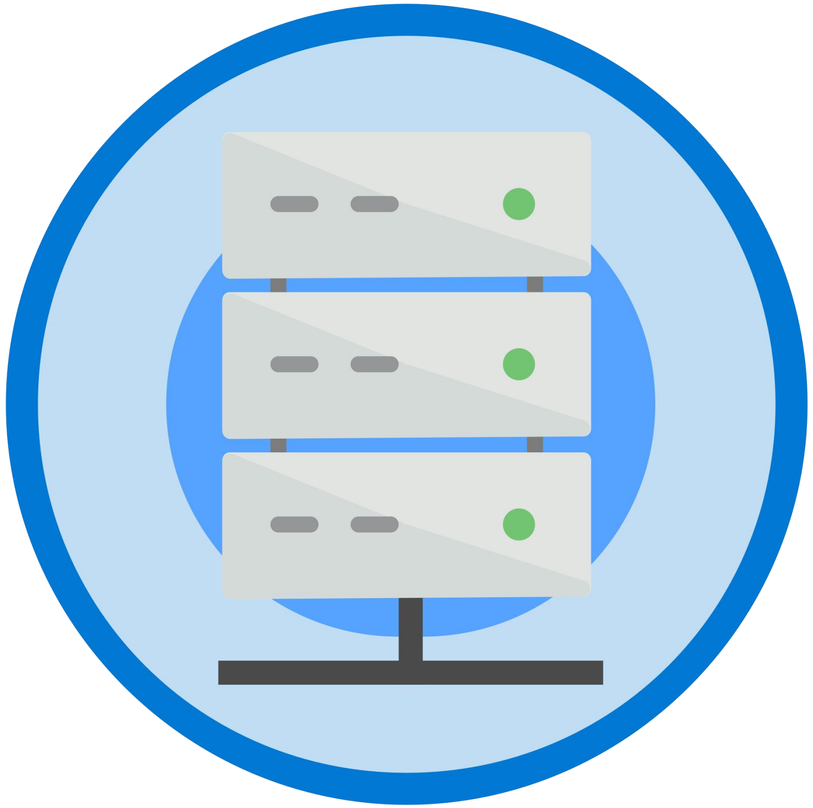
BAIT 3273 Cloud Computing

Week 4

Core Cloud Services
Azure architecture and service guarantees

Lesson Objectives:

- To explore the physical structure of Azure infrastructure
- To understand the service level agreements provided by Azure
- To learn how to provide your own service level agreements for your apps



Introduction

Microsoft Azure:

- **Reliable**, **redundant**, **energy-efficient** infrastructure
- More than **100 highly secure facilities** around the world
- Link through one of the **largest networks** on earth.
- Azure allows users to gain **global reach** with **local presence**

Regions

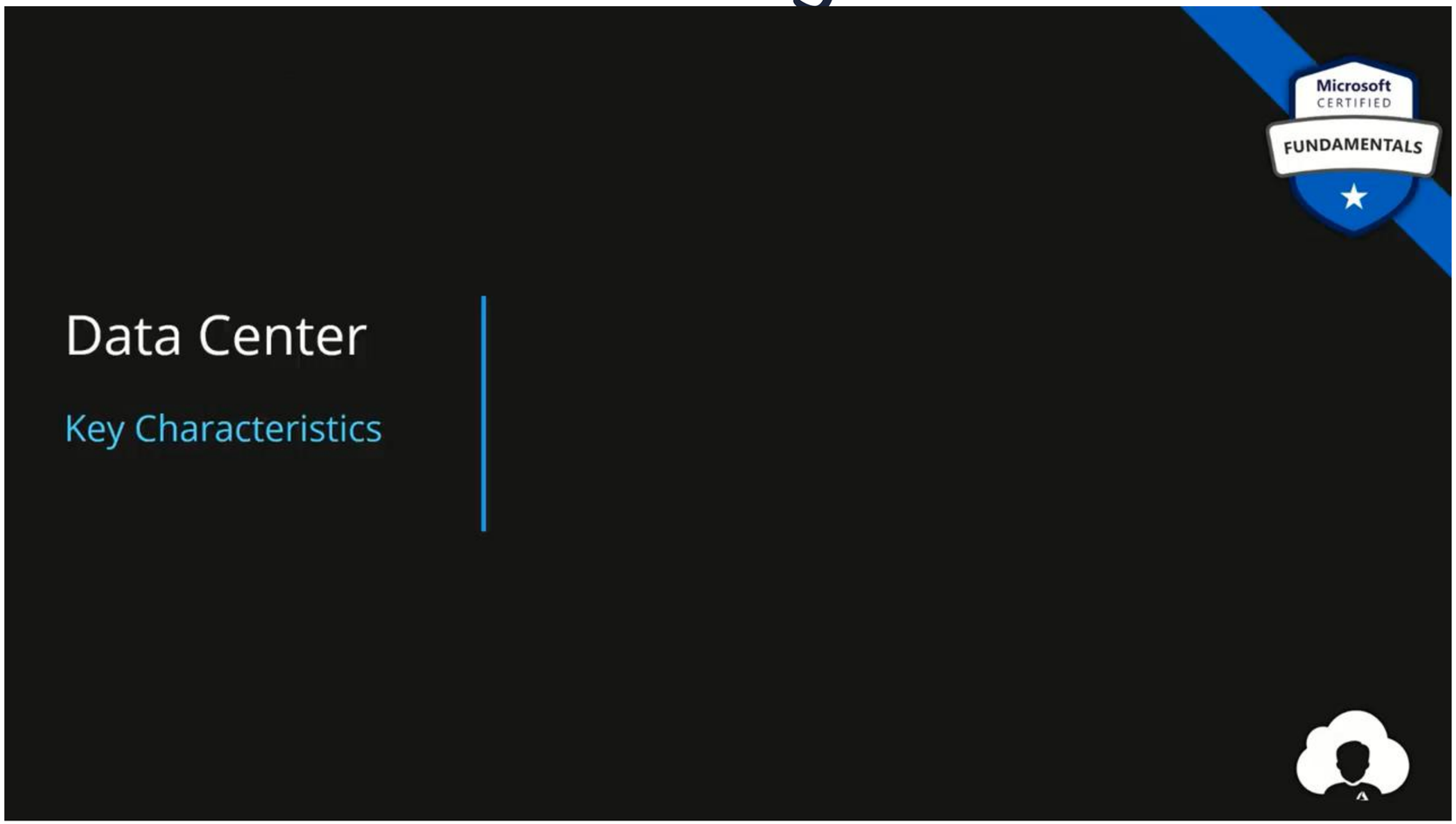
- Azure is made up of datacenters located around the world
- A region is a geographical area on the earth that contains at least one datacenters
- Azure customers will often need to choose a region when he/she deploys a resource.
- Regions provide better scalability, redundancy, and preserve data residency for the services.

Special Azure regions

Azure has specialized regions, which may be needed by the Azure customers when building applications for regulatory compliance or legal purposes. These includes:

- US DoD Central
- US Gov Virginia
- US Gov Iowa
- China East
- China North

Azure regions



Geographies

Azure divides the world into geographies that are defined by geopolitical boundaries or national borders.

Azure geography is a distinct market that usually includes two or more regions that preserves data residency and compliance boundaries.

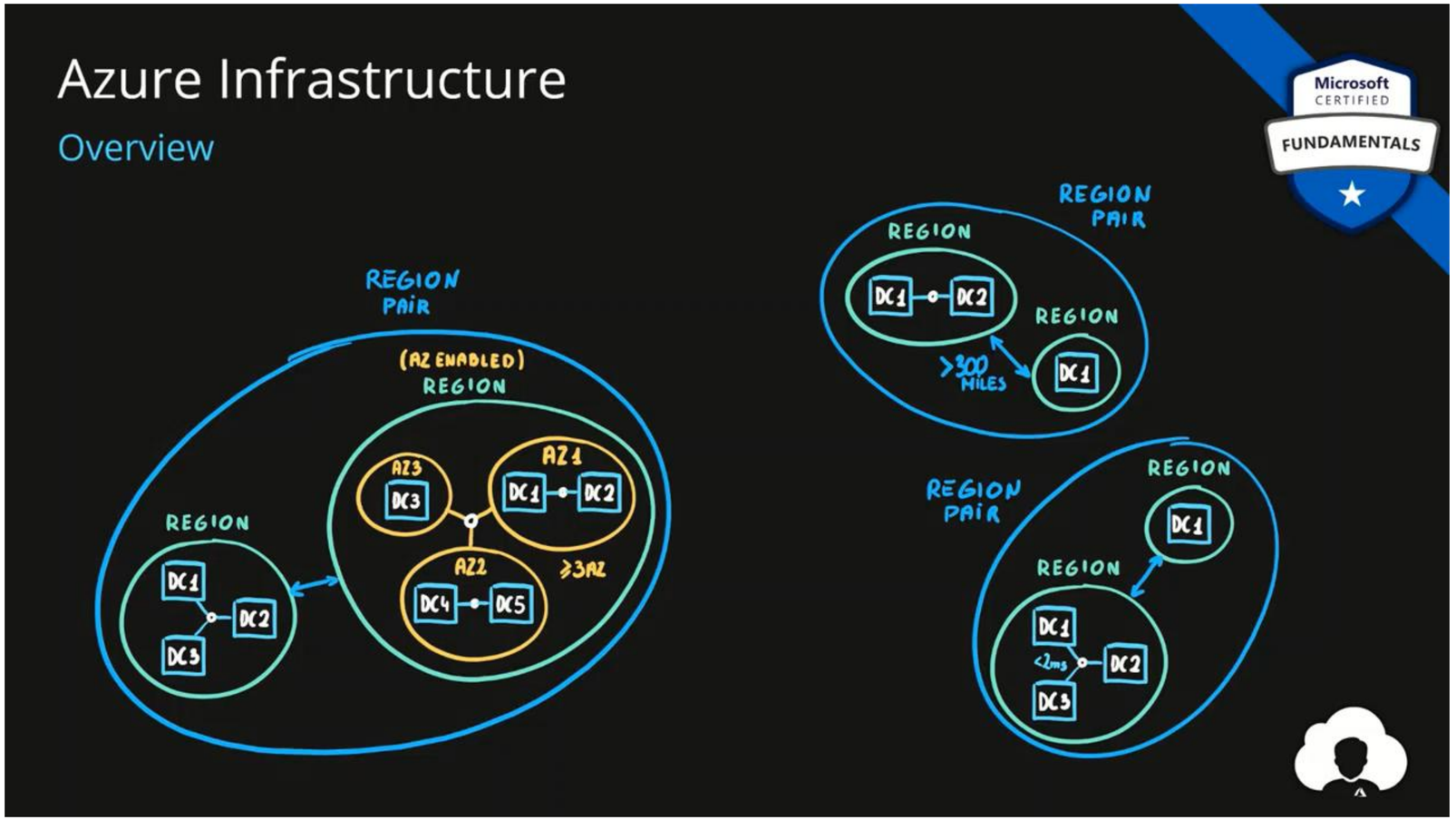
- Allow customers to keep their data and applications closed.
- Ensure data residency, sovereignty, compliance, and resiliency requirements.
- Fault tolerant, able to withstand the failure of the entire region.

Geographies

Azure geographies are broken up into:

- Americas
- Europe
- Asia Pacific
- Middle East and Africa

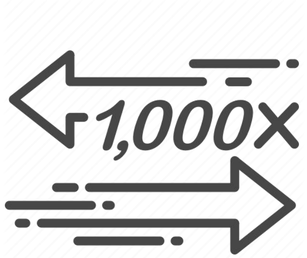
Azure geographies



Availability Zones

- Physically separate datacenters within a region
- Consists of one or more datacenters
- Set up to be an isolation boundary

- Fault Tolerant

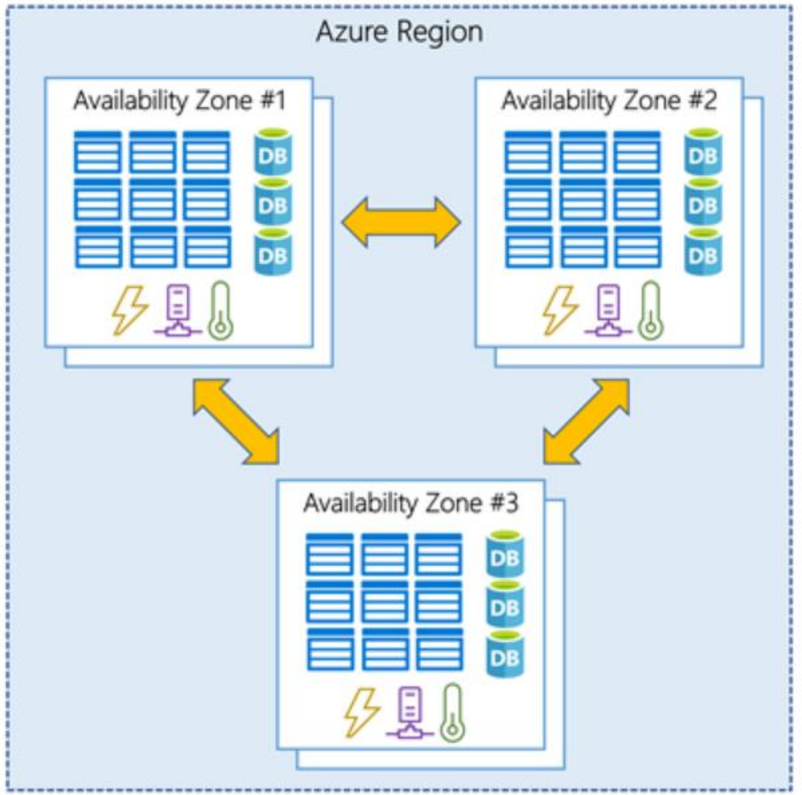


- Connected through high-speed networks

Availability Zones – Supported regions

Not every region supports Availability Zones. The following areas have at least three separate areas to ensure flexibility.

- Central US
- East US 2
- West US 2
- West Europe
- France Central
- North Europe
- South East Asia



Using Availability Zones

Run mission-critical applications

Build high-availability applications

Replicating resources in other zones

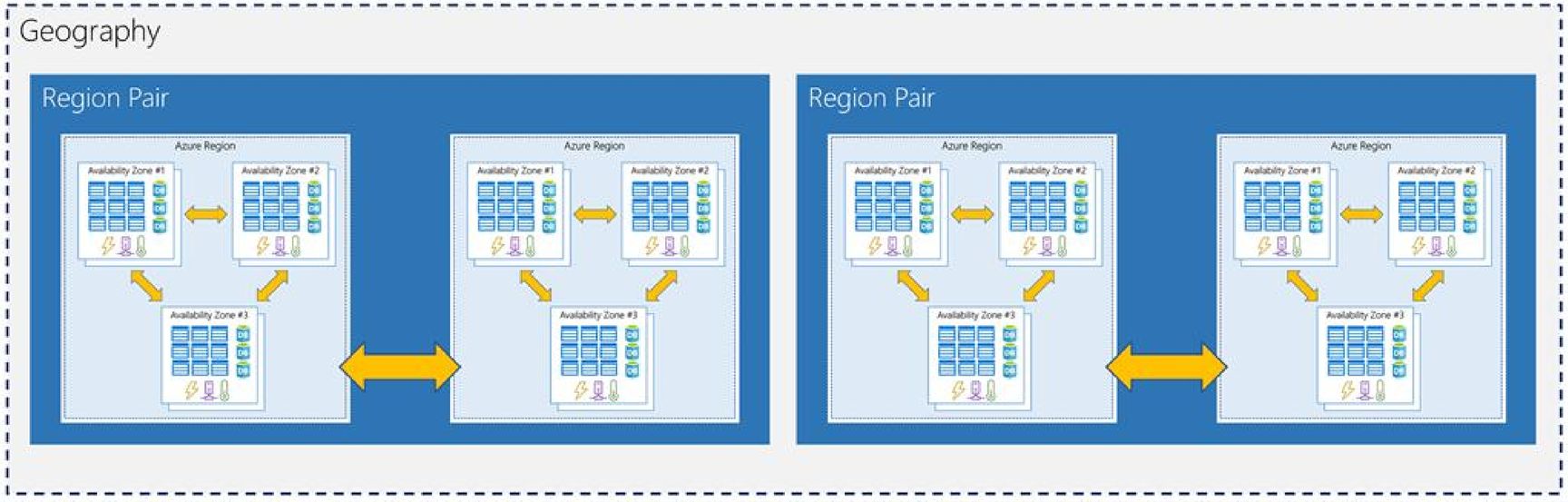
Divided into two categories:

- Zonal Services
- Zone-redundant services

Available Zone

Region Pairs

- Paring Azure region with another region in the same geographic area
- Each region must at least 300 miles away
- Replicate resources to other region within a region pair
- Whenever there is a region failure, all the services will be automatically fail over to another region in its region pair.

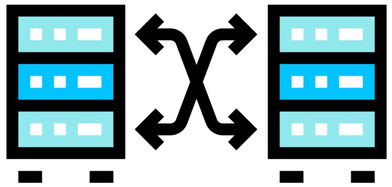


Region Pairs

Example of region pairs in Azure:

- West US and East US
- South East Asia and East

Pros: Asia

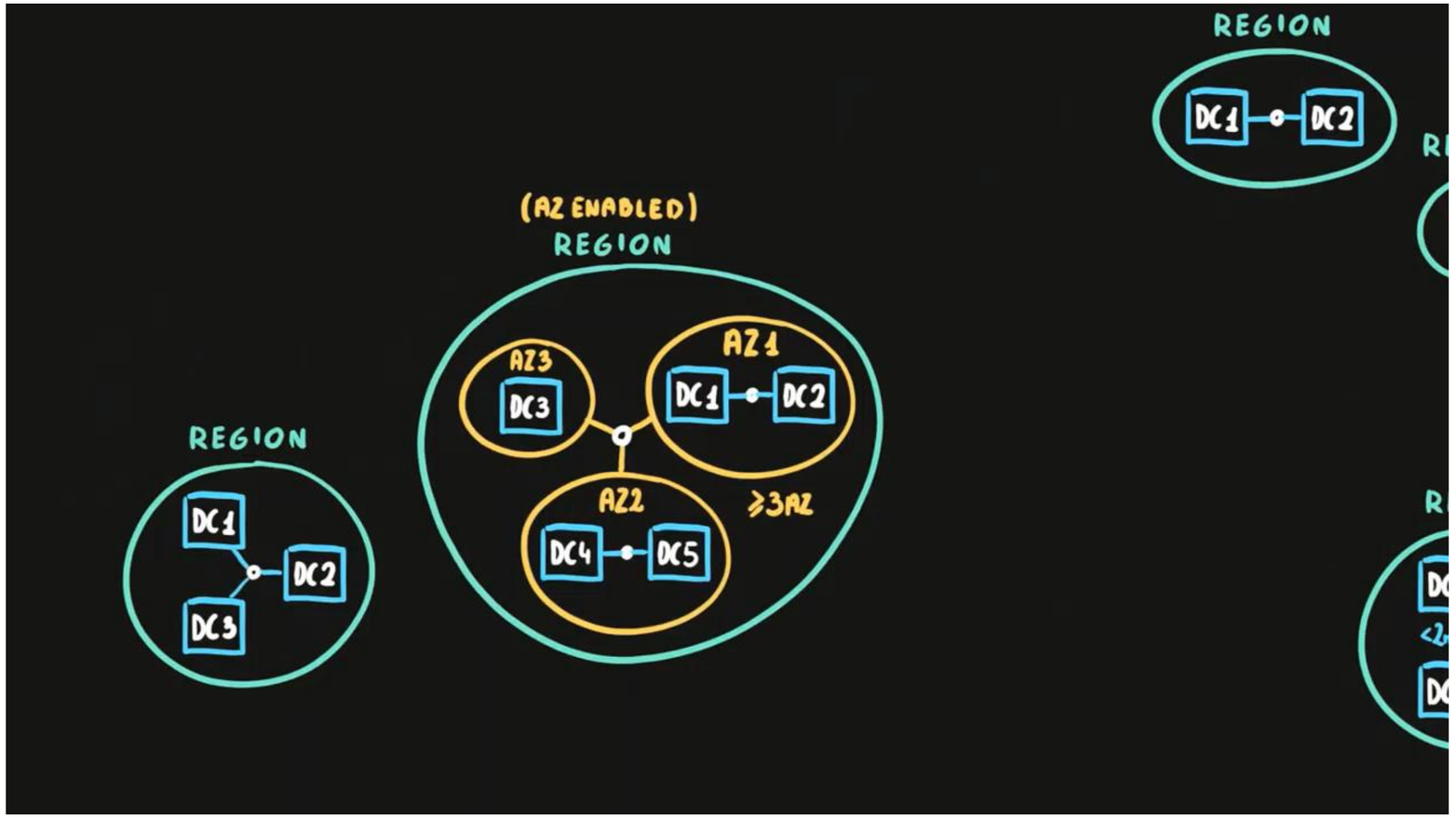


- Reliable
- Data redundancy

- High availability
- Minimize downtime

- Quick Restore

Region pairs



Service-Level Agreements (SLAs)

SLAs specify Microsoft's commitment to provide Azure users with high-quality services by adhering to specific performance standards. SLAs also specify what will happen if the service or product fails to meet the SLA's specification.

Characteristics of SLAs:

1. Performance Targets
2. Uptime and Connectivity Guarantees
3. Service Credits

Uptime and Connectivity Guarantees

SLA %	Downtime per week	Downtime per month	Downtime per year
99	1.68 hours	7.2 hours	3.65 days
99.9	10.1 minutes	43.2 minutes	8.76 hours
99.95	5 minutes	21.6 minutes	4.38 hours
99.99	1.01 minutes	4.32 minutes	52.56 minutes
99.999	6 seconds	25.9 seconds	5.26 minutes

Service Credits

Monthly Uptime Percentage	Service Credit Percentage
< 99.9	10
< 99	25
< 95	100

SLAs

Cloud Computing

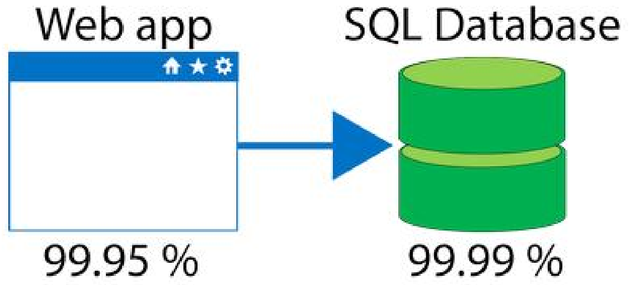
Service Level Agreement (SLA)

- What is SLA??

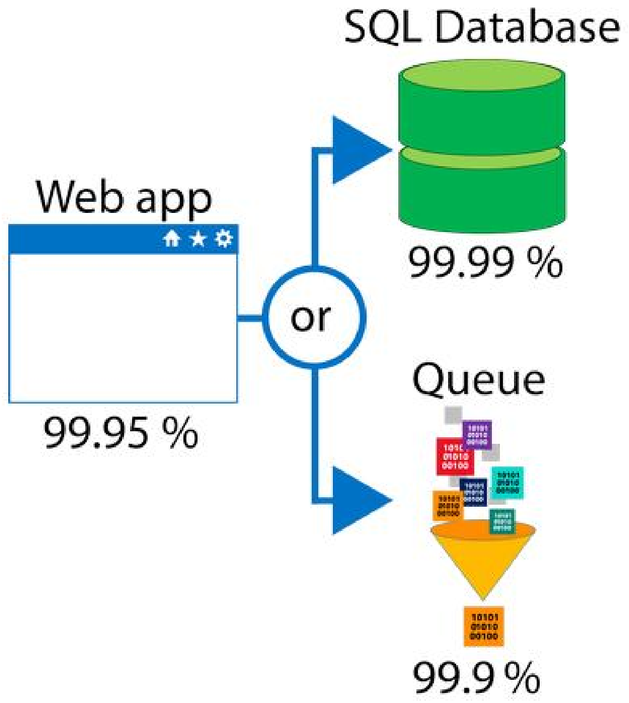
Compose SLAs across service

- Combination of different SLAs is known as Composite SLA
- Composite SLA may have higher or lower uptime values

Calculating Downtime



99.95% x 99.99% = 99.94%



1.0 - (0.0001 x 0.001) = 99.99999 %

99.95% x 99.99999% = ~99.95%

Composite SLAs

Pricing and Support – Composite SLA

Improve app reliability in Azure

SLAs can be used:

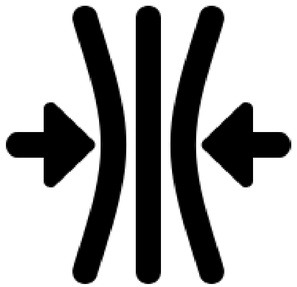
- Evaluate Azure solutions
- Specify performance targets
- Known as Application SLA

Improve app reliability in Azure

Understand app requirements

- Develop efficient and reliable solutions
- Select suitable services and resources
- Assist in creating achievable Application SLAs

Resiliency



- Disaster recovery
- Minimize downtime
- Prevent data loss
- Recover system from failure

Improve app reliability in Azure

Cost and complexity vs. high availability

- Availability
- Cost
- Solution Complexity

Considerations for defining application SLAs

Application with 99.99% performance targets:

- Azure solutions must be self-diagnosing and self-healing.
- Difficult to respond quickly to failures
- The smaller the time window, the tighter the tolerance.

Summary

- Azure provides **global reach** and **local presence**
- **Azure geographies** and **regions** bring applications closer to Azure customers
- **Availability Zones** and **Region pairs** ensure the services can keep functioning
- Azure specify **SLAs** to guarantee the uptime of Azure services

Check your knowledg

1. Deploying an app can be done directly to what level of physical granularity?

- ○ Region
- ○ Datacenter
- ○ Server rack

2. To use Azure datacenters that are made available with power, cooling, and networking capabilities independent from other datacenters in a region, choose a region that supports _____?

- ○ Geography distribution
- ○ Service-Level Agreements (SLAs)
- ○ Availability Zones

3. Application availability refers to what?

- ○ The service level agreement of the associated resource.
- ○ Application support for an availability zone.
- ○ The overall time that a system is functional and working.

Answer

1. Deploying an app can be done directly to what level of physical granularity?

- (●) Region ✓

  Correct. Azure organizes infrastructure around *regions*, which include multiple datacenters. You can pick the region you want resources deployed into. You can't select a specific datacenter or location within a datacenter.

- ( ) Datacenter
- ( ) Server rack

2. To use Azure datacenters that are made available with power, cooling, and networking capabilities independent from other datacenters in a region, choose a region that supports _____?

- ( ) Geography distribution
- ( ) Service-Level Agreements (SLAs)
- (●) Availability Zones ✓

  Availability Zones are datacenters set up to be an isolation boundary from others in the region, with their own power, cooling, and networking. If one zone in a region goes down, other Availability Zones in the region continue to work.

3. Application availability refers to what?

- ( ) The service level agreement of the associated resource.
- ( ) Application support for an availability zone.
- (●) The overall time that a system is functional and working. ✓

  Correct. The time that a system is working is referred to as the application availability.

Thanks!