Azure status history

This page contains all root cause analyses (RCAs) for incidents that occurred on November 20, 2019 or later. Each RCA will be retained on this page for 5 years. RCAs before November 20, 2019 aren't available.

Product: All

Region: All

Date: All

May 2021

5/10

Microsoft Azure Portal - Intermittent Portal Access Issues (Tracking ID GVD7-RDZ)

Summary of Impact: Between 15:24 UTC and 17:55 UTC on 10 May 2021, customers may have experienced intermittent 500-level errors or an intermittent latency when accessing the Azure portal. Azure services were not affected.

Preliminary Root Cause: The Azure portal frontend endpoints in the US North Central region experienced an increase in CPU usage, causing some instances to not serve traffic as fast as expected.

Mitigation: We rerouted traffic around the unhealthy region and scaled out CPU resources in adjacent regions to handle the increase in traffic.

Next steps: We sincerely apologize for the impact to affected customers. We will continue to investigate to establish the full root cause and prevent future occurrences. Stay informed about Azure service issues by creating custom service health alerts: https://aka.ms/ash-videos for video tutorials and https://aka.ms/ash-alerts for how-to documentation.

5/4

Azure Speech Service - West Europe - Mitigated (Tracking ID LLL3-LTZ)

Summary of Impact: Between 06:45 UTC and 11:35 UTC on 04 May 2021, a subset of customers using Azure Speech Service in West Europe may have experienced failures with online transcription, batch transcription, custom speech, and translation.

Preliminary Root Cause: We have determined that during recent deployment a part of the code lost access to KeyVault, preventing the App Service that Azure Speech Service is dependent on from running as expected.

Mitigation: We have restored the access to the KeyVault to mitigate this issue and enable the App Service to run as expected in turn bringing Azure Speech Service back to healthy state.

Next steps: We will continue to investigate to establish the full root cause and prevent future occurrences. Stay informed about Azure service issues by creating custom service health alerts: https://aka.ms/ash-videos for video tutorials and https://aka.ms/ash-alerts for how-to documentation.

April 2021

4/30

Issues accessing Azure Portal - HTTP 500-level Errors / Performance issues - Mitigated (Tracking ID 0TK3-HPZ)

Summary of Impact: Between 07:30 and 08:45 UTC on 30 Apr 2021, a subset of customers may have experienced intermittent HTTP 500 errors or general latency when trying to access the Azure Portal. There was no impact to Azure services during this time and retries to the portal may have been successful for some customers.

Preliminary Root Cause: At the start of business hours in the UK region, the Azure portal frontend endpoints in UK South began scaling up their instances to support the daily traffic. Our initial investigation show that the scaling process kicked in as expected but instances didn't serve traffic as fast as expected, leading to degraded customer experience.

Mitigation: The issue was self-healed once the new instances were able to service traffic. Even though our telemetry shows that the traffic patterns for the duration of the incident are similar to those observed during the past week, we provisioned additional instances and also increased the maximum instance count to be used for future scaling operations.

Next steps: We will continue to investigate to establish the full root cause and prevent future occurrences. Stay informed about Azure service issues by creating custom service health alerts: https://aka.ms/ash-videos for video tutorials and https://aka.ms/ash-alerts for how-to documentation.

4/20

RCA- Intermittent 503 errors accessing Azure Portal (Tracking ID HNS6-1SZ)

Summary of Impact: Between approximately 10:30 and 12:11 UTC, and again between 13:49 and 14:09 UTC on 20 Apr 2021, a subset of customers may have experienced intermittent HTTP 503 errors when trying to access the Azure Portal. There was no impact to Azure services during this time, and retires to the portal may have been successful for some customers.

Root Cause: The Azure portal frontend resources in UK South was taken out of rotation for maintenance the previous day, at 2021-04-19 19:08 UTC. For operational reasons related to an issue with that maintenance, the region was left out of rotation for a longer period than anticipated. This shifted traffic from UK South to UK West. This scenario was within acceptable operational limits, as the volume of Azure Portal traffic for that part of the world was declining at the end of the working day there.

The next day, the increase in traffic cause our instances in UK West to automatically scale-up, and it soon reached the maximum allowed number of instances, and stopped scaling up further. The running instances became overloaded, causing high CPU and disk activity, to a point where the instances became unable to process requests and began returning HTTP 503 errors.

Mitigation: At 12:11 UTC, we removed the region from the global Azure Portal rotation to restore functionality. In tandem we also scaled-out resources in other regions to ensure there was no impact related to the load rebalancing.

As part of the mitigation troubleshooting, the UK West instances were initially reimaged and retuned to rotation, as the impact from the UK south traffic was not fully understood, and thus it was believed this would resolve the issue. They were briefly brought online at 13:49 UTC, but the lack of scale caused a recurrence of the issue. UK West was taken out of rotation again at 14:09 UTC. pending a full RCA review.

Next Steps: We apologize for the impact to affected customers. We are continuously taking steps to improve the Microsoft Azure Platform and our processes to help ensure such incidents do not occur in the future. In this case, this includes (but is not limited to):

- Return UK South and UK West to rotation with increased autoscaling limits. [COMPLETED]
- Ensure autoscaling rules for adjacent regions are adjusted in the event of a region being taken out of rotation.
- Raise internal alerts to a higher severity to ensure an earlier response.
- Raise default thresholds for autoscaling to account for growth of Portal.
- Improve monitoring to take region out of rotation automatically (failures weren't consistent enough to reach the threshold for our alerts).
- Alert if a region is running at the maximum auto-scale limits.

Provide Feedback: Please help us improve the Azure customer communications experience by taking our survey: https://aka.ms/AzurePIRSurvey

4/1

RCA - DNS issue impacting multiple Microsoft services (Tracking ID GVY5-TZZ)

Summary of Impact: Between 21:21 UTC and 22:00 UTC on 1 Apr 2021, Azure DNS experienced a service availability issue. This resulted in customers being unable to resolve domain names for services they use, which resulted in intermittent failures accessing or managing Azure and Microsoft services. Due to the nature of DNS, the impact of the issue was observed across multiple regions. Recovery time varied by service, but the majority of services recovered by 22:30 UTC.

Root Cause: Azure DNS servers experienced an anomalous surge in DNS queries from across the globe targeting a set of domains hosted on Azure. Normally, Azure's layers of caches and traffic shaping would mitigate this surge. In this incident, one specific sequence of events exposed a code defect in our DNS service that reduced the efficiency of our DNS Edge caches. As our DNS service became overloaded, DNS clients began frequent retries of their requests which added workload to the DNS service. Since client retries are considered legitimate DNS traffic, this traffic was not dropped by our volumetric spike mitigation systems. This increase in traffic led to decreased availability of our DNS service.

Mitigation: The decrease in service availability triggered our monitoring systems and engaged our engineers. Our DNS services automatically recovered themselves by 22:00 UTC. This recovery time exceeded our design goal, and our engineers prepared additional serving capacity and the ability to answer DNS queries from the volumetric spike mitigation system in case further mitigation steps were needed. The majority of services were fully recovered by 22:30 UTC. Immediately after the incident, we updated the logic on the volumetric spike mitigation system to protect the DNS service from excessive retries.

Next Steps: We apologize for the impact to affected customers. We are continuously taking steps to improve the Microsoft Azure Platform and our processes to help ensure such incidents do not occur in the future. In this case, this includes (but is not limited to):

- Repair the code defect so that all requests can be efficiently handled in cache.
- Improve the automatic detection and mitigation of anomalous traffic patterns.

Provide Feedback: Please help us improve the Azure customer communications experience by taking our survey at https://aka.ms/AzurePIRsurvey .

March 2021

3/18

RCA - Azure Key Vault - Intermittent failures (Tracking ID 5LJ1-3CZ)

Summary of Impact: Between 23:00 UTC on 18 Mar 2021 and 02:15 UTC on 19 Mar 2021, a subset of customers experienced issues and/or encountered error message "InternalServerErrors" when accessing their vaults in West Europe and North Europe regions. These errors were directly impacting customers performing operations on the Control Plane or Data Plane for Azure Key Vault or for supported scenarios that used Customer Managed Keys for encryption at rest for Azure resource providers, in which case those resources were unavailable.

Timeline:

- 18 Mar 2021 23:00 UTC - First Impact Observed in West Europe
- 18 Mar 2021 23:10 UTC - West Europe Key Vault service fails over to North Europe
- 19 Mar 2021 00:00 UTC - North Europe Vaults impacted by same issue
- 19 Mar 2021 01:50 UTC - Mitigations completed by deploying new VMs. North Europe fully recovered
- 19 Mar 2021 02:15 UTC - West Europe fully recovered

Root Cause (updated 27 Apr 2021): Azure Key Vault's microservice that handles storage transactions in the West Europe region was impacted by network resource exhaustion that started at 18 Mar 2021 at 23:00 UTC. This was triggered by a surge of requests to the Data Plane for a specific type of resource which resulted in excessive operations to access the backend storage and over-utilization of network resources as a result of a code defect. The particular microservice was also under-provisioned in the West Europe and North Europe regions and had limited capacity to handle the increased load, which caused exceptionally high CPU usage. This is typically prevented by caching and service limits which will throttle the requests, but in this particular case there was a gap in our cache implementation and the lower capacity allocated to the service resulted in the incident in the West Europe region. As a result of this, the service health monitoring automatically failed over West Europe traffic to North Europe at 23:10 UTC. In North Europe, the same conditions led to the service degrading and eventually experiencing an outage on 19 Mar 2021 at 00:00 UTC.

Mitigation: As a first measure to remediate the situation, underlying Virtual Machines (VMs) supporting Azure Key Vault were rebooted. However, the CPU usage continued to be high in the VMs. Engineers then deployed new VMs with higher capacity to handle the increased CPU usage and redirected traffic to them. Once this was completed both regions recovered. Also as a preliminary measure to prevent recurrence in other regions, the capacity was reviewed and increased globally.

Next Steps (updated 27 Apr 2021): We apologize for the impact to affected customers. We are continuously taking steps to improve the Microsoft Azure Platform and our processes to help ensure such incidents do not occur in the future. In this case, this includes (but is not limited to):

- The Azure Key Vault team has immediately fixed the storage access patterns for the resource type that triggered the incident.
- Capacity has been reevaluated globally to ensure that the service is resilient to increased usage.
- New monitors have been added to watch for overly high resource usage and these monitors will trigger autoscaling.
- We are modifying the failover pattern for the service so that the paired region is not affected by failover traffic from another region.
- Constraints on resource usage and circuit breakers are being added for all down level dependencies so that the service can gracefully react to spikes in traffic and avoid extended incidents.

Provide Feedback: Please help us improve the Azure customer communications experience by taking our survey: https://aka.ms/AzurePIRSurvey

3/15

RCA - Authentication errors across multiple Microsoft services (Tracking ID LN01-P8Z)

Summary of Impact: Between 19:00 UTC on March 15, 2021 and 09:37 UTC on March 16, 2021, customers may have encountered errors performing authentication operations for any Microsoft services and third-party applications that depend on Azure Active Directory (Azure AD) for authentication. Mitigation for the Azure AD service was finalized at 21:05 UTC on 15 March 2021. A growing percentage of traffic for services then recovered. Below is a list of the major services with their extended recovery times:

22:39 UTC 15 March 2021 Azure Resource Manager.
01:00 UTC 16 March 2021 Azure Key Vault (for most regions).
01:18 UTC 16 March 2021 Azure Storage configuration update was applied to first production tenant as part of safe deployment process.
01:50 UTC 16 March 2021 Azure Portal functionality was fully restored.
04:04 UTC 16 March 2021 Azure Storage configuration change applied to most regions.
04:30 UTC 16 March 2021 the remaining Azure Key Vault regions (West US, Central US, and East US 2).
09:25 UTC 16 March 2021 Azure Storage completed their recovery and we declared the incident fully mitigated.

Root Cause and Mitigation: Azure AD utilizes keys to support the use of OpenID and other Identity standard protocols for cryptographic signing operations. As part of standard security hygiene, an automated system, on a time-based schedule, removes keys that are no longer in use. Over the last few weeks, a particular key was marked as "retain" for longer than normal to support a complex cross-cloud migration. This exposed a bug where the automation incorrectly ignored that "retain" state, leading it to remove that particular key.

Metadata about the signing keys is published by Azure AD to a global location in line with Internet Identity standard protocols. Once the public metadata was changed at 19:00 UTC on 15 March 2021, applications using these protocols with Azure AD began to pick up the new metadata and stopped trusting tokens/assertions signed with the key that was removed. At that point, end users were no longer able to access those applications.

Service telemetry identified the problem, and the engineering team was automatically engaged. At 19:35 UTC on 15 March 2021, we reverted deployment of the last backend infrastructure change that was in progress. Once the key removal operation was identified as the root cause, the key metadata was rolled back to its prior state at 21:05 UTC.

Applications then needed to pick up the rolled back metadata and refresh their caches with the correct metadata. The time to mitigate for individual applications varies due to a variety of server implementations that handle caching differently. A subset of Storage resources experienced residual impact due to cached metadata. We deployed an update to invalidate these entries and force a refresh. This process completed and mitigation for the residually impacted customers was declared at 09:37 UTC on 16 March 2021.

Azure AD is in a multi-phase effort to apply additional protections to the backend Safe Deployment Process (SDP) system to prevent a class of risks including this problem. The first phase does provide protections for adding a new key, but the remove key component is in the second phase which is scheduled to be finished by mid-year. A previous Azure AD incident occurred on September 28th, 2020 and both incidents are in the class of risks that will be prevented once the multi-phase SDP effort is completed.

Next Steps: We understand how incredibly impactful and unacceptable this incident is and apologize deeply. We are continuously taking steps to improve the Microsoft Azure platform and our processes to help ensure such incidents do not occur in the future. In the September incident, we indicated our plans to "apply additional protections to the Azure AD service backend SDP system to prevent the class of issues identified here."

- The first phase of those SDP changes is finished, and the second phase is in a very carefully staged deployment that will finish mid-year. The initial analysis does indicate that once that is fully deployed, it will prevent the type of outage that happened today, as well as the related incident in September 2020. In the meantime, additional safeguards have been added to our key removal process which will remain until the second phase of the SDP deployment is completed.
- In that September incident we also referred to our rollout of Azure AD backup authentication. That effort is progressing well. Unfortunately, it did not help in this case as it provided coverage for token issuance but did not provide coverage for token validation as that was dependent on the impacted metadata endpoint.
- During the recent outage we did communicate via Service Health for customers using Azure Active Directory, but we did not successfully communicate for all the impacted downstream services. We have assessed that we have tooling deficiencies that will be addressed to enable us to do this in the future.
- We should have kept customers more up to date with our investigations and progress. We identified some differences in detail and timing across Azure, Microsoft 365 and Dynamics 365 which caused confusion for customers using multiple Microsoft services. We have a repair item to provide greater consistency and transparency across our services.

Provide Feedback: Please help us improve the Azure customer communications experience by taking our survey at https://aka.ms/AzurePIRSurvey .

3/9

Argentina and Uruguay - Issue Accessing Azure Resources (Tracking ID 8NVQ-HD8)

Summary of Impact: Between 17:21 and 17:37 UTC on 09 Mar 2021, a network infrastructure issue occurred impacting traffic into and out of Argentina and Uruguay. During this time, customers in these areas may have experienced intermittent issues connecting to Azure resources.

Root Cause: A regional networking fiber cut resulted in a brief loss of connectivity to Microsoft resources.

Mitigation: An automated failover of network traffic to an alternative fiber route mitigated the issue.

Stay informed about Azure service issues by creating custom service health alerts: https://aka.ms/ash-videos for video tutorials and https://aka.ms/ash-alerts for how-to documentation.

February 2021

2/26

RCA - Azure Storage and dependent services - Japan East (Tracking ID PLWV-BT0)

Summary of Impact: Between 03:26 UTC and 10:02 UTC on 26 Feb 2021, a subset of customers in Japan East may have experienced service degradation and increased latency for resources utilizing Azure Storage, including failure of virtual machine disks. Some Azure services utilizing Storage may have also experienced downstream impact.

Summary Root Cause: During this incident, the impacted storage scale unit was under heavier than normal utilization. This was due to:

- Incorrect limits set on the scale unit which allowed more load than desirable to be placed on it. This reduced the headroom that is usually available for unexpected events such as sudden spikes in growth which allows time to take load-balancing actions.
- Additionally, the load balancing automation was not sufficiently spreading the load to other scale units within the region.

This high utilization triggered heavy throttling of storage operations to protect the scale unit from catastrophic failures. This throttling resulted in failures or increased latencies for storage operations on the scale unit.

Note: The original RCA mistakenly identified a deployment as a triggering event for the increased load. This is because during an upgrade, the nodes to be upgraded are removed from rotation, temporarily increasing load on remaining nodes. An upgrade was in queue on the scale unit but had not yet started. Our apologies for the initial mistake.

Background: An internal automated load balancing system actively monitors resource utilization of storage scale units to optimize load across scale units within an Azure region. For example, resources such as disk space, CPU, memory and network bandwidth are targeted for balancing. During this load balancing, storage data is migrated to a new scale unit, validated for data integrity at the destination and finally the data is cleaned up on the source to return free resources. This automated load-balancing happens continuously and in real-time to ensure workloads are properly optimized across available resources.

Detailed Root Cause: Prior to the start of impact, our automated load-balancing system had detected high utilization on the scale-unit and was performing data migrations to balance the load. Some of these load-balancing migrations did not make sufficient progress, creating a situation where the resource utilization on the scale unit reached levels that were above the safe thresholds that we try to maintain for sustained production operation. This kick-started automated throttling on incoming storage write requests to protect the scale unit from catastrophic failures. When our engineers were engaged, they also detected that the utilization limits that were set on the scale unit to control how much data and traffic should be directed to the scale unit was higher than expected. This did not give us sufficient headroom to complete load-balancing actions to prevent customer facing impact.

Mitigation: To mitigate customer impact as fast as possible, we took the following actions:

- Engineers took steps to aggressively balance resource load out of the storage scale unit. The load-balancing migrations that were previously unable to finish were manually unblocked and completed, allowing a sizeable quantity of resources to be freed up for use. Additionally, load-balancing operations were tuned to improve its throughput to more effectively distribute load.
- We prioritized recovery of nodes with hardware failures that had been taken out of rotation to bring additional resources online.

These actions brought the resource utilization on the scale unit to a safe level which was well below throttling thresholds. Once Storage services were recovered around 06:56 UTC, dependent services started recovering. We declared full mitigation at 10:02 UTC.

Next steps: We sincerely apologize for the impact this event had on our customers. Next steps include but are not limited to:

- Optimize the maximum allowed resource utilization levels on this scale unit to provide increased headroom in the face of multiple unexpected events.
- Improve existing detection and alerting for cases when load-balancing is not keeping up, so corrective action can be triggered early to help avoid customer impact.
- Improve load-balancing automation to handle certain edge-cases under resource pressure where manual intervention is currently required to help prevent impactful events.
- Improve emergency-levers to allow for faster mitigation of impactful resource utilization related events.

Provide Feedback: Please help us improve the Azure customer communications experience by taking our survey: https://aka.ms/AzurePIRSurvey

2/16

Azure Frontdoor - Europe - Timeouts connecting to resources (Tracking ID ZN8_-VT8)

Summary of Impact: Between approximately 12:00 UTC and 13:30 UTC a subset of customers using Azure Frontdoor in Europe may have experience timeouts and/or issues connecting to resources.

Root Cause: Engineers determined that a backend network device became unhealthy, and traffic was not automatically rerouted. This resulted in Azure Front Door requests to fail.

Mitigation: We manually removed the faulty backend network device and rerouted network traffic. This mitigated the issue.

Stay informed about Azure service issues by creating custom service health alerts: https://aka.ms/ash-videos for video tutorials and https://aka.ms/ash-alerts for how-to documentation.

◂ 1 ... 3 4 5 ... 9 ▸