## <https://azure.microsoft.com/en-us/support/legal/sla/virtual-machines/v1_9/>

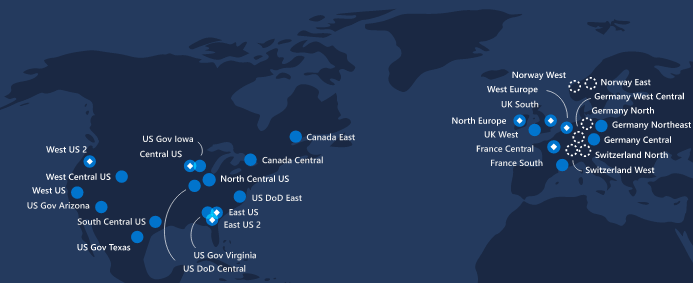
SLA for VM’s

## Microsoft Azure High Availability Key Features

There are four key features of Microsoft Azure that provide the core framework of availability to applications that are hosted in the Microsoft Azure public cloud. They are:

* Availability Sets
* Availability Zones
* Fault domain
* Update domain

Let’s take a look at each one of these fundamental availability features and mechanisms to Microsoft Azure and see how they contribute to the availability of business-critical applications hosted in Microsoft Azure.



Microsoft Azure provides resiliency across many layers of the infrastructure

## Microsoft Azure Availability Sets

The first cornerstone of high availability in the Azure public cloud is the Availability Set. When thinking about any resource that is housed in the public cloud that is critical for business continuity, making the resource redundant or able to withstand failure is extremely important. Redundancy means that even if a particular resource fails, another resource is available to take the place and assume responsibility for the workload the failed resource was processing.

Even though the Azure public cloud is extremely resilient to a failure in a particular Azure datacenter, this does not make a single VM or other resource running in the specific data center highly available or resilient to failure. For high availability and resiliency, Azure has something called an Availability set.

**What is an availability set?**

An Azure availability set is a logical grouping capability for isolating VM resources from each other when they are deployed. The availability set is configured to run across multiple physical servers, compute racks, storage units, and network switches. With the availability set implemented, any failure in the underlying infrastructure in Azure will only result in a subset of the VMs being impacted. The overall solution stays operational.

**What is availability set?**

Availability Set. A group with two or more virtual machines in the same Data Center is called Availability Set, this ensures that at least one of the virtual machines hosted on Azure will be available if something happens.

When a new VM is deployed in Azure, the availability set is defined as a parameter in the new VM creation. The new VMs are isolated across multiple physical infrastructures. Again, this ensures that if there is a failure, one of the instances of the VM will continue running since it will be on different physical infrastructure.

## Availability Zones

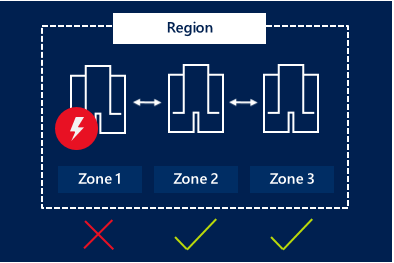
The availability zone takes the availability concept to a new level. This is a fairly new addition to the available Azure high availability mechanisms. Instead of simply protecting from underlying hardware failure for a VM, the availability zone protects applications and data from entire data center failures. The availability zone is a unique physical location within Azure with each zone made up of one or more data centers equipped with different supplying infrastructure. This includes different power, cooling, and networking. For resiliency purposes, there is a minimum of three separate zones in all the availability zone enabled regions.

Azure zone-redundant services replicate applications and data across availability zones to protect from single points of failure. Azure offers an SLA of 99.99% uptime for a virtual machine housed in the availability zone.

The availability zone technology combines the fault domain and update domain in a way that provides distributes both updates and fault domains across data centers.

In a scenario where three or more VMs are created in three zones in an Azure region, the VMs are distributed across three fault domains and three update domains. Azure services that support availability zones fall into two categories:

* Zonal services – resources are pinned to a specific zone
* Zone-redundant services – replication of the platform happens automatically across zones



Availability zone provides redundancy at the data center level

For the utmost in redundancy, applications need to be built using the combination of Availability Zones with Azure region pairs. Data can be synchronously replicated using Availability Zones within an Azure region for high availability. Additionally, data can be asynchronously replicated across Azure regions for DR purposes.

## Azure Fault Domain

**What are Azure Fault domains?**

The fault domain is a logical group of the underlying hardware that shares a common power source and network switch. This is very similar to the concept of a fault domain on-premises. When virtual machines are created in Azure, the Azure platform automatically distributes VMs across fault domains. This helps to limit the impact of potential physical hardware failures that may include power disruptions or potential network outages that may affect the availability of a virtual machine.

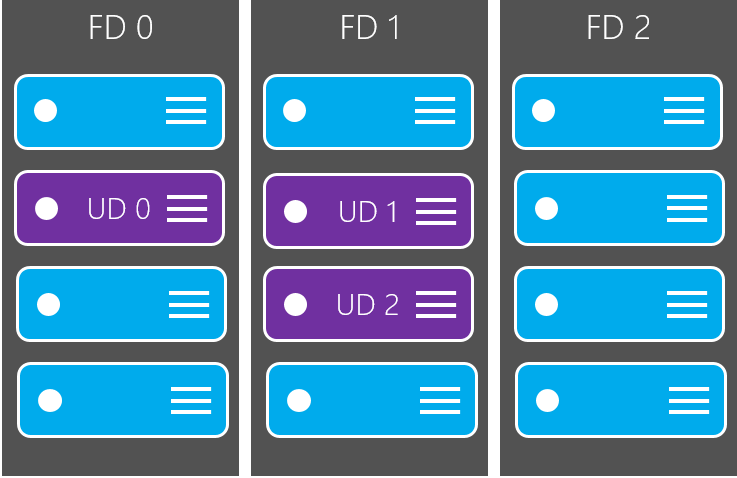
## Update Domain

Another key concept in terms of high availability in Microsoft Azure is the update domain.

**What are update domains?**

This is a logical group of the underlying hardware that undergoes maintenance at the same time. This can include any reboot operations. When virtual machines are created in an availability set, Azure automatically distributes VMs across update domains. This helps to ensure that at least one instance of your application will stay running as Azure components undergo periodic maintenance.

According to Microsoft’s documentation of the update domain, only one update domain is rebooted at a given time, however, the update domains may not be rebooted and proceed sequentially.



Logical representation of an update domain in Microsoft Azure

## Concluding Thoughts

Microsoft’s Azure public cloud environment is an extremely resilient public cloud environment that provides redundancy at many levels down to the level of the rack a VM runs inside the Azure data center and when the underlying server hosts are rebooted.

Microsoft has obviously put a lot of thought into how resiliency and high-availability are potentially affected by various operations and maintenance of applications with Microsoft Azure. Businesses running critical applications inside Azure have these various layers to ensure the availability of the solution regardless of the outage or maintenance operation at hand.