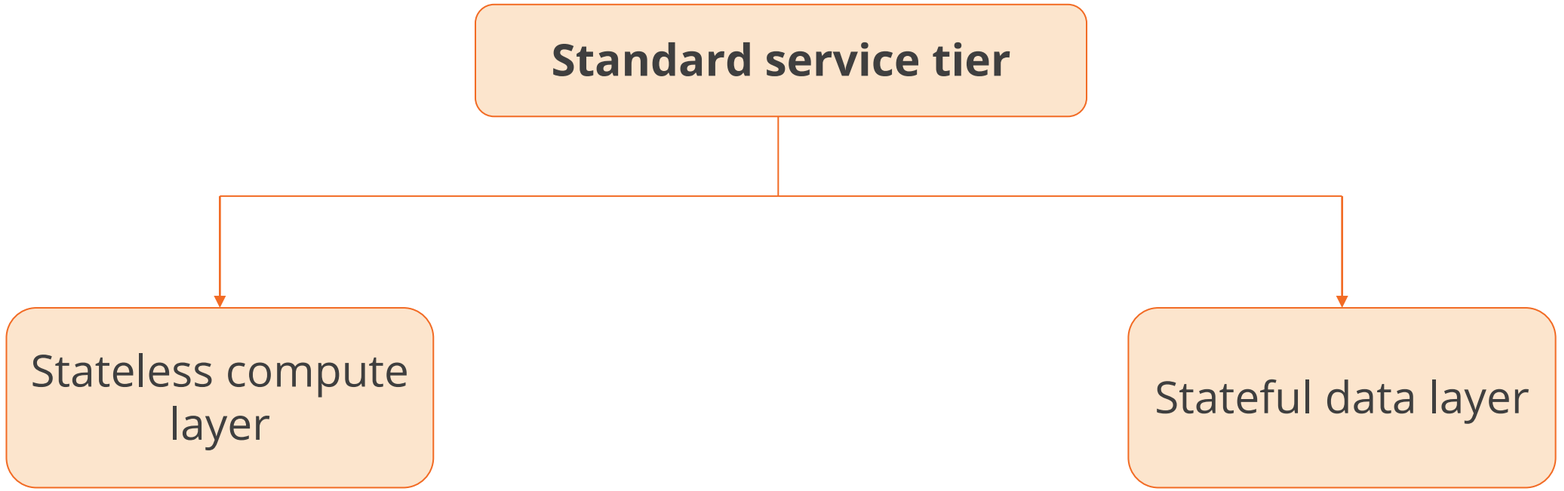
Standard Availability Model

The following figure shows four nodes with separate compute and storage layers:



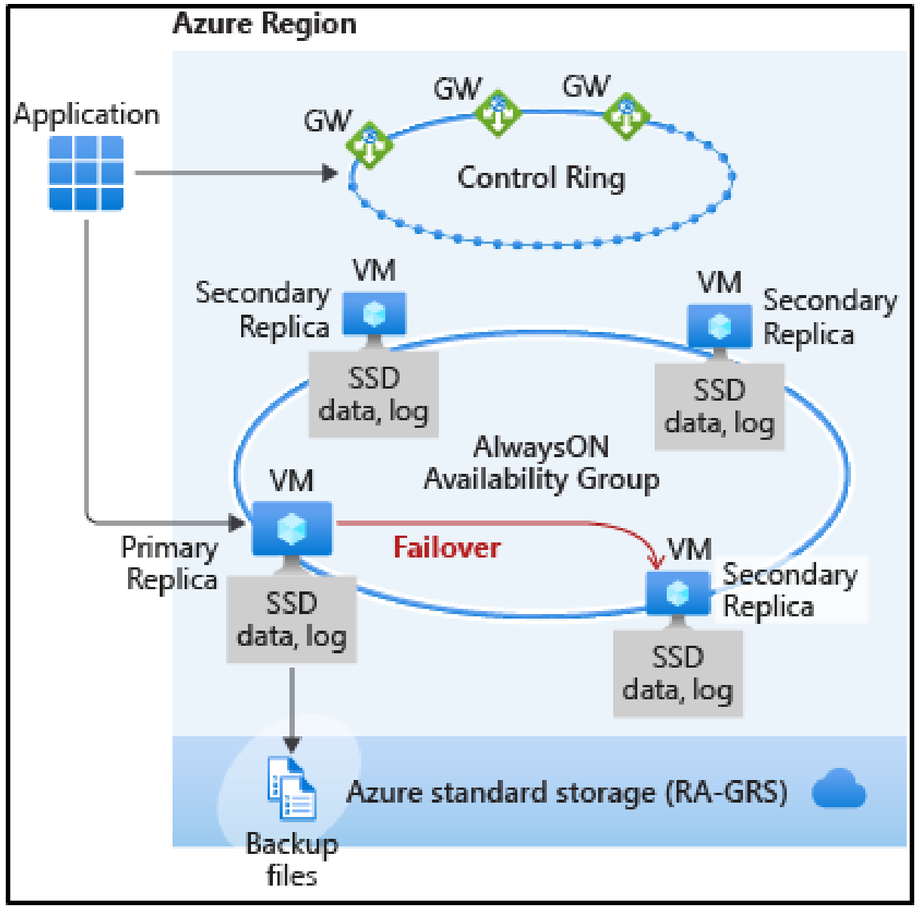
Standard Availability Model

The standard and general purpose service tier availability include two layers:



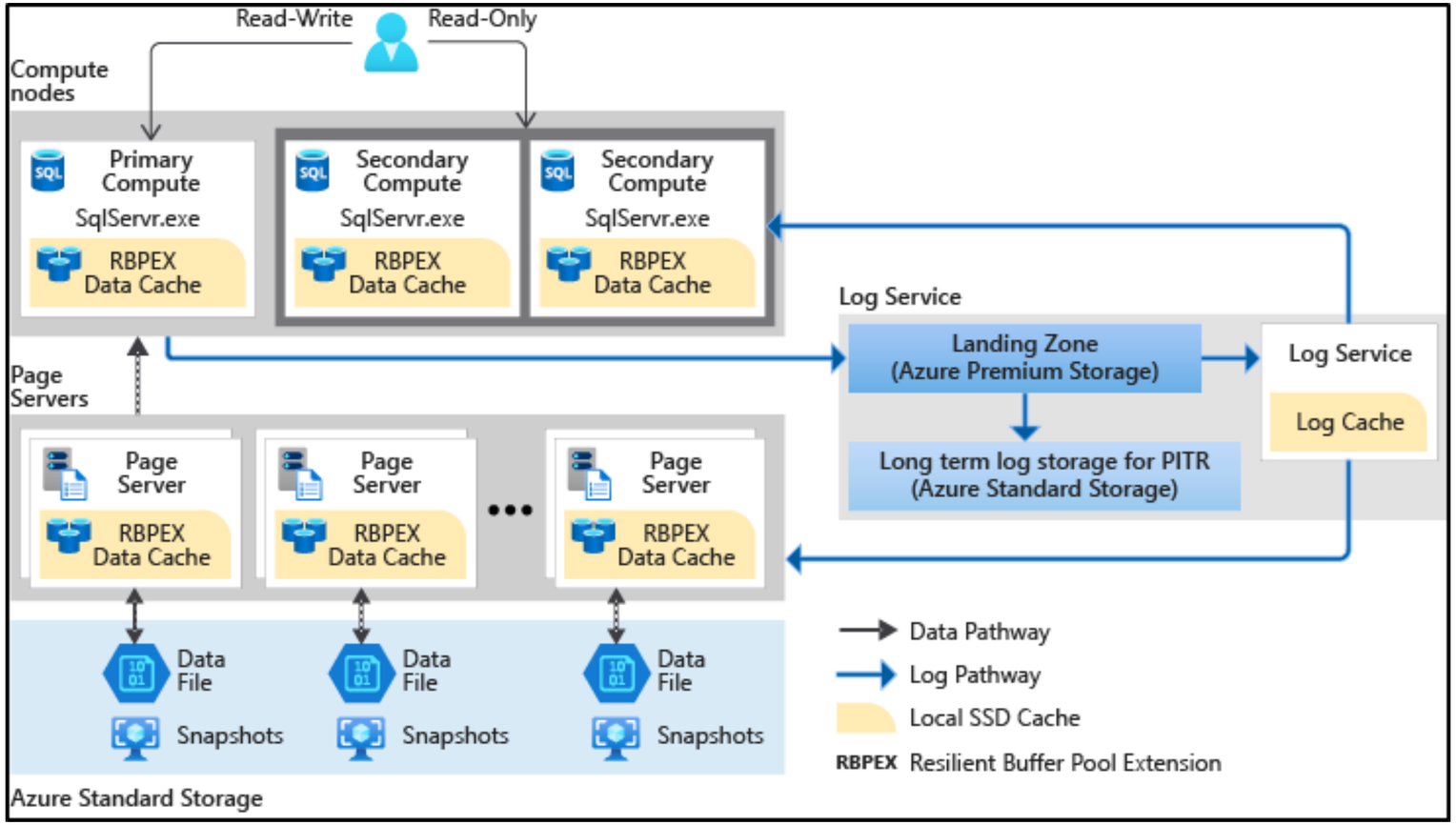
Premium and Business Critical Service Tier Availability

The premium model of high availability is shown below:



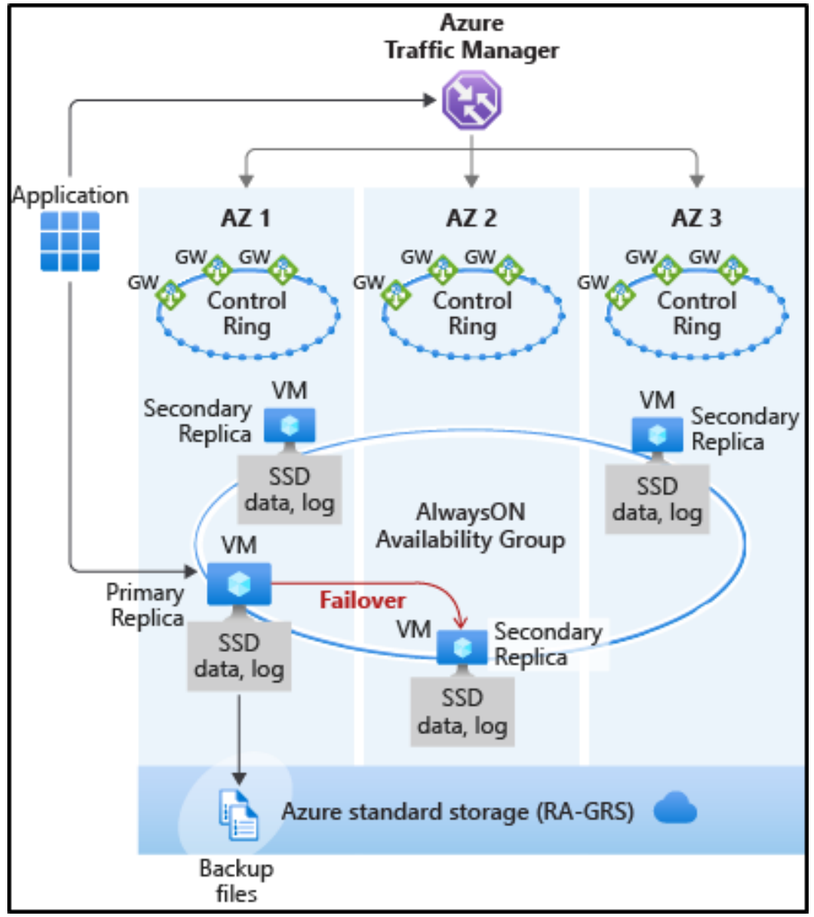
Hyperscale Service Tier Availability

The diagram given below will help users to understand the hyperscale service tier:



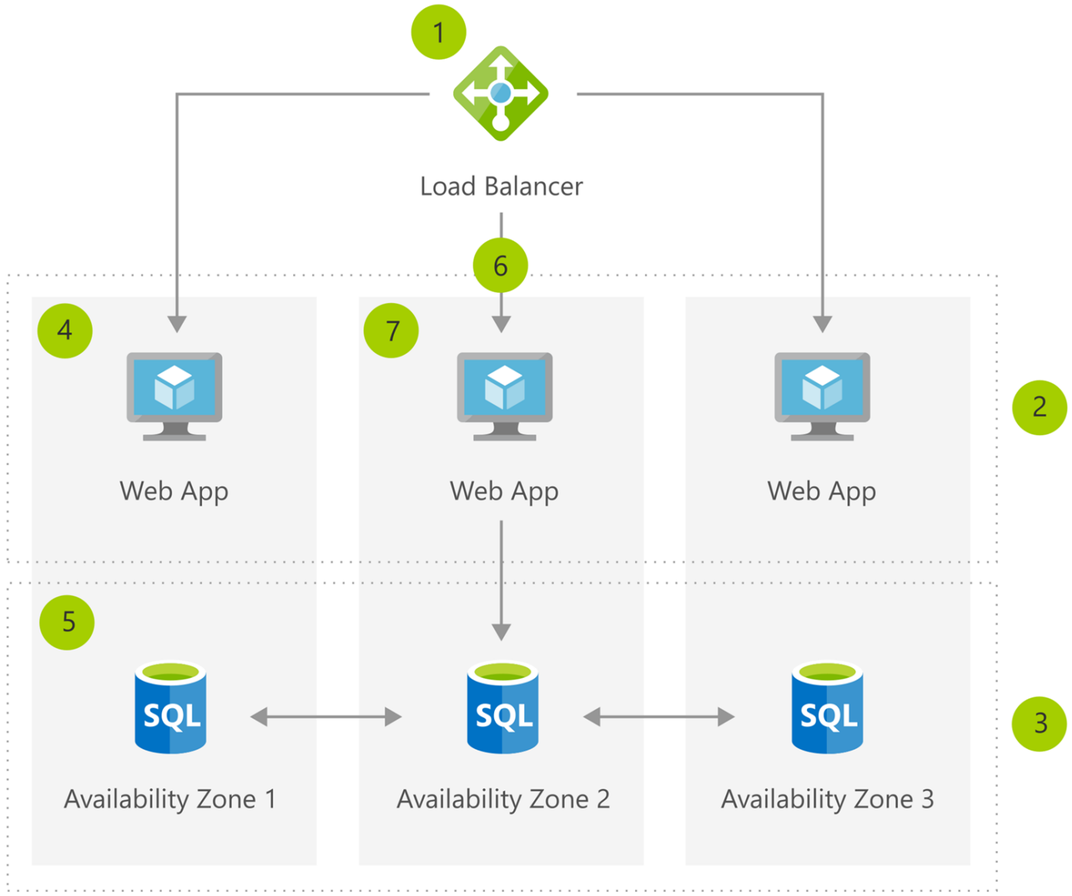
Zone Redundant Configuration

The following figure shows the zone redundant version of high availability:



High Availability for Business Continuity and Disaster Recovery

High availability for business continuity and disaster recovery can be achieved by:



- Creating a zone-redundant load balancer
- Creating front-end subnet
- Creating DB subnet
- Creating VMs in three availability zones
- Configuring zone-redundant SQL DB
- Adding VMs to the load balancer's backend pool
- Deploying the application on VMs for redundancy and high availability

Assisted Practice

Virtual Machine Availability set

Duration: 10 Min.

Problem Statement:

You've been requested to offer a method for achieving high availability for a virtual machine-hosted application. To do so, create an availability set, which is a logical grouping of VMs that helps Azure to understand how your application is designed to offer redundancy and availability.

Assisted Practice: Guidelines

Steps to create a virtual machine availability set are:

1. Login to your Azure portal
2. Search for virtual machines
3. Configure availability set while creating a VM

Assisted Practice

Virtual Machine Availability Zone

Duration: 10 Min.

Problem Statement:

You have been asked to showcase the availability zone option available on your virtual machine as an Azure Architect to increase the level of control you have over the availability of the applications and data on your VMs.

Assisted Practice: Guidelines

Steps to create a virtual machine availability zone are:

1. Login to your Azure portal
2. Search and select virtual machines
3. Select availability zone
4. Configure the availability zone while creating a VM

Azure Front Door

Azure Front Door

Azure Front Door is a global, scalable entry-point that uses the Microsoft global edge network to create fast, secure, and highly scalable web applications.

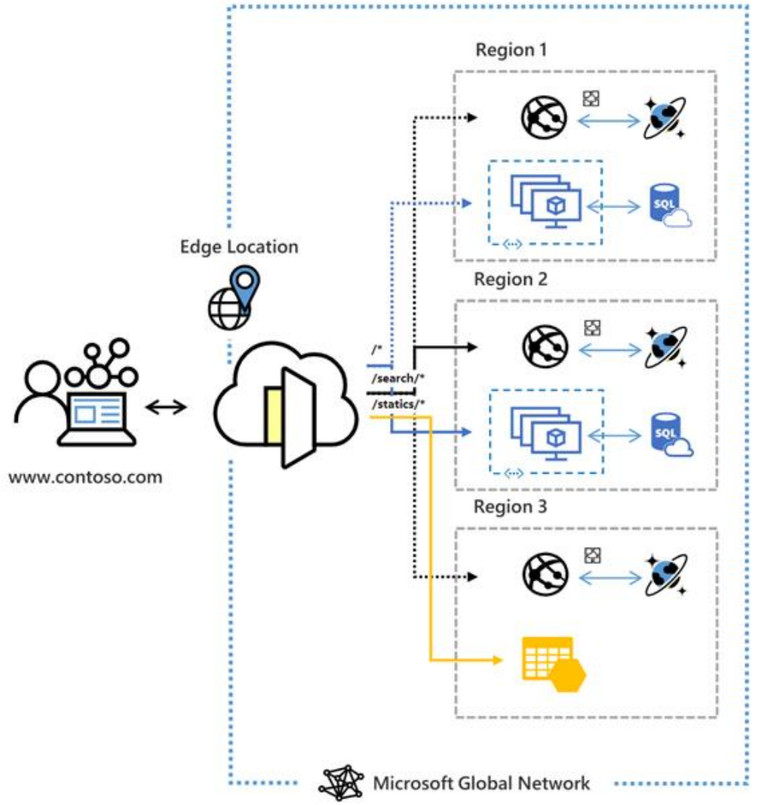
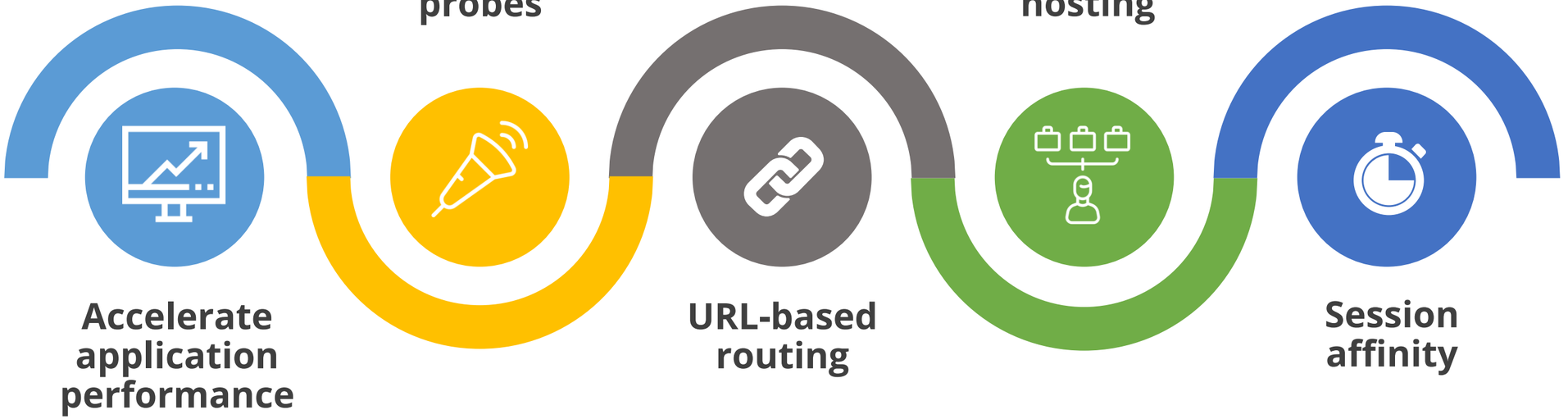


image source: https://docs.microsoft.com/en-in/

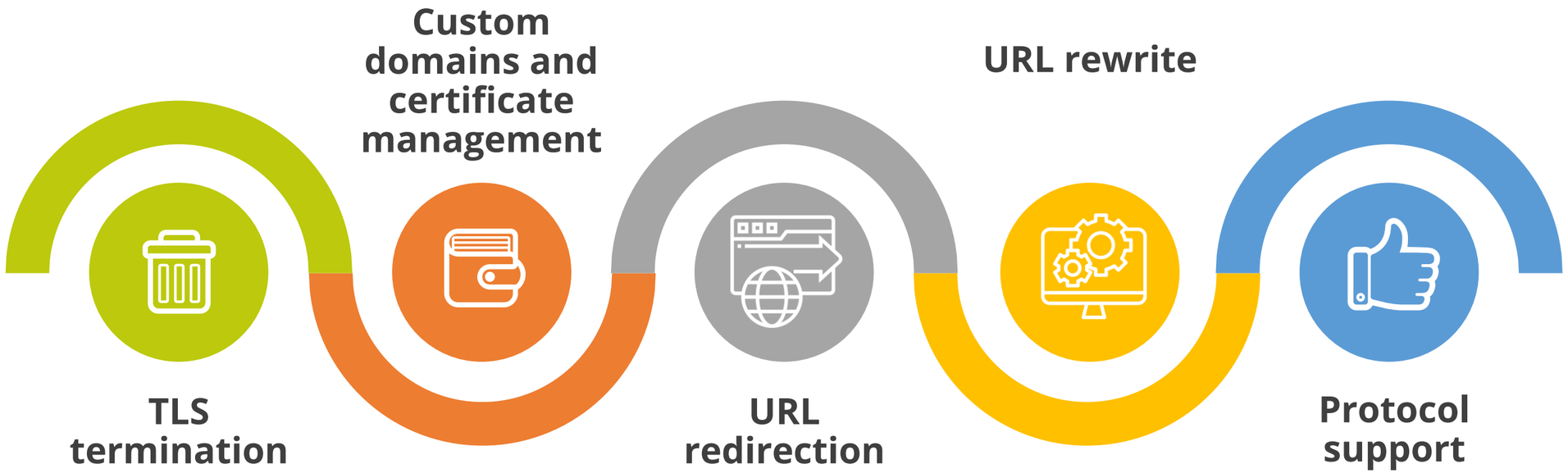
Benefits of Azure Front Door

It allows the user to identify, manage, and track the global route for web traffic by optimizing for best client-side performance and instant global failover for high availability.

Azure Front Door Features



Azure Front Door Features



Assisted Practice

Azure Front Door **Duration: 10 Min.**

Problem Statement:

You've been requested to suggest a global, scalable entry-point that uses the Microsoft global edge network to create fast, secure, and widely scalable web applications.

Assisted Practice: Guidelines

Steps to create a Azure Front Door service are:

1. Configuring Azure Front Door
2. Configuring Frontend and Backend
3. Adding a routing rule

Azure Traffic Manager

Azure Traffic Manager

Traffic Manager is a network service that is used to route users to web app endpoints (deployments) in potentially different data centers located around the world.

Benefits

It ensures our apps are:

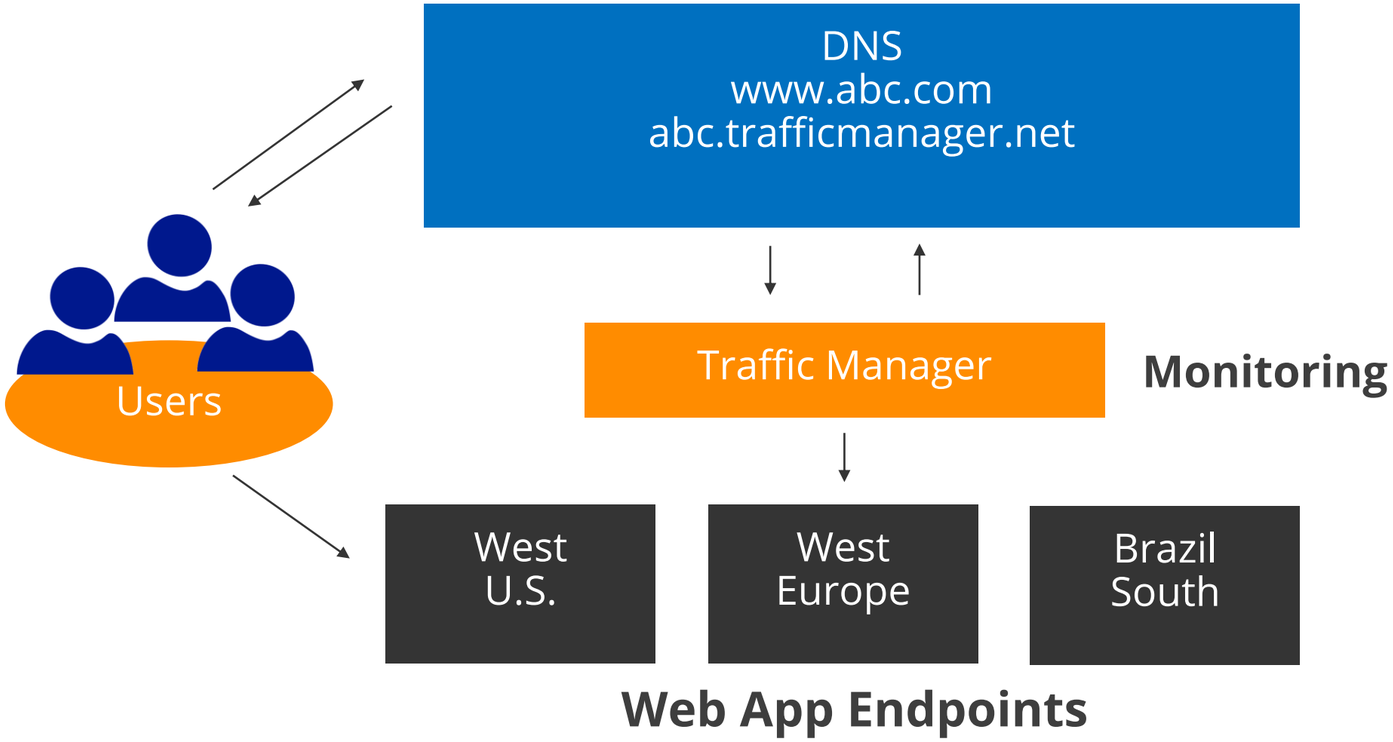
- Highly scalable
- Highly available

Prerequisites

One should have two or more deployments of a web app to leverage the features of Azure Traffic Manager.

Traffic Manager Overview

The below diagram shows the overview of a traffic manager:



Traffic Manager Features

Increase application availability

Improve application performance

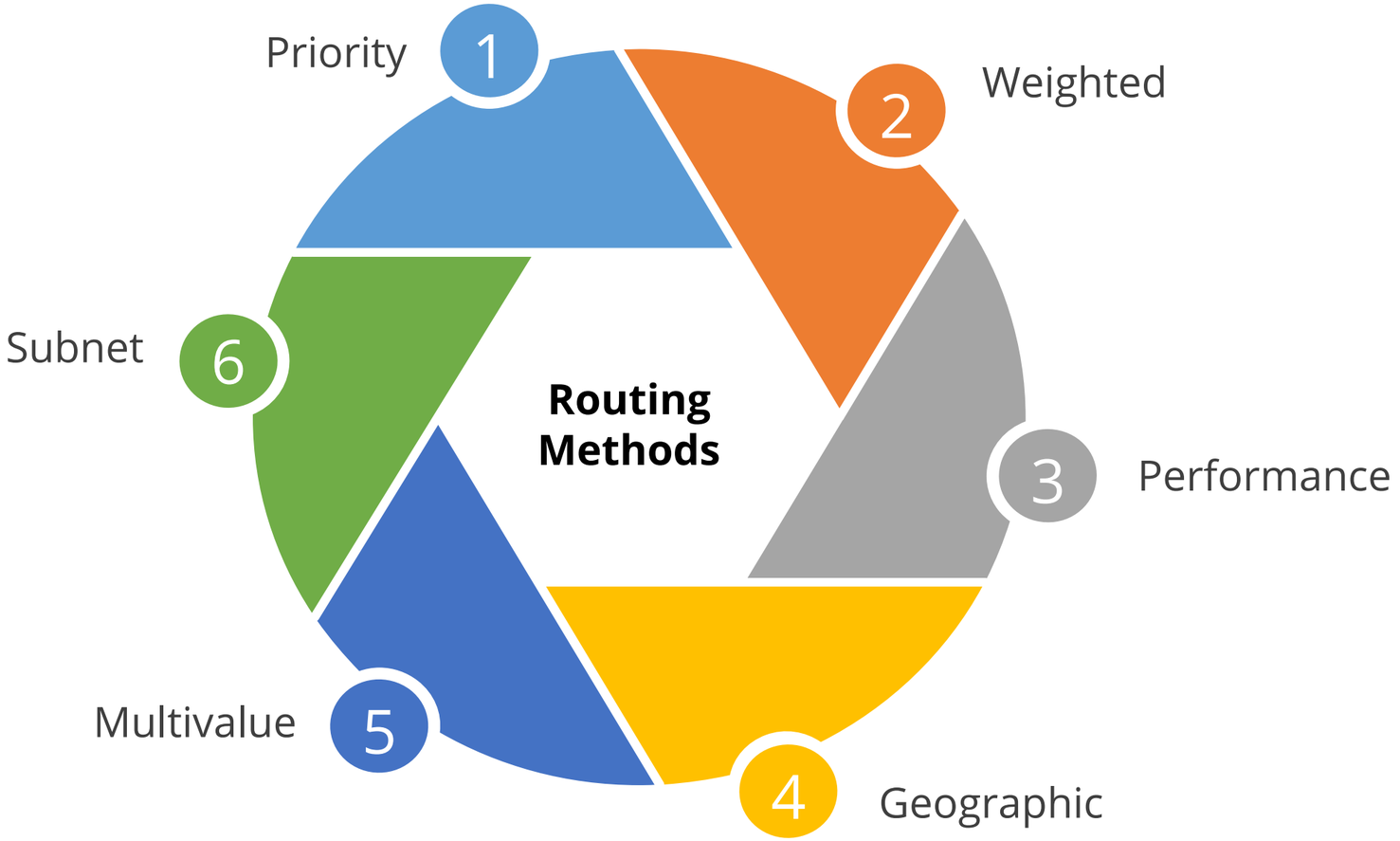
Service maintenance without downtime

Distribute traffic for complex deployments

Combine hybrid applications

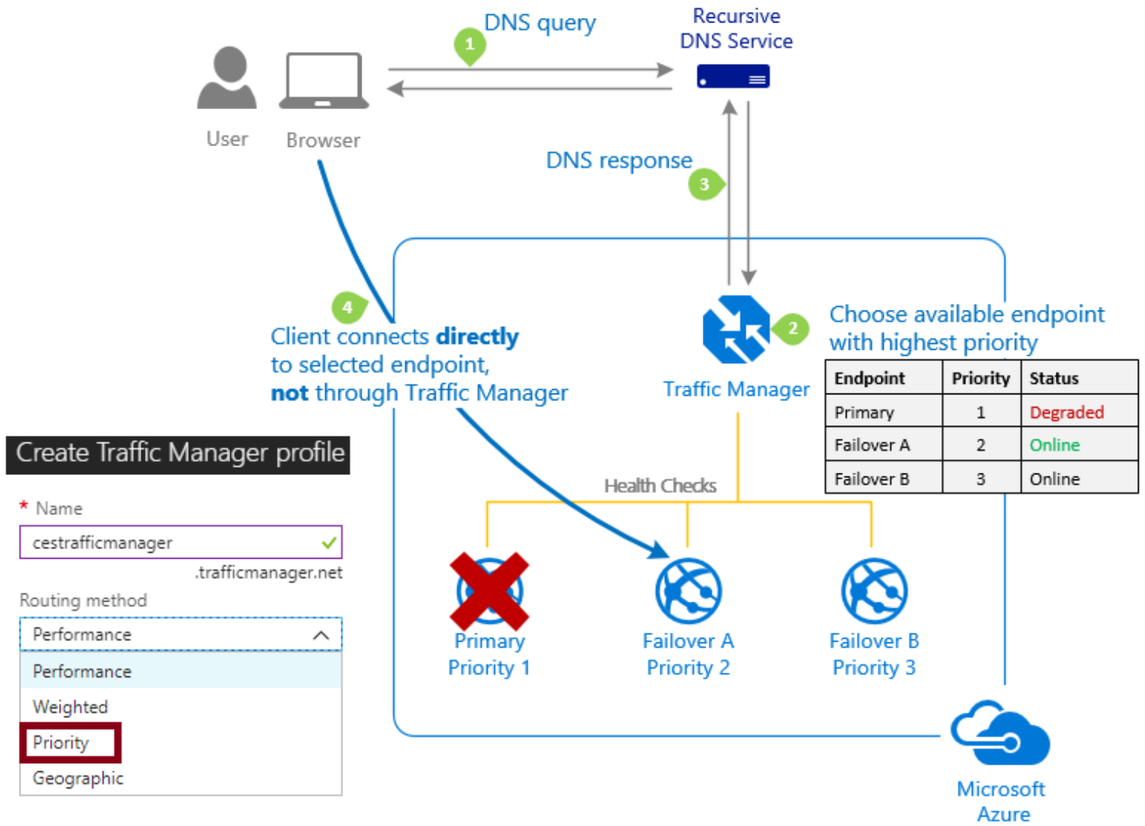
Traffic Manager Routing Methods

These are the features of routing methods:



Priority Routing

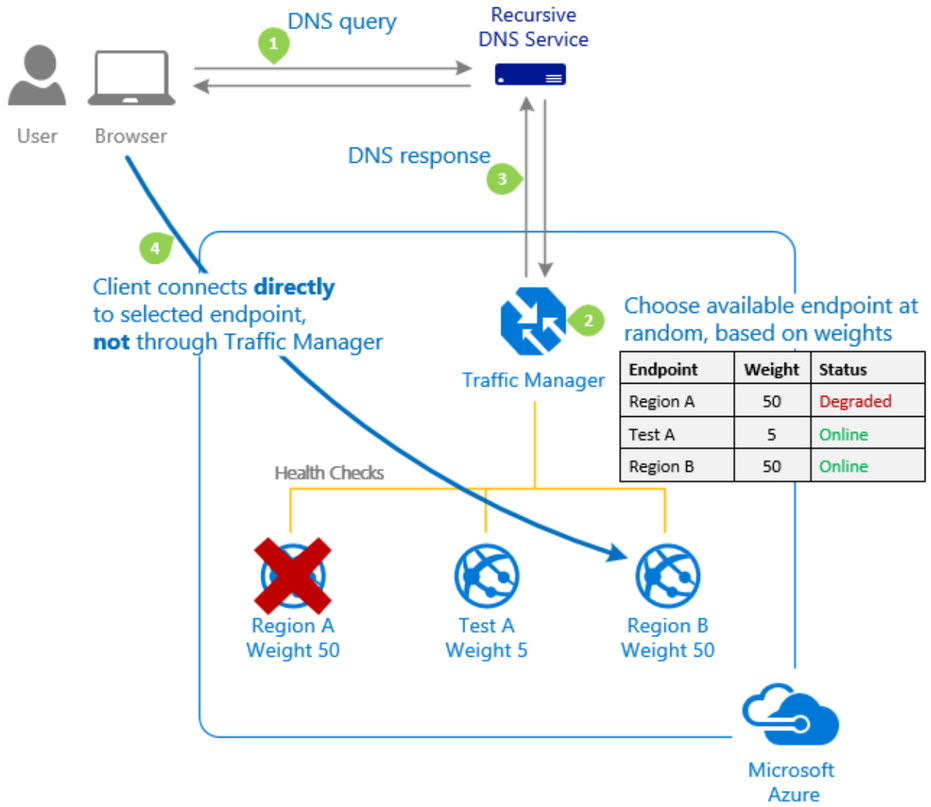
The Priority traffic-routing method allows Azure customers to implement this failover pattern:



Source: https://docs.microsoft.com/

Weighted Routing

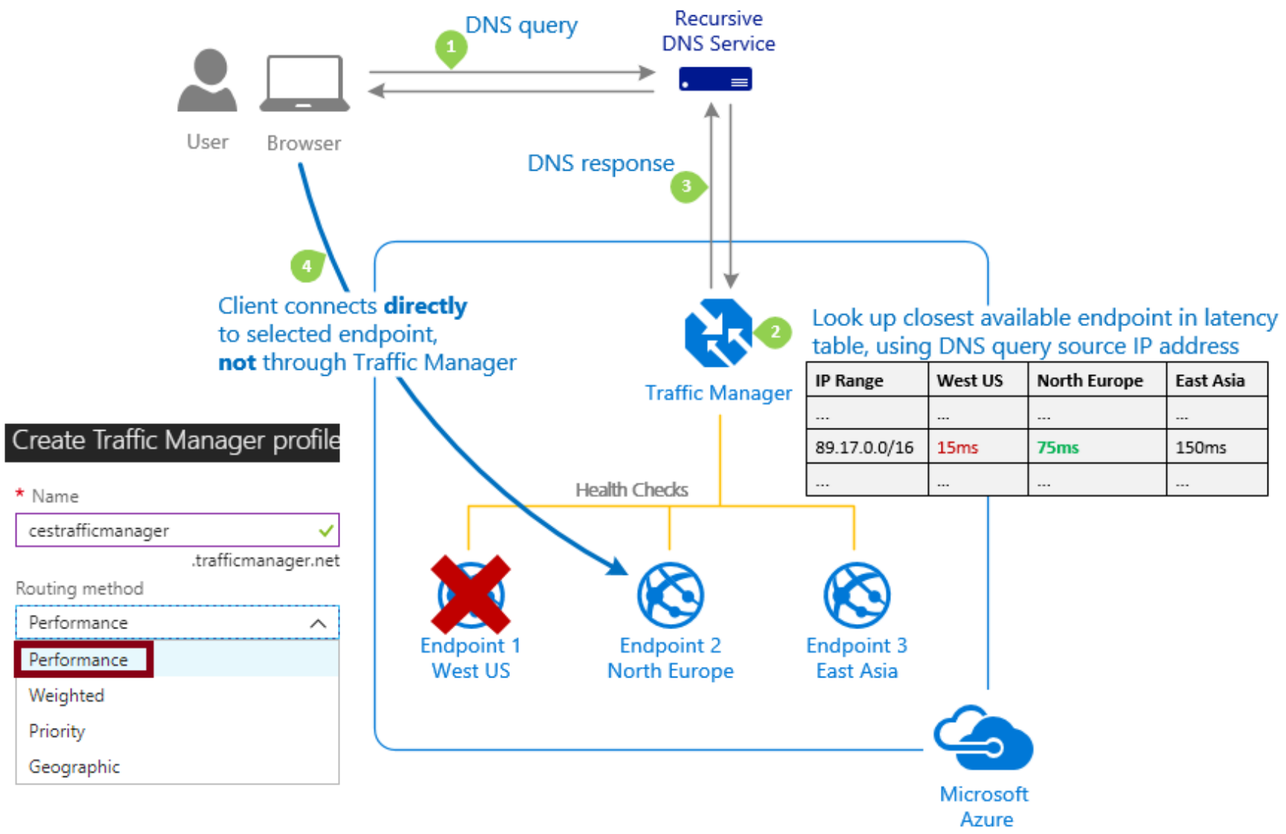
The Weighted traffic-routing method distributes traffic evenly or uses a predefined weighting.



Source: https://docs.microsoft.com/

Performance Routing

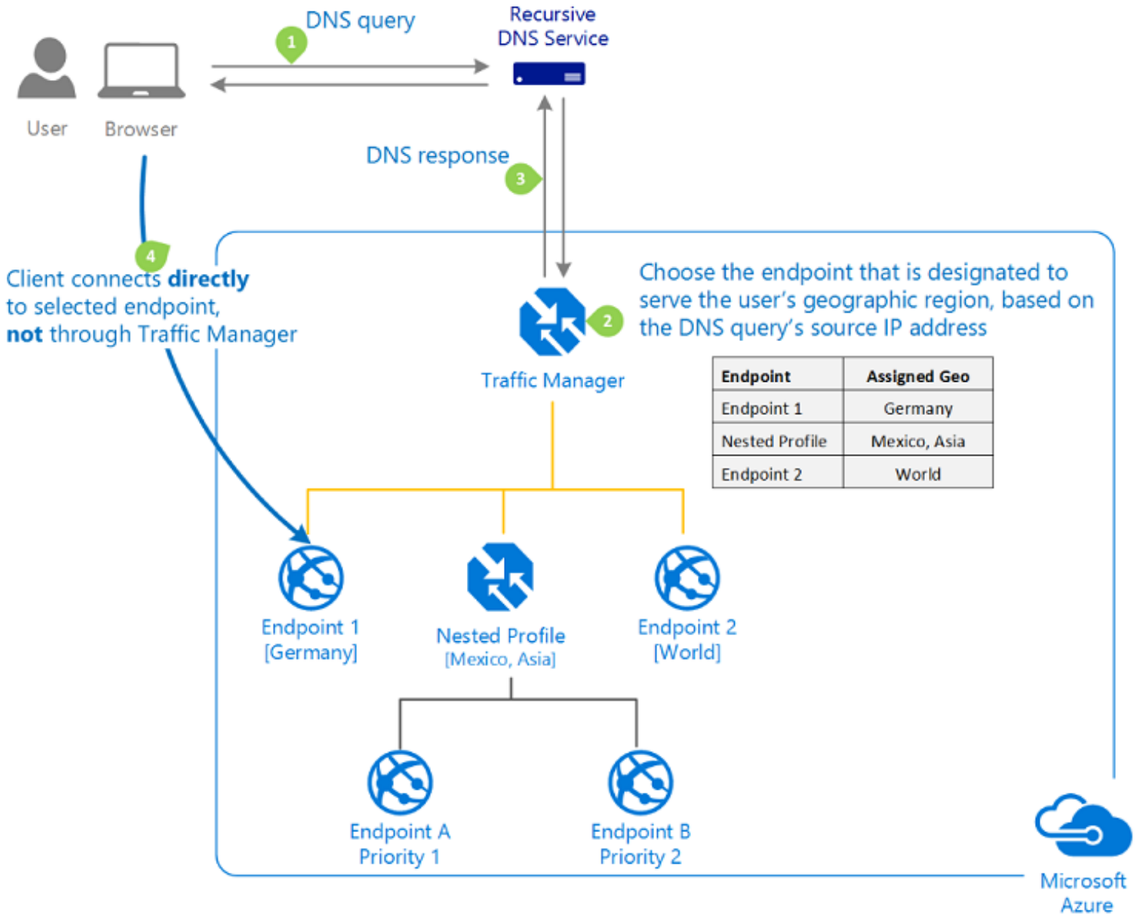
The Performance traffic-routing method helps users route traffic to their nearest location.



Source: https://docs.microsoft.com/

Geographic Routing

The Geographic traffic-routing method helps in measuring traffic from different regions.



Endpoint	Assigned Geo
Endpoint 1	Germany
Nested Profile	Mexico, Asia
Endpoint 2	World

Source: https://docs.microsoft.com/

Points to Remember

- All Azure Traffic Manager profiles use the shared domain ***.trafficmanager.net***.
- Web apps in the same data centers will automatically be load balanced.
- Web apps spread across multiple data centers will need to be load balanced.
- Endpoints need to be in the same subscription to provide load balancing between the apps.

Distributing Network Traffic

The network traffic table is explained below:

Service	Azure load balancer	Application gateway	Traffic manager	Azure front door
Technology	Transport layer (Level 4)	Transport layer (Layer 7)	DNS resolver	Layer 7 or HTTP/HTTPS
Protocols	Any TCP or UDP protocol	HTTP, HTTPS, HTTP/2, and Web sockets	DNS resolution	Split TCP-based anycast protocol
Backends and Endpoints	Azure VMs and Azure VM scale sets	Azure VMs, Azure VM scale sets, Azure app services, IP addresses, and hostnames	Azure cloud services, Azure app services, Azure app services slots, and public IP addresses	Internet-facing services hosted inside or outside of Azure
Network connectivity	External and Internal	External and Internal	External	External and Internal

Assisted Practice

Azure Traffic Manager

Duration: 25 Min.

Problem Statement:

Sim-Edu Corporation wants to use Azure Traffic Manager and wants to see its working functionality by checking how high availability works with Azure Traffic Manager. You've been assigned the task of implementing Traffic Manager, which will route client requests to the proper service endpoint using a traffic-routing method.

Assisted Practice: Guidelines

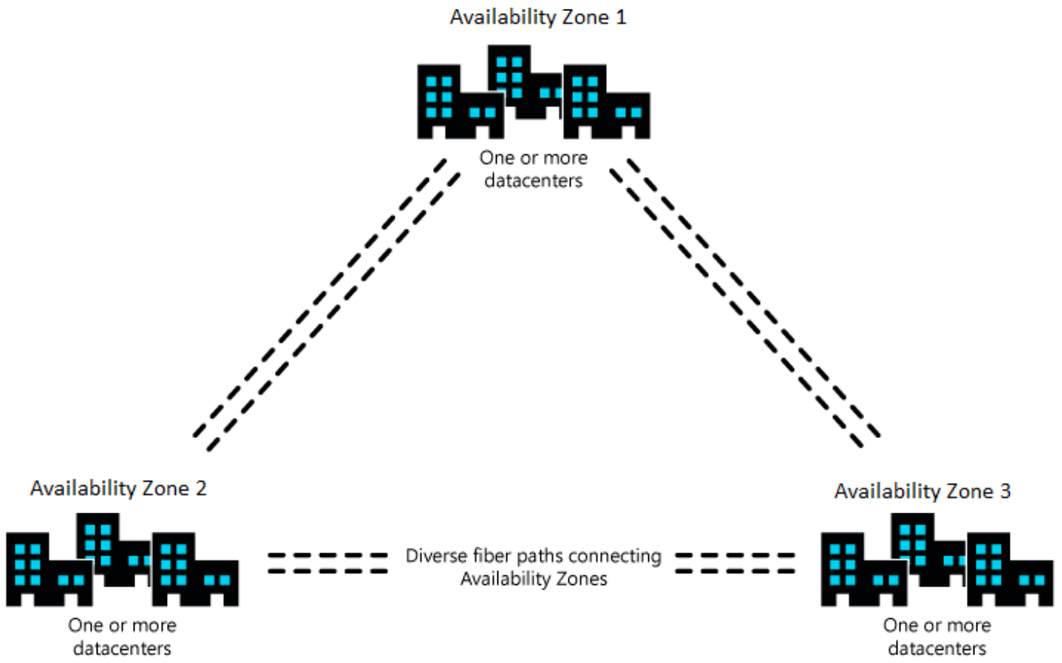
Steps to implement the traffic manager:

1. Creating virtual machines
2. Adding DNS names to virtual machines
3. Creating Traffic Manager
4. Testing Traffic Manager

Recommend a High Availability Solution for Compute

Building Solutions for High Availability Using Availability Zones

Availability Zones are physical locations inside an Azure region that are distinct from one another.

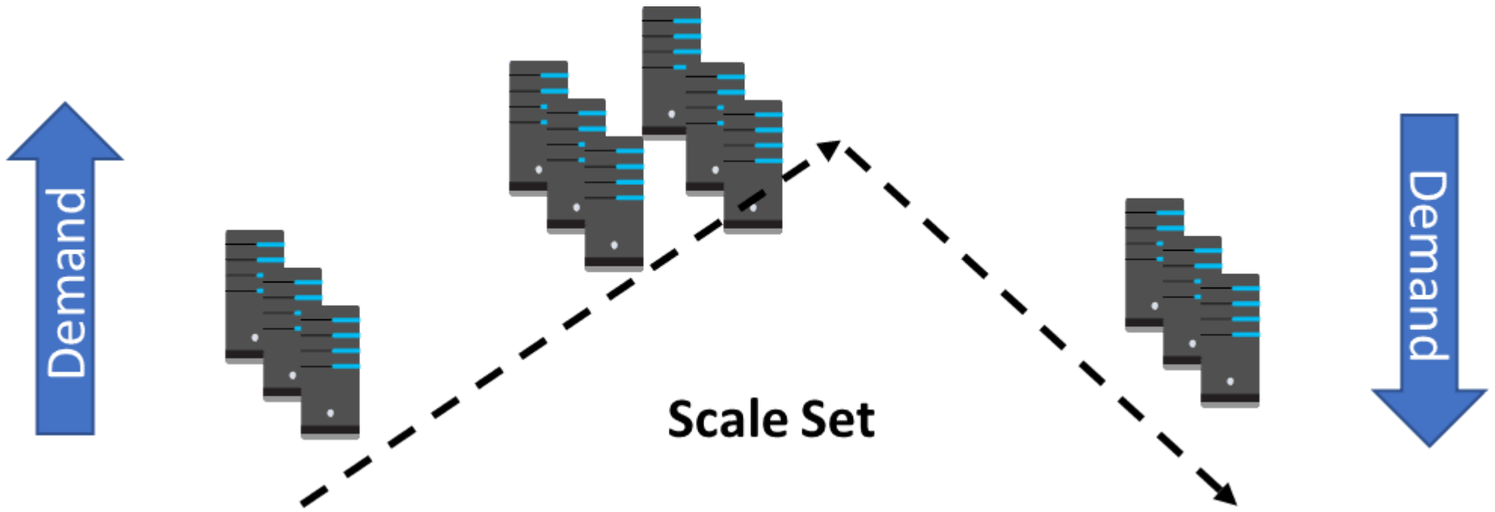


- A zone is made up of one or more data centers with independent power, cooling, and networking.
- The physical separation of availability zones within a region limits the impact on applications.

Zones are designed to support services, capacity, and availability from other zones in the region.

Scale Sets

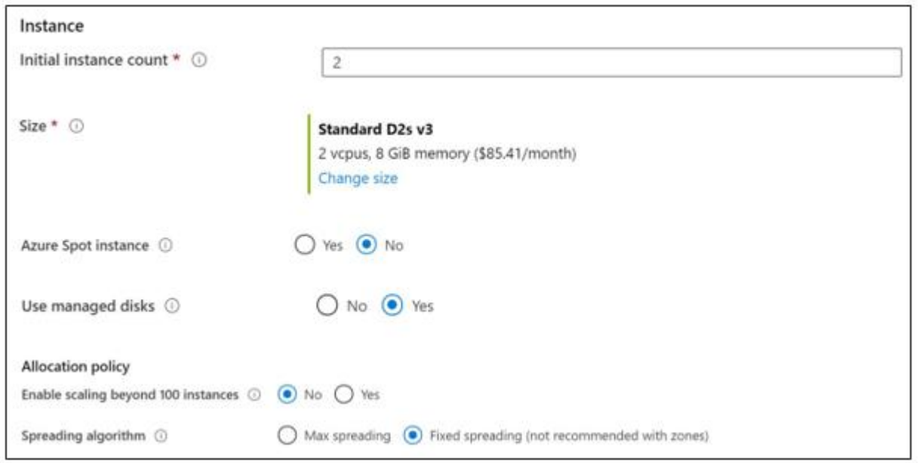
Azure virtual machine scale sets enable the user to deploy a set of identical VMs.



Scale Sets Benefits

- All VM instances are created from the same base OS image and configuration.
- Scale sets are used to run multiple instances of the user's application.
- Scale sets can automatically increase or decrease the number of VM instances to accommodate changing client demand (called Autoscale).
- Scale sets are highly scalable and can contain up to 1,000 VM instances. (600 when using custom images)

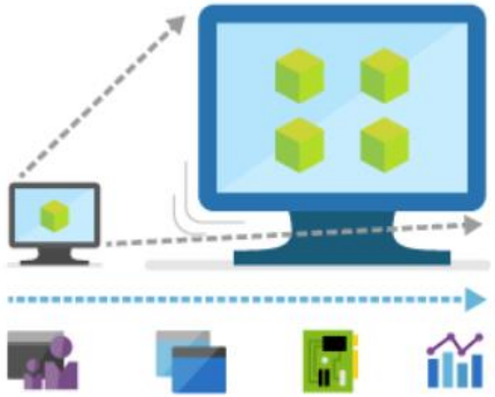
Implementing Scale Sets

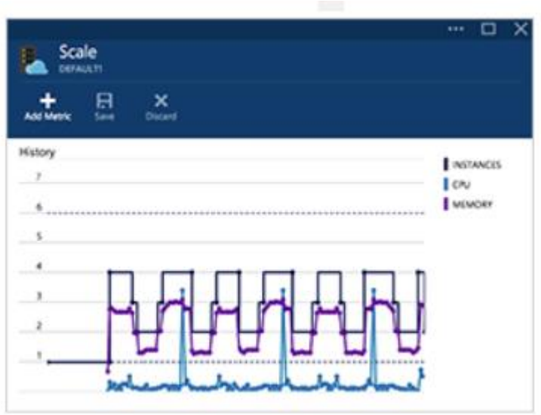


- Instance count: Number of VMs in the scale set (0 to 1000)
- Instance size: The size of each virtual machine in the scale set
- Azure spot instance: Can save up to 80%
- Use managed disks
- Enable scaling beyond 100 instances

Autoscale

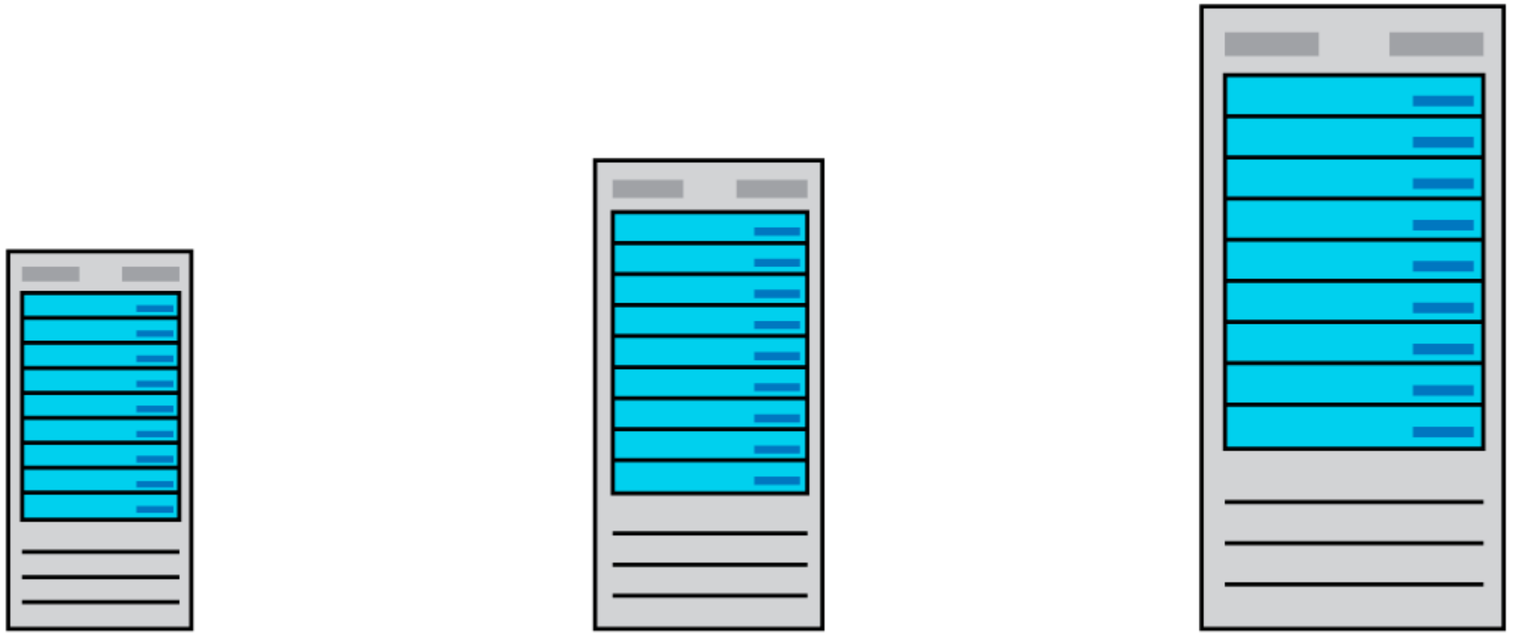
- Defines rules to adjust capacity automatically
- Schedules events to increase or decrease at a fixed time
- Reduces monitoring and optimizes performance





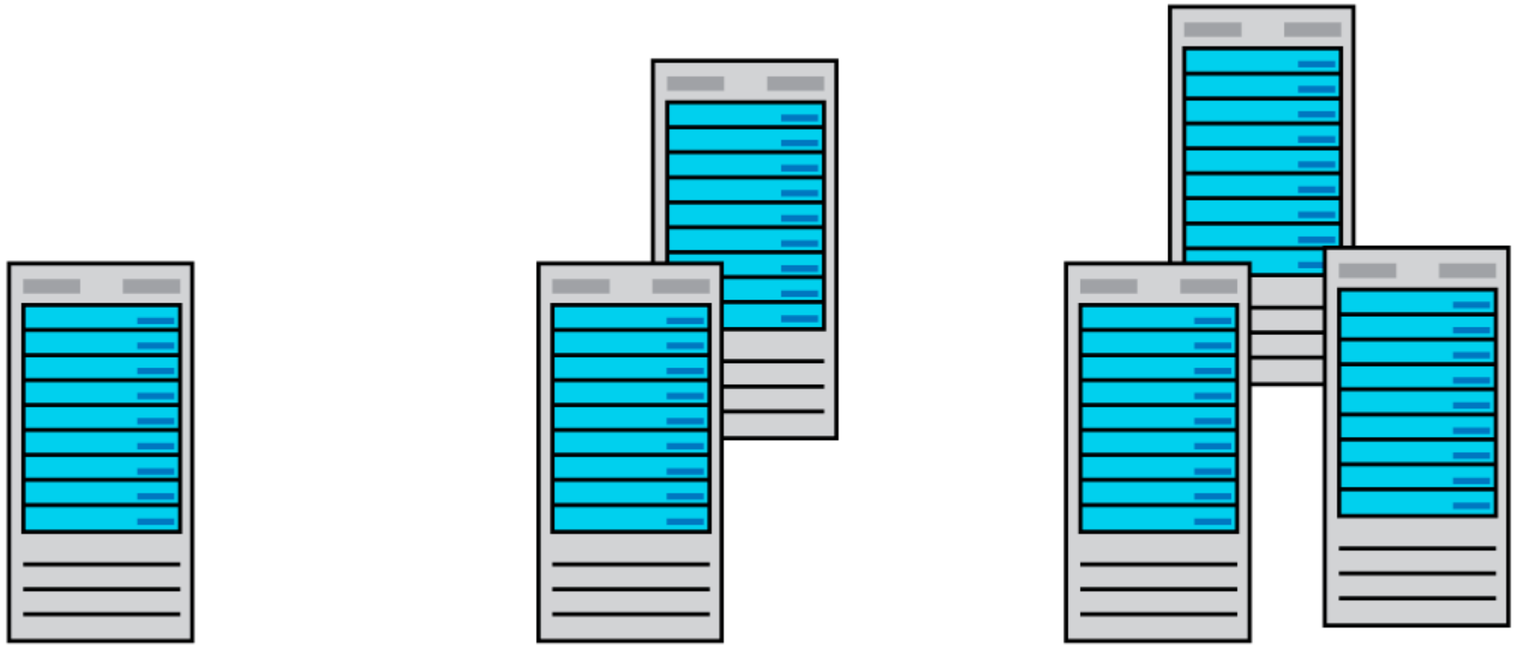
Vertical Scale

Vertical scaling is also known as scale-up and scale-down, which means increasing or decreasing virtual machine sizes in response to a workload.



Horizontal Scale

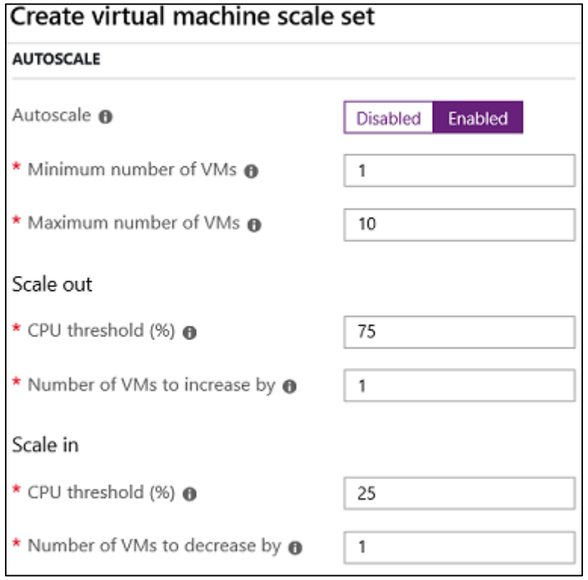
Scale-out (increase) the number of VMs in the set



Scale-in (reduce) the number of VMs in the set

Implementing Autoscale

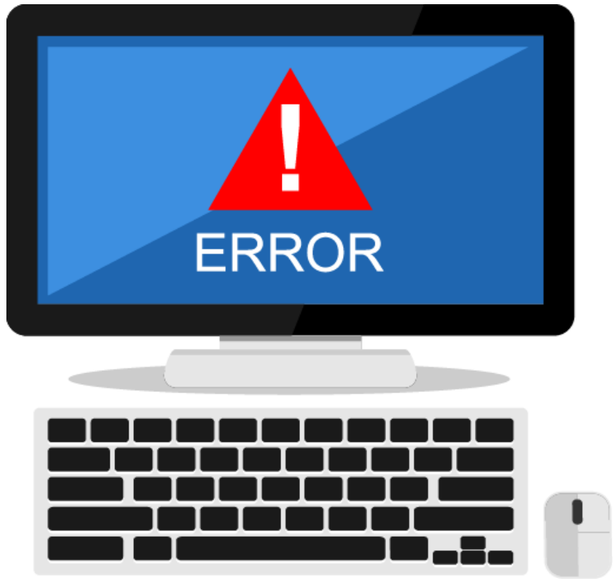
It defines the minimum, maximum, and default numbers of VM instances.



It creates more advanced scale sets with scale-out and scale-in parameters.

Storage Account Replication

Storage Account Replication stores multiple copies of user data to protect from planned and unplanned events, such as:

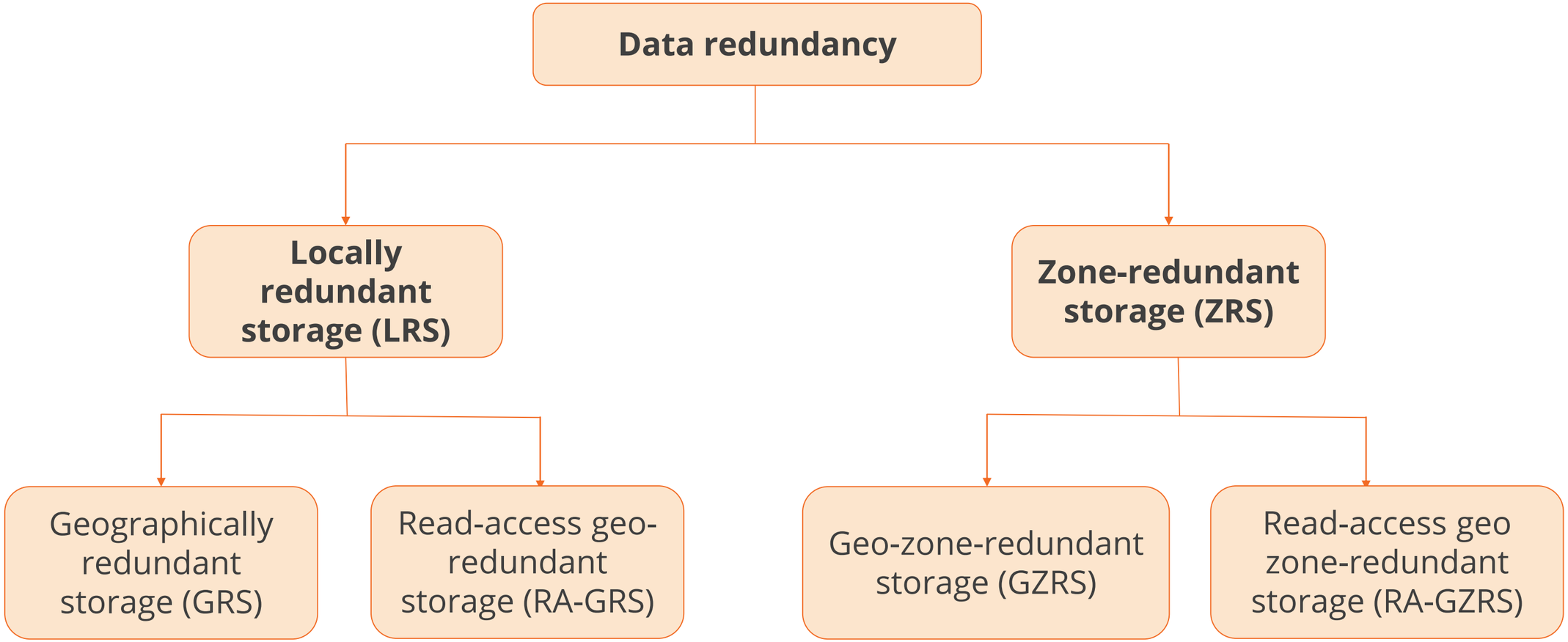


- Hardware failures
- Network or power outages
- Massive natural disasters

It ensures user storage accounts have their backup even in failure.

Storage Account Replication

The types of replication models are as follows:



Data in an Azure Storage account is always replicated three times in the primary region.

Assisted Practice

Virtual Machine Scale Set

Duration: 10 Min.

Problem Statement:

You were tasked with identifying different options for deploying and configuring Azure virtual machines and making it highly available using VM Scale set. You need to investigate compute and storage resiliency and scalability options that are available when using Azure virtual machine scale sets. You also want to explore the ability to automatically configure virtual machines and virtual machine scale sets by using the Azure Virtual Machine Custom Script extension.

Assisted Practice: Guidelines

Steps to create a virtual machine scale set are:

1. Register the Microsoft resource providers
2. Deploy zone-resilient Azure virtual machine scale sets by using the Azure portal
3. Configure Azure virtual machine scale sets by using virtual machine extensions

Recommend a High Availability Solution for Relational Data Storage

Hyperscale Availability Model

The availability model in Hyperscale includes four layers:

1

Compute Layer

A stateless compute layer executes sqlservr.exe and primarily contains transitory and cached data.

2

Storage Layer

Page servers form a stateless storage layer. It is a storage engine for the sqlservr.exe processes operating on the compute replicas

Hyperscale Availability Model

The availability model in Hyperscale includes four layers:

3

Transaction Log

The transaction log landing zone and transaction log long-term storage together make a stateful transaction log storage layer.

4

Data Storage

This layer makes use of Azure Storage data availability and redundancy characteristics.

Accelerated Database Recovery (ADR)

ADR is a SQL Server database engine feature that redesigns the SQL Server database engine recovery process to greatly improve database availability, particularly in the presence of long-running transactions.

Benefits of ADR

Fast and consistent database recovery

Long-running transactions have no impact on overall recovery time with ADR, allowing for quick and consistent database recovery regardless of the number of ongoing transactions or their magnitude.

Benefits of ADR

Instantaneous transaction rollback

Transaction rollback is instantaneous with ADR, regardless of how long the transaction has been active or how many modifications it has completed.

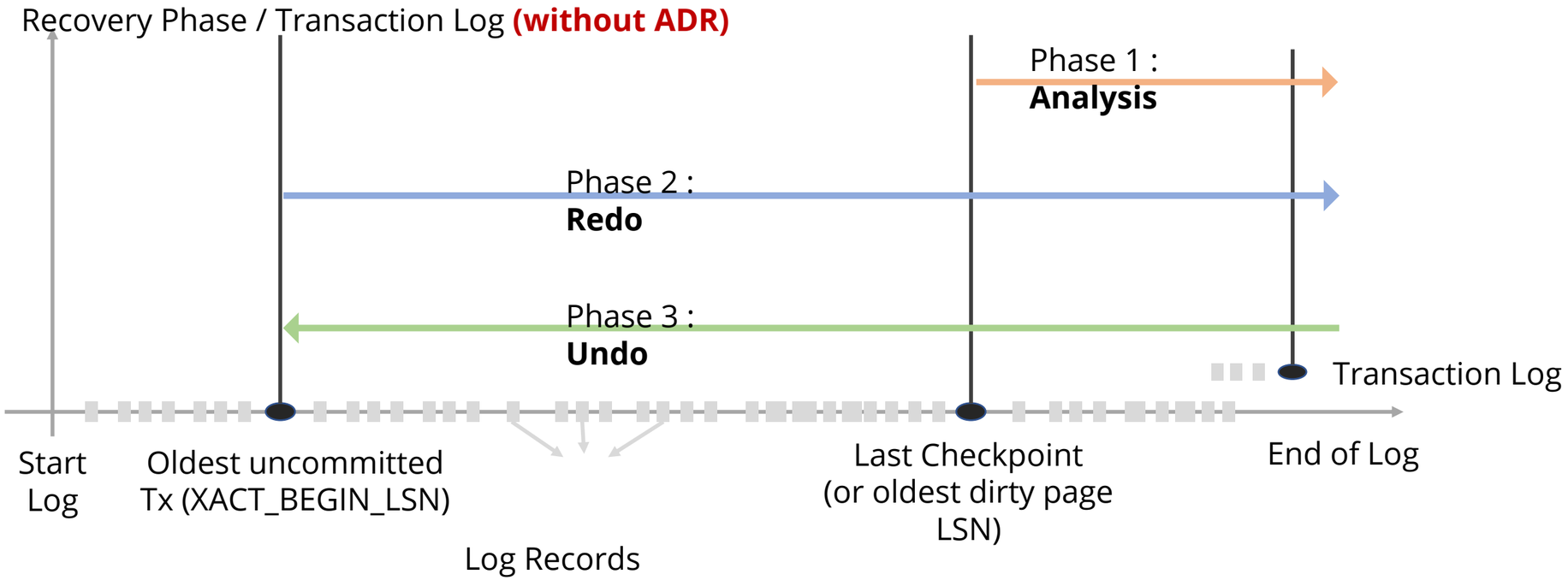
Benefits of ADR

Aggressive log truncation

Even in the presence of active long-running transactions, the transaction log is aggressively trimmed with ADR, preventing it from expanding out of control.

Standard Database Recovery Process

The ARIES recovery model is used to recover databases, and it consists of three steps, as shown in the image below .



Standard Database Recovery Process

Analysis phase

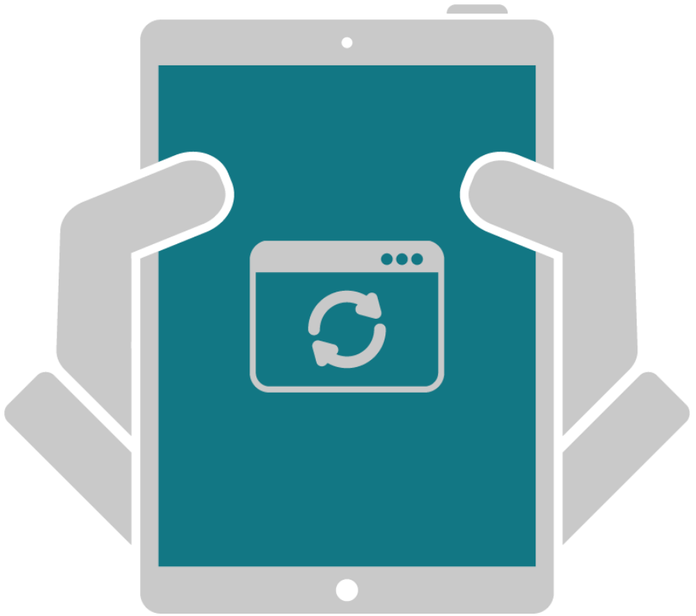
To identify the status of each transaction at the time the database was shut down

Standard Database Recovery Process

Redo phase

To restore the database to its state before the accident by undoing all committed operations

Standard Database Recovery Process

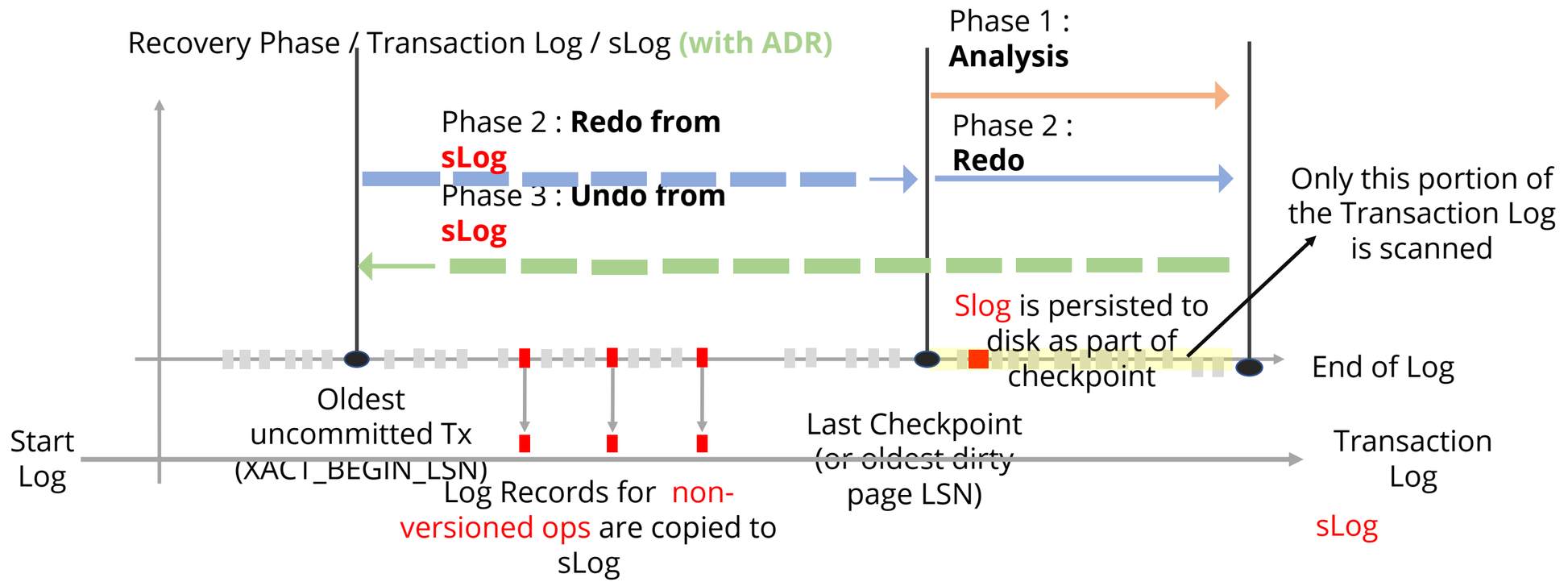


Undo phase

Traverses the log backward for each transaction that was active at the time of the crash, redoing the actions that this transaction executed

Accelerated Database Recovery Process

The three phases of the ADR recovery procedure are the same as those of the present recovery process.



Accelerated Database Recovery Process

Analysis phase

The technique is the same as before for non-versioned operations, with the addition of recreating sLog and copying log records.

Accelerated Database Recovery Process

Redo phase

It is divided into two parts:

Phase 1:

Redo from sLog (oldest uncommitted transaction up to the last checkpoint): As it just needs to process a few records from the sLog, redo is a quick operation.

Phase 2:

The redo from the Transaction Log: It begins at the last checkpoint (instead of the oldest uncommitted transaction).

Accelerated Database Recovery Process

Undo phase

Using sLog to undo non-versioned operations and Persisted Version Store (PVS) with Logical Revert to provide row level version-based Undo, the Undo phase with ADR is virtually quick.

Four Key Components of ADR

Persisted version Store (PVS):

It is a new SQL Server database engine.

Logical Revert:

It is an asynchronous mechanism that performs row-level version-based undo.

sLog:

It is a secondary in-memory log stream that stores log records for non-versioned operations.

Cleaner:

It is an asynchronous operation that runs regularly and removes outdated page versions.

Accelerated Database Recovery Patterns

The workloads listed below benefit the most from ADR:

- Workloads with long-running transactions
- Workloads that have seen cases where active transactions are causing the transaction log to grow significantly
- Workloads that have experienced long periods of database unavailability due to long - running recovery

Assisted Practice

Azure SQL DB – Geo-Replication

Duration: 10 Min.

Problem Statement:

Demonstrate the use Azure SQL DB Geo-Replication to construct readable secondary databases of individual databases on a server in the same or a different data center.

Assisted Practice: Guidelines

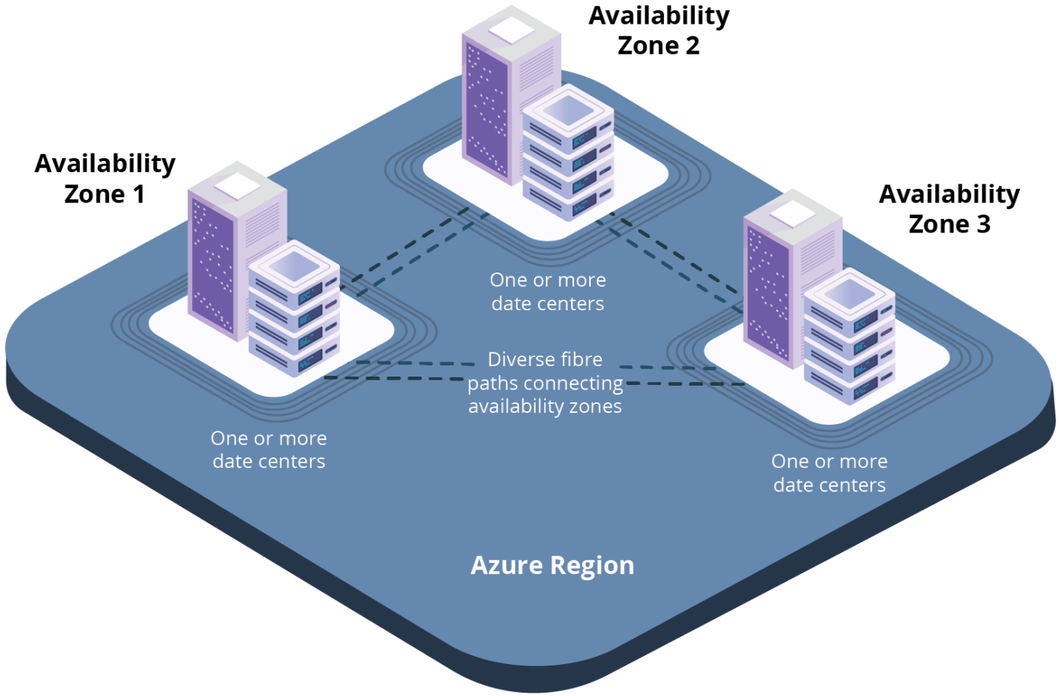
Steps to create a secondary SQL database in the Resource Group database in Azure using the geo-replication:

1. Create Azure SQL database
2. Create a secondary database using the geo-replication

Recommend a High Availability Solution for Non-Relational Data

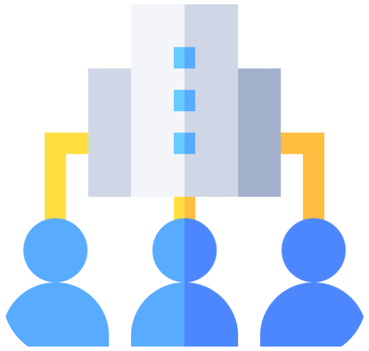
High Availability Using Availability Zones

Azure regions and availability zones are meant to assist users in achieving resiliency and reliability for their goal workloads.



Datacenters are placed within a latency-defined perimeter in each Azure region.

Delivering Reliability in Azure



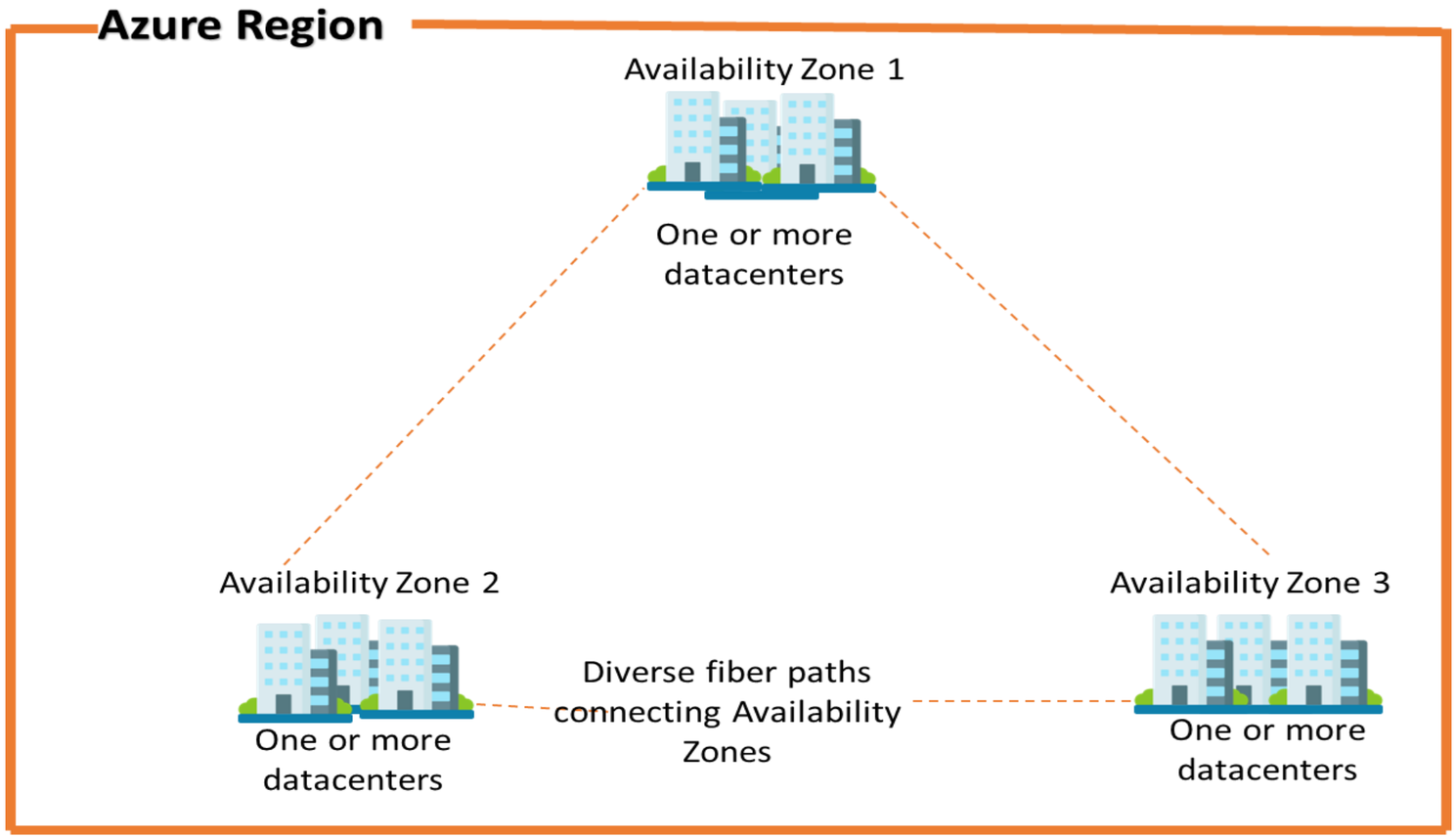
Resilient foundation

Resilient services

Resilient applications

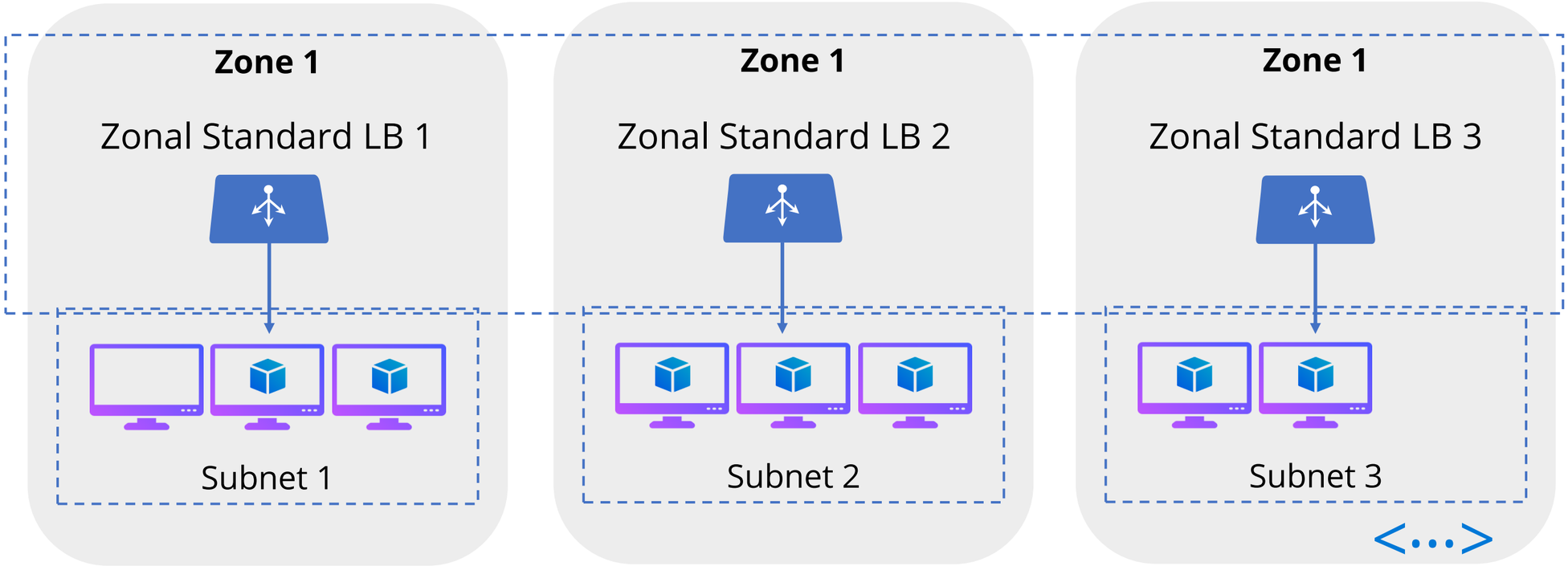
Zonal vs. Zone-Redundant Architecture

An Azure Availability Zone is made up of a fault domain and an update domain.



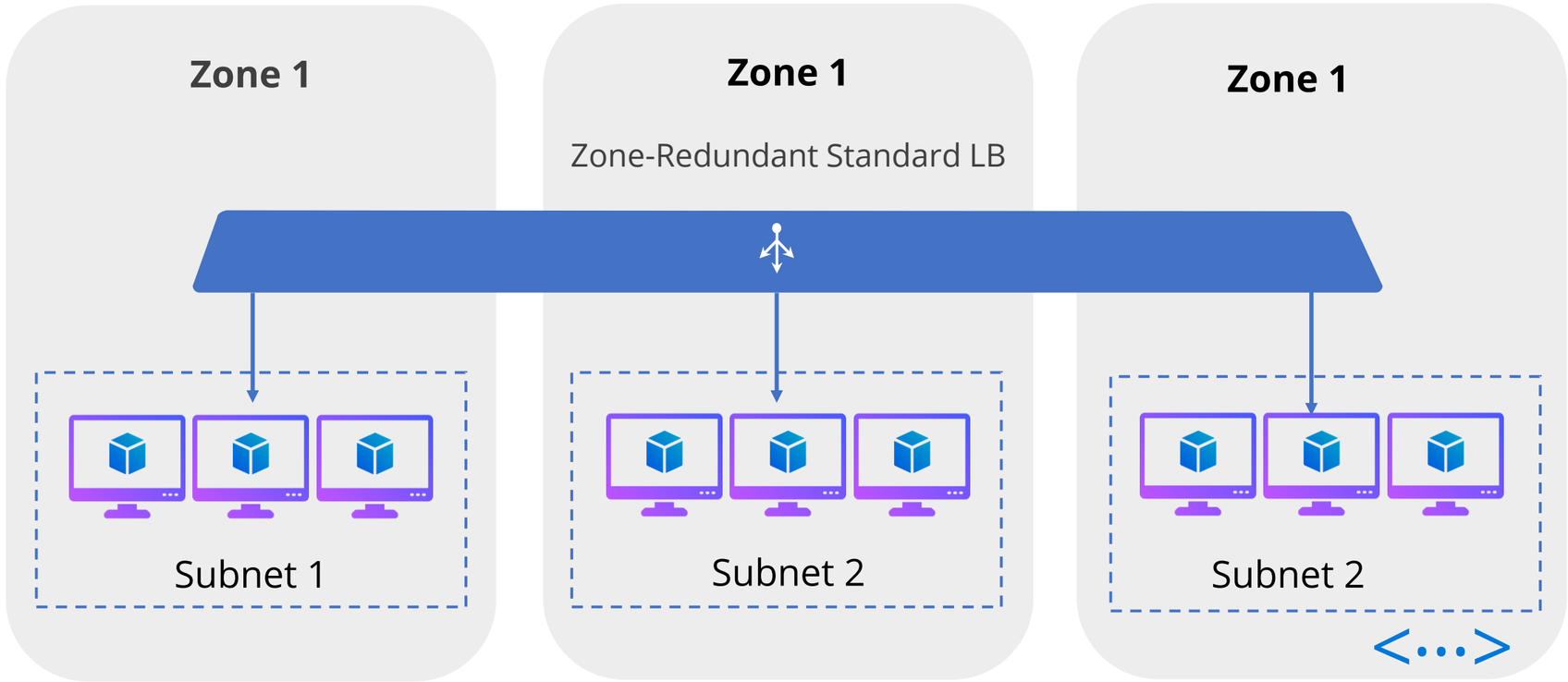
Zonal Architecture

To achieve more restrictive latency or performance criteria, a resource can be deployed to a specific, self-selected Availability Zone using a zonal architecture.



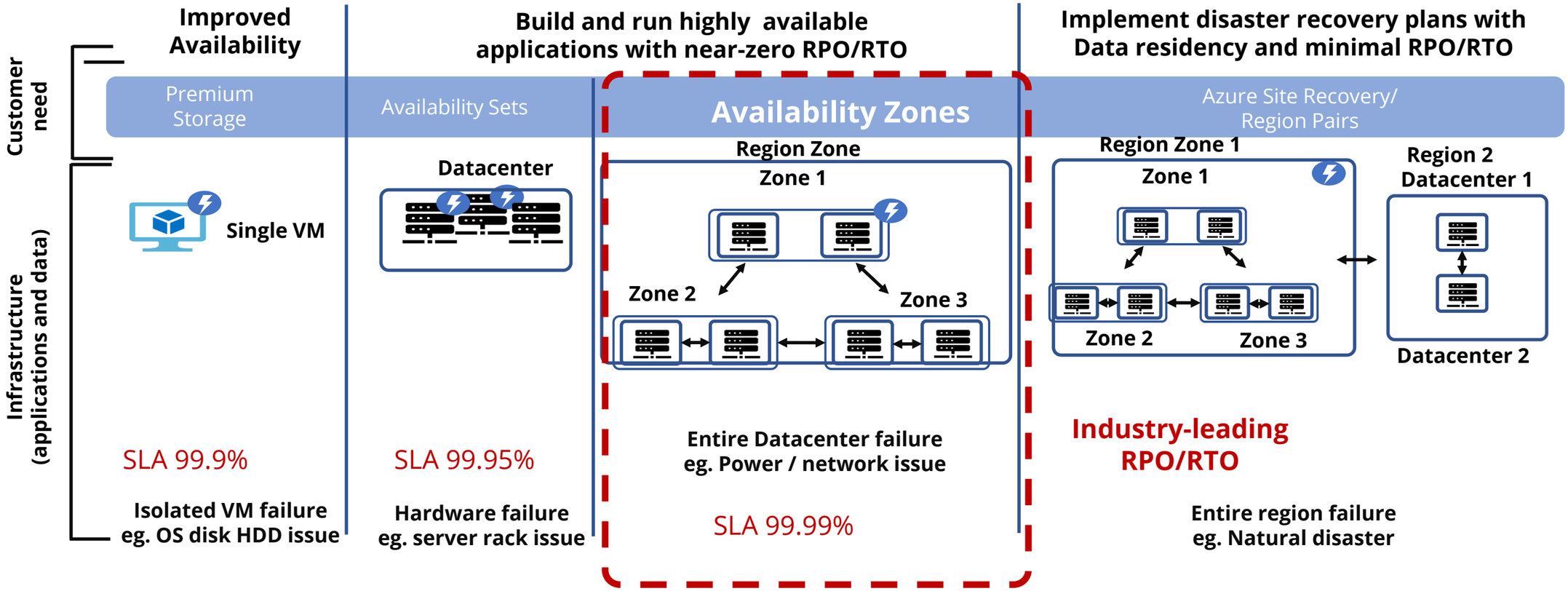
Zone-Redundant

Workload distribution is a characteristic of the service that is managed by Azure.



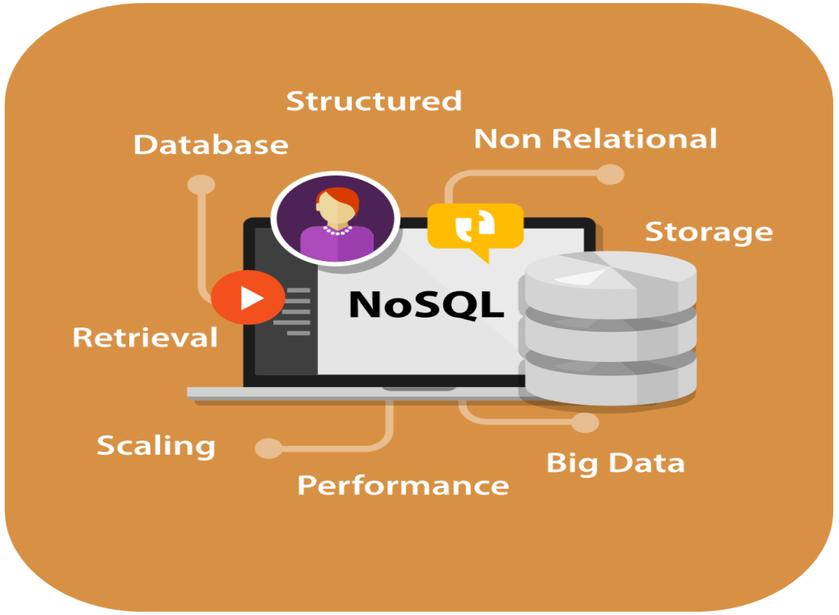
SLA Offered by Availability Zones

The figure below describes various levels of HA provided by a single VM, as well as Availability Sets and Availability Zones.



Relational vs. NoSQL Data

The two types of database systems used in cloud-native apps are relational and NoSQL.



They are designed to store data in different ways and can be accessed in multiple ways.

NoSQL

High-performance, non-relational data stores are referred to as NoSQL databases.

They stand out for their ease of use, scalability, resilience, and availability.

NoSQL Databases

There are various types of NoSQL databases:

MongoDB

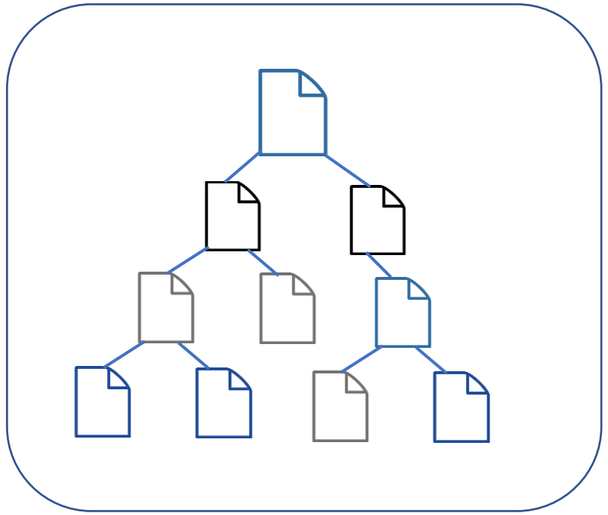
Key value store

Tabular

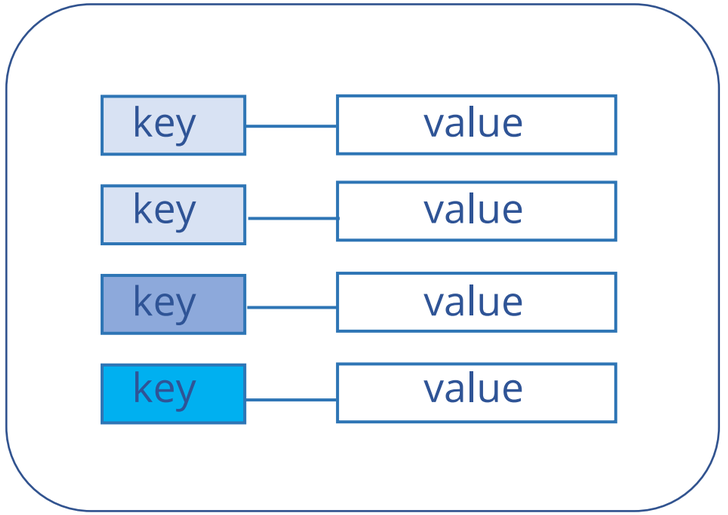
Document based

NoSQL Databases

There are four types of NoSQL databases models:



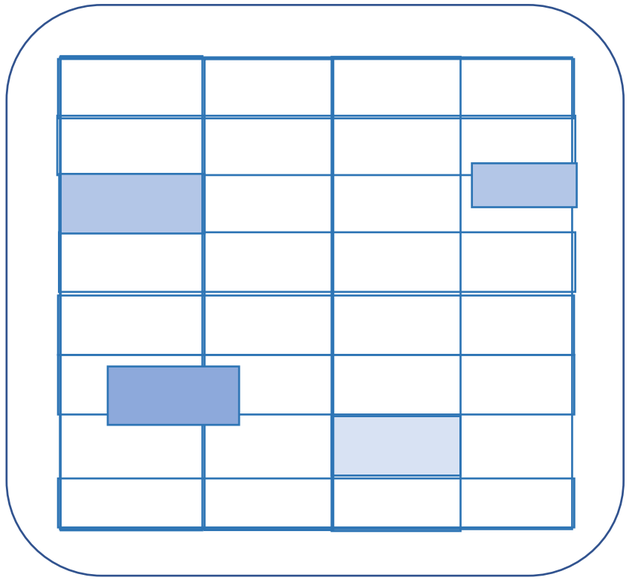
Document Store



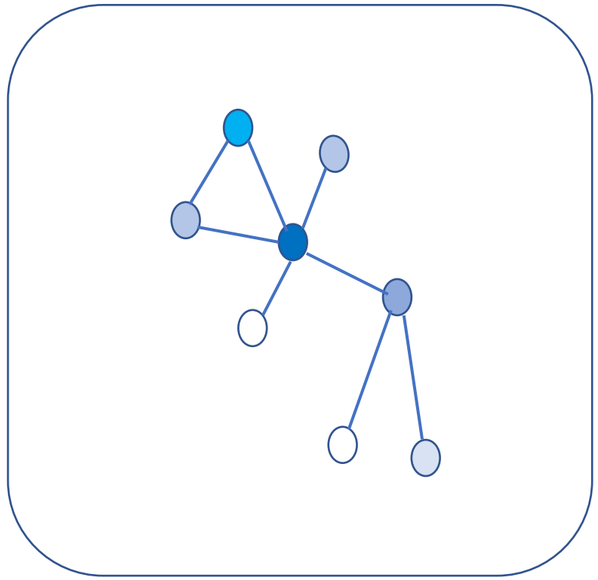
Key-Value Store

NoSQL Databases

There are four types of NoSQL databases models:



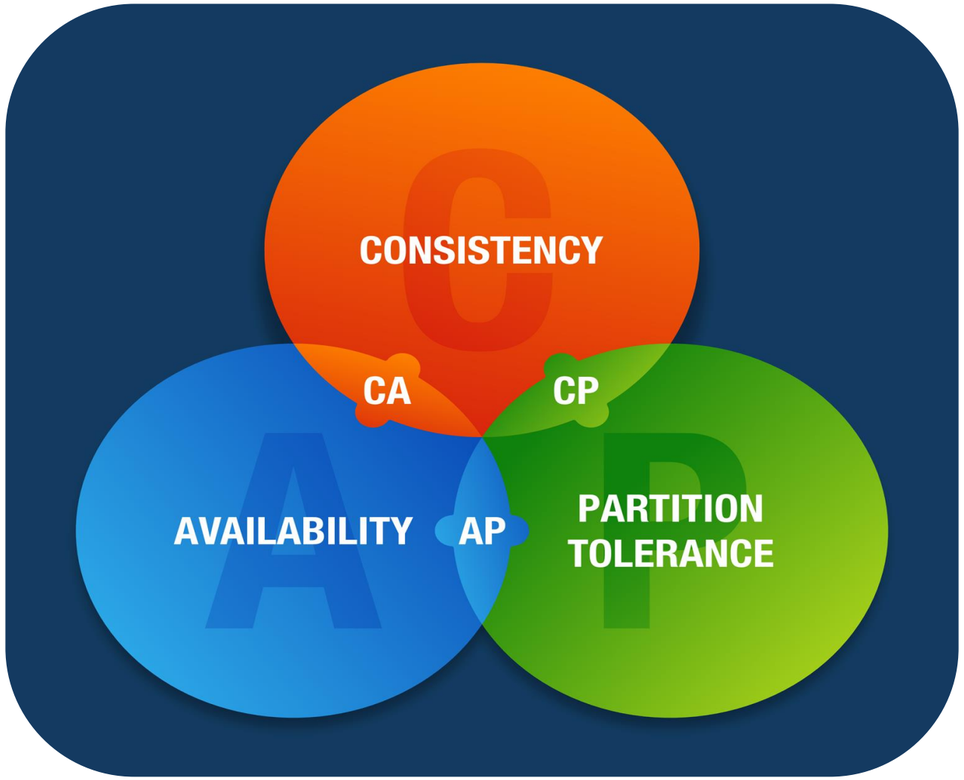
Wide-Column Store



Graph Store

CAP Theorem

A set of concepts that applies to distributed systems that maintain state to better understand the differences between various types of databases



Properties of CAP Theorem

There are three types of properties:

Consistency

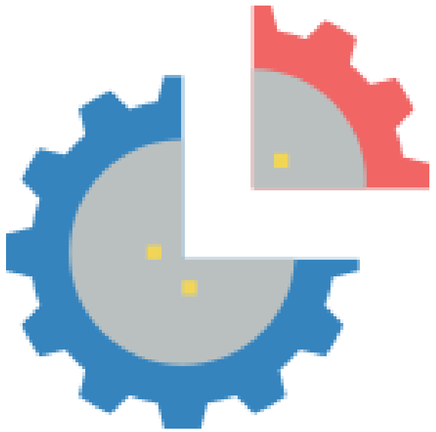
Even if the system must block the request until all replicas update, every node in the cluster replies with the most latest data.

Properties of CAP Theorem

Availability

Every node responds right away, even if the data it returns is not the most recent.

Properties of CAP Theorem



Partition tolerance

Even if a replicated data node fails or loses connectivity with other replicated data nodes, the system will continue to function.

Database as a Service

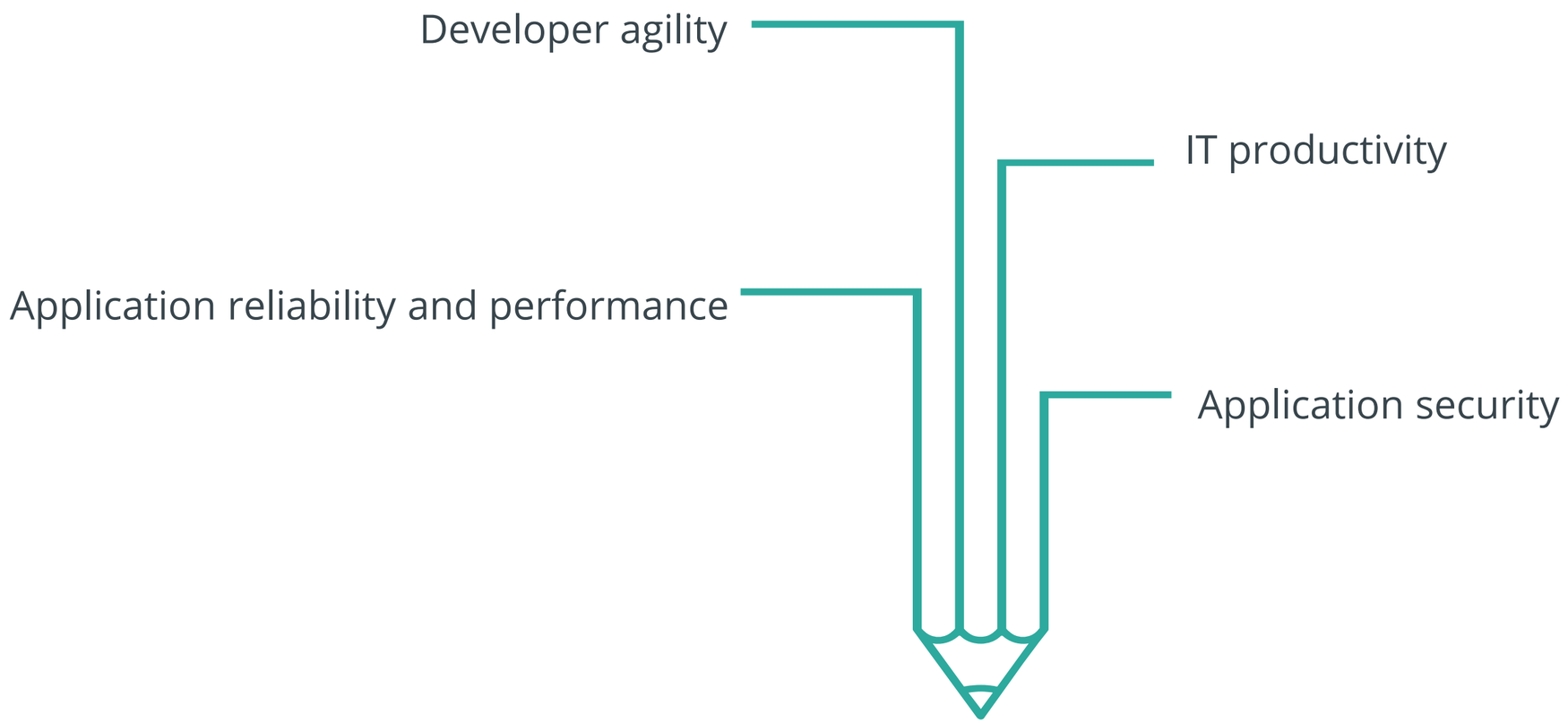
Users may set up, run, and scale databases using a common set of abstractions (primitives) without having to know or care about the actual implementations.

DBaaS on IaaS

The DBaaS solution would request resources from the underlying IaaS, which would handle the provisioning of computing, storage, and networking as needed, and effectively eliminate the need for IT involvement.

Benefits of DBaaS

The benefits of DBaaS are:



Azure Relational Databases

Azure provides four managed relational databases as a service (DBaaS) packages for cloud-native microservices that require relational data, as depicted in the figure:

SQL Database	My SQL	MariaDB	PostgreSQL
Shared Database Services Platform	•Built-in high availability with a 99.99% service level agreement •Automatic upgrades, patching, and backups •On-demand scaling with pay-as-you-go pricing •Intelligence: Advisors, tuning , and monitoring •Active geo-replication with readable secondary databases •Enterprise-grade security to protect data at rest and in motion		
Azure Compute layer			
Azure Storage layer			
Globally available across 54 Azure regions			

Azure SQL Database

It is a Microsoft SQL Server Database Engine-based, fully managed relational database-as-a-service (DBaaS).

Deployment Options

1. A Single Database is a fully managed SQL Database that runs in the Azure cloud on an Azure SQL Database server.

2. A Managed Instance of the Microsoft SQL Server Database Engine is a fully managed instance that delivers near-100 percent compatibility.

3. Serverless computing in Azure SQL Database is a compute tier for a single database that scales automatically based on workload requirements.

Identify Storage Types for High Availability

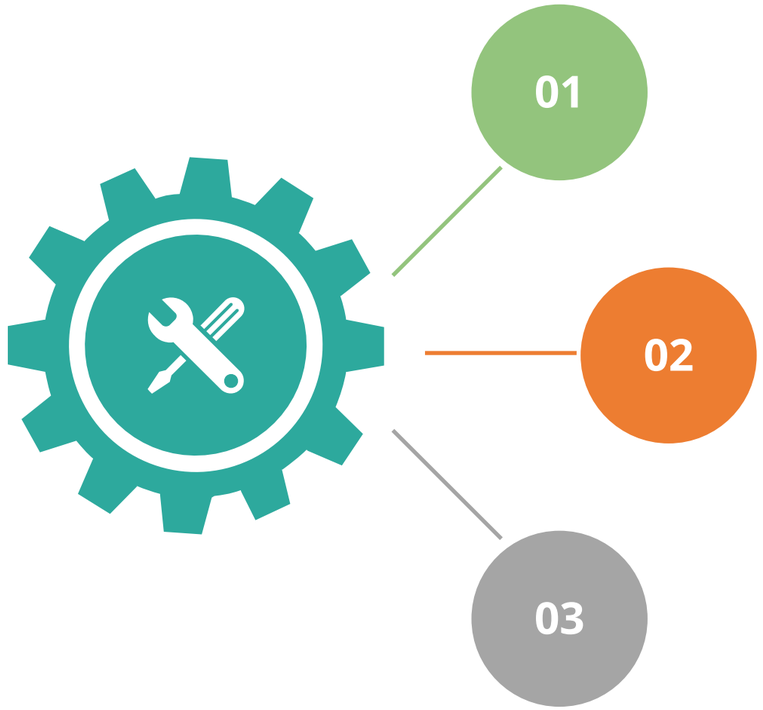
Storage Account Failover

Storage Account Failover prevents the user against unplanned service outages.

- It is an integral part of the disaster recovery plan.
- It helps the user in the event of the primary endpoint becoming unavailable.

Storage Account Failover

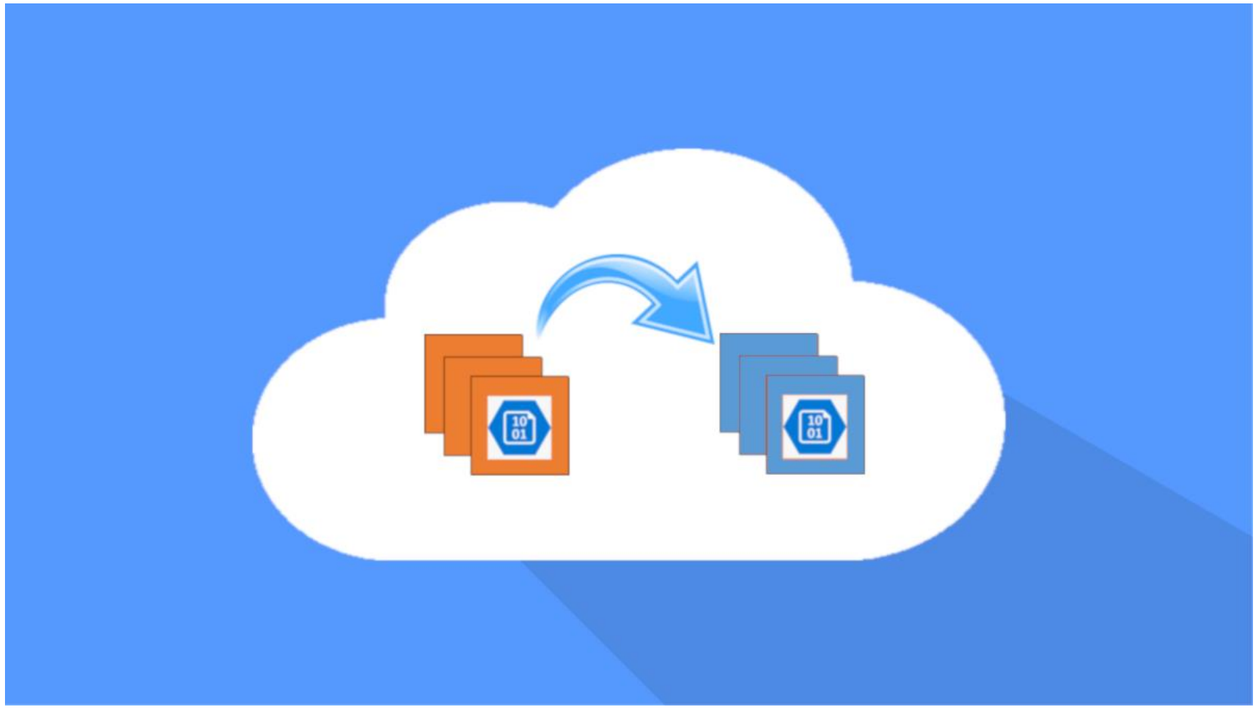
Azure Storage supports account failover for geo-redundant storage accounts. Storage Account Failover is available for the following account types:



- General-purpose V1
- General-purpose V2
- Blob storage account

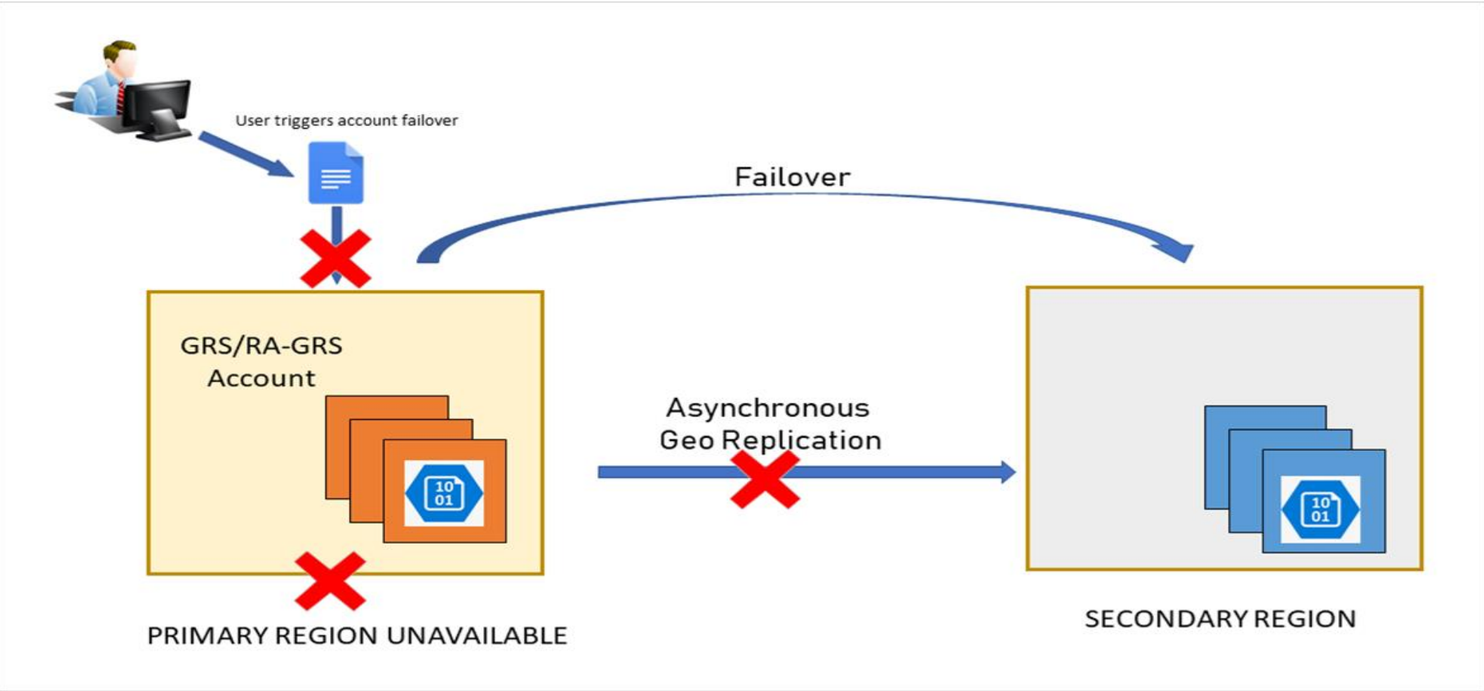
Storage Account Failover

If the primary endpoint becomes impossible to access, the user can start the storage account failover process.



Working of an Account Failover

The failover updates the secondary endpoint to become the primary endpoint for the storage account.



Once the failover is complete, clients can begin writing to the new primary endpoint.

Key Takeaways

- Standard and Premium are the two types of high availability models.
- The user can construct and manage a group of load-balanced VMs using Azure virtual machine scale settings.
- Storage Account Failover prevents the users against unplanned service outages.
- Storage Account Replication ensures user storage accounts have their backup even in the case of failure.

Designing solution for High Availability

Duration: 25 min.

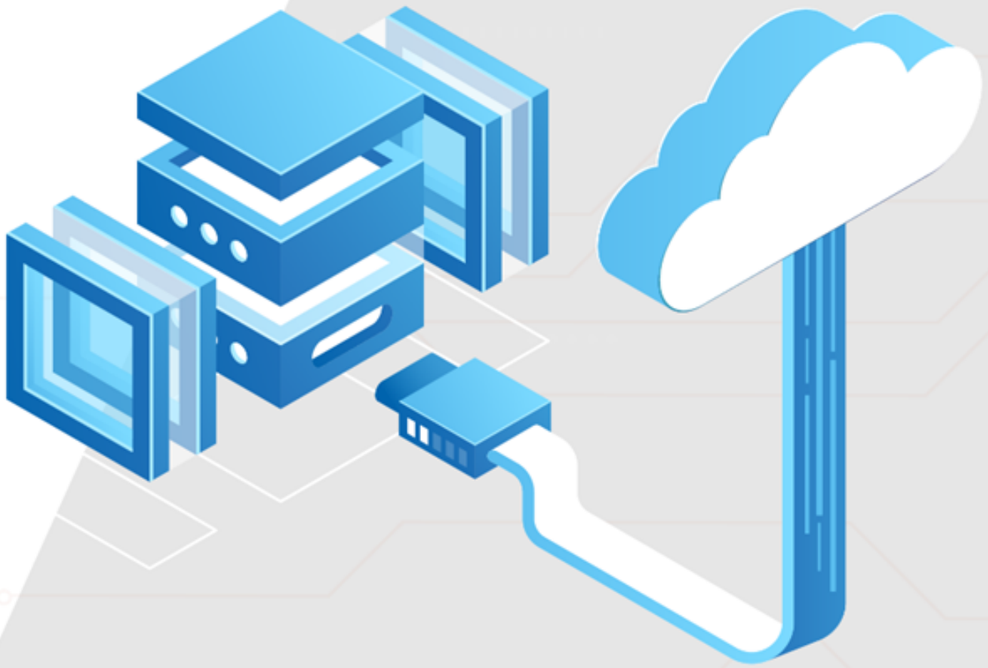
Project Agenda: To design a solution for high availability for the given scenario

Description: You have been given a project to choose the right candidate service for a relational database that can be used for an OLTP application. You need to ensure that during an adverse event like a catastrophic failure or natural disaster, the users are still able to read the data. You need to configure Geo-replication so that you have a secondary database in place.

Perform the following:

1. Login to your Azure account
2. Configure SQL Database and Geo-replication for a secondary database

LESSON-END PROJECT



Thank you