LTIMindtree

Whitepaper

Achieving Resiliency in Microsoft Azure Cloud

Authors: **Akhtar Shaikh, Mohammed Kasim**

Contents

Abstract

In today's fast-moving world, users want mobility and connect anytime-anywhere, across multiple platforms and multiple locations. Any downtime or data loss leads to depletion in business. Cloud model is an affordable solution that provides low-cost disaster recovery and high replication. It offers huge advantages in terms of agility, reduced risk, low cost, and quick deployment. Planning for failure and disasters in the cloud requires understanding of what can cause an outage, responding to failures in a way that minimizes downtime or data loss, recognizing the failure quickly, and having a plan to recover or restore the service if something does go wrong. Additionally, we must also consider the extent of data loss that can be tolerated by the application without causing adverse business consequences, and then implement a strategy to achieve cloud resiliency.

In this paper, we have focused on resilience design best practices for various application deployments with varied resilience requirements and defined the required level of resilience based on business needs. The right decisions related to cloud are critical for organizations to reduce their overall spending and increase their ability to respond to cloud-related risks, threats, and opportunities. We will be providing guidance on designing resilient Infrastructure-as-a-Service (IaaS) applications on Azure and also provide sample application design patterns for varied levels of resilience, solutions to architects and developers with best architectural practices of designing resiliency for those who want to move applications from on-premises environments to Azure or are building solutions that will be deployed on Azure Cloud Services. We also outline the process for achieving resilience using a structured approach throughout the lifetime of an application by providing architectural assistance and technical suggestions on design patterns and their practicality in the context of cloud computing - from design and implementation to deployment and operations.

1. Overview of Resiliency

Resiliency is the ability of a system to recover from failures and continue to function. It's not about avoiding failures but responding to failures in a way that avoids downtime or data loss. The goal of resiliency is to return the application to a fully functioning state. Resiliency is often used in highly critical business environments for business function recovery and continuity concerns. Designing fault tolerance for highly critical applications that run on traditional infrastructure is a familiar process, and there are verified best practices to ensure high availability. However, cloud-based architectures tend to fail in a different way than traditional, machine-based architectures.

Building a reliable application in the cloud is not the same as building a reliable application in an enterprise setting. While historically you may have purchased high-end hardware to scale up, in a cloud environment, you must scale out instead of scaling up. Costs for cloud environments are kept low using commodity hardware. Instead of trying to prevent failures altogether, the goal is to minimize the effects of a failure within the system. Although resiliency can be achieved at many different levels, including the application level and the cloud infrastructure level, here we focus on the cloud infrastructure level.

2. Key Aspects of Resiliency

The three important aspects of resiliency are high availability, disaster recovery, and backup.

High Availability (HA): Keeps the application online in the events of high impact maintenance, hardware, datacenter failure, or fluctuations in load.

Disaster Recovery (DR): Protects regional failure by replication of virtual machine data from primary region to DR region using native tools like Azure Site Recovery.

Backup: Replication of virtual machine and data across one or more regions using Azure Backup.

3. Characteristics of Resiliency

Fault tolerance

The ability to remain up and running in the event of a component or service dysfunction. Typically, redundancy is built into cloud services architecture, so if one component fails, a backup component takes its place.

Availability

The ability to keep services up and running for long periods of time, with very little downtime, depending on the service in question.

Scalability

The ability to increase or decrease resources for any given workload. It can add additional resources to service a workload (known as scaling out), or add additional capabilities to manage an increase in demand to the existing resources (known as scaling up).

Self-healing

Enable auto-healing on Azure instances to avoid human intervention and take preventive actions before an unplanned failure.

Automation

Automation eradicates human errors and helps to build resilient architecture with minimal downtime.

Redundancies

With Azure, deploying and managing redundant systems is handled by Microsoft from a hardware perspective. You'll need to focus your planning and resources more on the software side. Designing a redundant application is critical.

Reporting

Being able to measure the health of your systems on a regular basis is a step that is often overlooked. Understanding the performance and health of your applications not only helps provide a positive user experience, but also helps ensure that your application is available and meets the pre-defined SLAs.

4. How to Design a Resilient Cloud Architecture

Single VM Instance Availability

Any application which is not designed to run on multiple VMs or for scaling out, falls under the single VM scenario. These are less expensive options with high availability.

A single Azure VM cannot be banked upon for high availability, although Microsoft gives the guarantee of 99.9% uptime, it must be met with some preconditions to achieve it. For example:

- It should be provisioned on premium disk.
- Microsoft notifies the administrator five days before any pre-planned maintenance.
- There is no protection of Application VM from unplanned maintenance.

To achieve an SLA of more than 99.9%, we must opt for Availability Set, DR, or Backup strategy.

Reasons for outages on a single VM:

- Unplanned hardware maintenance event.
- Unexpected downtime.
- Planned maintenance by Microsoft.

Availability set

For our application to be redundant, it is recommended to keep two or more virtual machines in an availability set. This configuration in a datacenter ensures that during either a planned or an unplanned maintenance event, at least one virtual machine is available and meets the 99.95% Azure SLA. Currently, there is a single-VM SLA of 99.9% with premium storage. The basis for the calculation is 30 days per month or 43,200 minutes. For example, a 0.05% downtime corresponds to 21.6 minutes. When two or more VMs are part of the same availability set, each virtual machine in the availability set is assigned an update domain and a fault domain by the underlying Azure platform.

- Update domains guarantee that multiple VMs are not rebooted at the same time during the planned maintenance of an Azure infrastructure. Only one VM is rebooted at a time.
- Fault domains guarantee that VMs are deployed on hardware components that do not share a common power source and network switch. When servers, a network switch, or a power source undergo an unplanned downtime, only one VM is affected.

By default, the virtual machines configured within the availability set are separated across up to three fault domains for Resource Manager Deployments. Although placing virtual machines into availability set cannot protect the application from operating system or application-specific failures, it does limit the impact of physical hardware failures, network outages, or power interruptions.

Availability zones

Availability zones are unique physical locations which offer to protect the applications and data from datacenter failures. Each zone is made up of one or more datacenters equipped with independent power, cooling, and networking. There's a minimum of three separate zones in all enabled regions to ensure resiliency. The physical separation of availability zones protects applications and data from datacenter failures. Zone-redundant services replicate applications and data across these zones to protect them from single-points-of-failure.

The different zones within a single Azure region enable deployment of applications across two or three availability zones. If there are any issues in power sources and/or network, it would only affect one zone's infrastructure. However, application deployment within the other Azure regions will still be fully functional with some reduced capacity, as some VMs in one zone might be lost. But VMs in the other two zones are still up and running.

High availability can be built into the application architecture by co-locating compute, storage, networking, and data resources within a zone and replicating it to other zones. Azure services that support availability zones fall into two categories:

- Zonal services: All resources are pinned to a specific zone (for example, virtual machines, managed disks, IP addresses).
- Zone-redundant services: Platform replicates automatically across zones (for example, zone-redundant storage, SQL Database).

Things to be considered for using Availability Zones:

- Azure Availability Sets within an Availability Zone can't be deployed. We have to choose either an Availability Zone or an Availability Set as deployment frame for a VM.
- Basic Load Balancer cannot be used to create failover cluster solutions based on Windows Failover Cluster Services or Linux Pacemaker. Azure Standard Load Balancer SKU should be used instead.
- Azure Availability Zones do not give any guarantees of a certain distance between the different zones within one region.
- The network latency between different Azure Availability Zones, within different Azure regions might be different from one Azure region to another.
- Using Azure managed disks is mandatory for deploying into Azure Availability Zones.

5. Essentials of Disaster Recovery in Cloud

Disaster Recovery (DR) is focused on recovering from a catastrophic loss of application functionality. For example, if an Azure region hosting your application becomes unavailable, you need a plan for running your application or accessing your data in another region. Business and technology owners must determine how much functionality is required during a disaster. This level of functionality can take a few forms: completely unavailable, partially available via reduced functionality or delayed processing, or fully available.

Azure site recovery provides a simple way to replicate VMs from either Azure or on-premise. You don't need to provide any additional resources in the secondary region, as it automatically creates the required resources in the target region, based on the source VM settings. It provides continuous replication and enables you to perform application failover quickly. You can also run DR drills by test failover without affecting your production workloads or ongoing replication. A test failover will create the VMs in an isolated network and will be destroyed during clean-up upon testing or validation.

Azure Site Recovery can protect various workloads as stated below:

• **Azure VMs:** Site Recovery can replicate any workload running on a supported Azure VM.

• **Hyper-V virtual machines:** Site Recovery can protect any workload running on a Hyper-V VM.

• **Physical servers:** Site Recovery can protect physical servers running on Windows or Linux.

• **VMware virtual machines:** Site Recovery can protect any workload running in a VMware VM.

6. Azure Backup

Azure Backup is the Azure-based service you can use to back up and restore your data in the Microsoft cloud. Azure Backup replaces your existing on-premises or off-site backup solution with a cloud-based solution that is reliable, secure, and cost-competitive. Furthermore, backup strategy should be defined with a conscious understanding of the business requirements for restoring application data. Also, processes for creating and restoring backup copies of data, either in whole or in part, should be a part of resiliency plan.

Key benefits of Azure Backup: Automatic Storage Management, Unlimited Scaling, Unlimited Data Transfer and Data Encryption.

7. Conclusion

It is important to understand the crucial elements of resiliency while designing applications on Azure Cloud and the process of designing and deploying highly resilient applications/ infrastructure. One must grasp the shared responsibility of the customer and Microsoft and the expectations that go along with the SLA while deploying applications in the cloud.

Designing, implementing, and operating a resilient architecture can be complex. It needs appropriate knowledge and expertise at each step to achieve application availability goals and meet business requirements for resiliency. At any point of failure in the cloud, the techniques and strategies for managing them are different than when failure occurs on premises.

We have observed that despite all the efforts in making the platform reliable, apps can suffer from downtime because of unplanned events such as power failures, data corruption, ransomware attacks, and natural disasters. A highly available, resilient application absorbs fluctuations in availability, load, and temporary failures in the dependent services and hardware. Planning for and implementing resiliency concepts permits the application to continue to operate at an acceptable user and systemic response level as defined by business requirements or application SLA.

About the Authors

Akhtar Shaikh

Cloud Solution Architect, LTIMindtree

With more than 13 years of overall IT experience in design, consulting, implementation, maintenance and management of IT infrastructure, Akhtar works as a Solution Architect with LTIMindtree, providing consultancy and design solutions for Microsoft and AWS technologies. Specialized in Azure, AWS and GCP cloud, Akhtar is responsible to drive RFP/RFQ/RFI to build comprehensive cloud solution response. As a Microsoft Certified Trainer, he conducts webinars and cloud sessions for internal resources.

Mohammed Kasim

Cloud Solution Architect, LTIMindtree

Mohammed has 14+ years of experience in overall IT covering Infrastructure, Platform and Cloud Solutions. As a Cloud Solution Architect, he is responsible for Azure Business and Azure Practice at LTIMindtree. He provides consultancy and design solutions for Microsoft Azure and technology roadmaps for large enterprises. He is also a Microsoft Certified Trainer, CompTIA CTT+ Classroom Trainer and delivered multiple Azure trainings at LTIMindtree.