Azure status history

Product: All

Region: All

Date: Most recent

August 2019

8/28 Azure Email Alerts - Mitigated

Summary of impact: Between 20:22 UTC on 28 Aug 2019 and 01:35 UTC on 29 Aug 2019, a subset of customers may not have received email alerts sent from a subset of Azure services during the impact window.

Preliminary root cause: Engineers determined that instances of a backend email processing service became unhealthy during platform maintenance. This prevented requests from completing successfully, causing impact to downstream services.

Mitigation: Engineers halted the maintenance and validated the service health to mitigate the issue.

Next steps: Engineers 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.

8/23 RCA - Service Management Operations - West/North Europe

Summary of Impact: Between 06:20 and 10:46 UTC on 23 Aug 2019, a subset of customers in North Europe and West Europe may have received failure notifications when performing service management operations - such as create, update, delete - for resources hosted in these regions.

Root Cause: During this incident, the memory consumption on Azure Resource Manager's (ARM) worker roles exceeded operational thresholds. This should have initiated a clearing process which attempts to recover additional memory when thresholds are reached. However, the memory cache was unrecoverable at the time given that it was referenced by other (active) objects in the system. CPU utilization on the affected worker roles also increased, which in turn prevented certain processes from completing.

Analysis has indicated that all threads on the affected worker roles were busy during the impact window, thus manifesting in impact to service management operations to ARM-dependent services and/or resources.

In addition, worker roles recycle on a weekly cadence which, in this scenario, further contributed to thresholds being reached. The nature of the issue required manual intervention to fully mitigate, which further delayed mitigation.

Mitigation: Engineers performed a manual restart of the affected worker roles to mitigate the issue.

Next Steps: We sincerely 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 included (but was not limited to):

- Automatically recycling worker roles on a more frequent cadence [In Progress]
- Separating distinct cache types [In Progress]
- Adding enhanced logging information to understand memory cache usage in worker roles [In Progress]
- Reducing cache footprint [In Progress]

Provide Feedback: Please help us improve the Azure customer communications experience by taking our survey https://aka.ms/HT4G-PC0.

8/16 RCA - Azure Services - South Central US

Summary of impact: Between 14:16 UTC and 21:36 UTC on 16 Aug 2019, a subset of customers using resources in South Central US may have experienced difficulties connecting to resources hosted in this region.

Root cause: An outage occurred during a regional infrastructure update, which is part of a South Central US datacenter refresh initiative. At approximately 14:00 UTC on 8/16/2019, facility engineers proceeded with removal of infrastructure following a pre-approved method of procedure (MOP). However, an error in fiber trunk validation resulted in production impacting connectivity loss.
The loss of these physical links resulted in connectivity issues with a scale unit in the datacenter that hosted several Azure infrastructure services.

Mitigation: To mitigate, services with multi-region deployments successfully failed out of the region to minimize impact. Residual impact was mitigated by restoring full fiber connectivity to the affected scale unit and placing it back into production rotation.

Next steps: We sincerely 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):

- Halt infrastructure refresh projects globally until below repair items are addressed - Complete
- Method of Procedure (MoP) detailed review as related to infrastructure refresh projects – in progress
- Improve existing physical infrastructure change management process with additional collaboration & validation between facility and infra engineering - in progress.

Provide feedback: Please help us improve the Azure customer communications experience by taking our survey https://aka.ms/KT5G-BC8

8/14 RCA - USGov Iowa - Service Availability and Downstream Impact

Summary of impact: Between 16:25 EDT on 14 Aug 2019 and 01:00 EDT on 15 Aug 2019, a subset of customers in USGov Iowa may have experienced issues accessing services in that region. Engineers determined that an underlying compute issue was impacting SQL, API Management, Azure Site Recovery, Media Services, Service Bus, and StorSimple.

A subset of App Service customers may have experienced Service Management issues across these regions: USDoD Central, USDoD East, USGov Arizona, USGov Iowa, USGov Texas, and USGov Virginia.

Root cause and mitigation: Automated health monitoring is utilized in all Azure regions to predict and react to telemetry signals that indicate that a failure condition has occurred on a resource. In this instance, health monitoring indicated that several physical host nodes in the scale unit had reached a threshold for high memory utilization, which resulted in those host nodes moving to a state for repair. An investigation is ongoing into the specific cause of the increased memory footprint on these hosts, but as this is the expected behavior for the platform health monitoring, this was functioning as expected. A secondary issue was detected on review that was causing the build-up of the number of nodes requiring non-automated or manual repair. When nodes are taken out of rotation and put into repair, automation processes the node and will perform needed tasks to achieve repair and return the node to service. In this case, the automation was triggering and was putting the nodes into a manual repair mode, waiting for human intervention. With the increased number of nodes failing and requiring human intervention, this impacted the Service Fabric seed nodes of Azure SQL DB in this stamp. Quorum was lost on seed nodes and databases hosted on this instance of Azure SQL DB became unavailable. In parallel, SQL DB and Compute engineering teams worked to restore these services and nodes in the scale unit. Manually applying the repair to these nodes was successful, after which, the cluster fabric and SQL DB services were able to recover, mitigating the issue.

After an initial recovery for the SQL DB instance, additional node recovery efforts mistakenly again impacted one of the Service Fabric seed nodes leading to Quorum loss on the seed nodes. This was due to a method of recovery for the systems that were requiring repair. A repair item has been created to address this conflict in recovery. Engineers manually recovered the nodes and brought them back online to restore connectivity to the dependent services, which in turn mitigated the downstream impact once the dependent services were restored.

Next steps: We sincerely 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):

- Improve the automation repair process to track the number of nodes required to operate the Azure SQL DB service and alert when the number of available nodes comes down below the safe operational thresholds
- Improve alerting from the Service Fabric layer quorum failures to upstream services when failures rate exceeds normal safe operational thresholds

Provide feedback: Please help us improve Azure customer communications experience by taking our survey - https://aka.ms/HLR7-N98

8/9 RCA - Azure Portal - Issues Loading Dashboard Tiles

Summary of impact: Between approximately 16:07 and 19:17 UTC on 09 Aug 2019, customers using the Azure portal may have received failure notifications under the following scenarios:

- Customers who had pinned tiles to their dashboards that provided them with a list of resources such as All Resources, All VMs, etc. would have either seen a "No resources found" message or just a grey tile. All other tiles on dashboards such as pinned resources continued to function as expected.
- Access of the following two pages on the Azure Portal. a) "My permissions" page that enables customers to see what permissions they have in a given subscription and b) "Resource Providers" option in the resource menu for any given subscription that lets customers view, register and unregister Resource Providers in their subscriptions. Trying to open one of these two pages would have resulted in the page load timing out after 30 seconds.

Root cause: The issue was caused by a code change that introduced a bug in an underlying API that the affected experiences relied on. The Azure Portal relies on feature flags to enable or disable features and the above regression would occur only when one such flag was set to the disabled state. While engineering had tests to cover these scenarios, those tests were only run with the flag set to enabled which was not the flag state we had in production.

Mitigation: Portal engineering received alerts from our monitoring and alerting infrastructure. A roll back to the previous production build that did not contain this regression was initiated. This roll back was completed in multiple regions by 09 Aug 2019 19:09 UTC and affected experiences were back to working as expected at this point of time.

Next steps: We sincerely 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):

- Added new tests that runs with all possible states of the particular feature flag that resulted in this incident to ensure that a similar regression is caught before changes get deployed to production.

Provide feedback: Please help us improve the Azure customer communications experience by taking our survey https://aka.ms/LTJ7-JP0

8/8 RCA - Azure CDN - Connectivity Issues

Summary of impact: Between 14:05 and 19:00 UTC on 08 Aug 2019, you were identified as a customer using Akamai CDN who may have experienced HTTP 504 (Gateway Timeout) errors when attempting to connect to resources.

Root cause: Engineers determined that a network update to add new peer servers to the US East region caused connections to fail between Akamai CDN and Azure resources. A network filtering policy on a redundant Microsoft peering device was improperly configured, which caused traffic sourced from Akamai CDN and destined for origins hosted on Azure storage in East US and East US 2 to be intermittently dropped.

Mitigation: Engineers rolled back the updates to mitigate the issue. After the root cause was identified and the network filtering policy was updated and consistent across the pair of devices, traffic was successfully routed over this path without drops.

Next steps: We sincerely 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):

- Inventory device fleet to ensure no further inconsistencies exist for network filtering policies
- Update device configuration drift platform to account network filtering policy inconsistencies across device pairs
- Add additional automation and checks to update network filtering policies across device fleet to remove the risk of inconsistencies

Provide feedback: Please help us improve the Azure customer communications experience by taking our survey https://aka.ms/KK63-J98

July 2019

7/30 RCA - Issues logging in to the Azure Portal with a Microsoft Account

Summary of Impact: Between 21:00 UTC and 22:24 UTC on 30 Jul 2019, a subset of customers may have experienced intermittent error messages and failures when logging in to the Azure Portal with a Microsoft account. This issue affected a subset of MSA users who could not authenticate or manage their accounts. Retries may have been successful.

Root Cause: Engineers were performing standard maintenance on a standby module when the active module became unstable. It was determined that combination of factors including a device mismatch and code bug, resulted in the active device becoming unstable.

Mitigation: Engineers routed the user traffic to other redundant paths to recover the service, which restored Microsoft Account services for customers.

Next Steps: We sincerely 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):

- Updates to processes around manageability of all devices.
- Improve and increase the redundancy of the supporting infrastructure, including the manageability of failed devices.

Provide Feedback: Please help us improve the Azure customer communications experience by taking our survey https://aka.ms/HVZ7-JP8

7/8 RCA - Azure Active Directory - Password Changes

Summary of impact: Between 18:09 and 22:32 UTC on 08 Jul 2019, a subset of customers using Azure Active Directory may have experienced password change issues. For hybrid customers, passwords would have appeared to have changed successfully on-prem, but the sync with the backend AAD would have failed. For cloud only customers, password changes with AAD would have resulted in failures. In either case, customers may not have been able to log into AAD or the Azure portal. The following link was a workaround to reset passwords: https://passwordreset.microsoftonline.com

Root cause: Engineers determined that a recent deployment of a certificate caused backend instances responsible for password change functionality to reject the password change operation. The certificate deployment happened according to a schedule, however due to a bug, backend instances were not able to load the new certificate. In addition, the certificate rotation had an unrelated bug that increased the exposure of the issue to a larger number of customers than intended.

Mitigation: Engineers rolled back the recent deployment task and restarted associated frontend instances to mitigate the issue.

Next steps: We sincerely 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):
- Improved automation and test coverage for rotation of the particular type of certificate to prevent re-occurrence of the same issue
- Verification of all certificate rotation procedures in AAD to eliminate exposure control issues

7/4 RCA - Connectivity Issues - UK South

Summary of impact: Between 06:18 UTC and 16:25 UTC on 04 July 2019, a subset of customers leveraging Storage in UK South may have experienced service availability issues. In addition, resources with dependencies on Storage may also have experienced downstream impact in the form of availability issues. These impacted services include App Services, Virtual Machines, Backup, KeyVault, App Insights, Logic Apps, and Azure Data Factory v2.

Root cause: During the last two weeks of June, the customer traffic on a single storage scale unit in the UK south region had increased more quickly than normal, and the automatic load balancing system had responded by beginning to shift load to other scale units. However, on 06:18 UTC on July 4th, 2019, the increased levels of resource utilization pushed the scale unit above its natural operating limits, which resulted in higher latency and failures. Services dependent on the storage scale unit experienced a high number of failures and latency which manifested in availability issues.

Mitigation: To recover the scale unit to healthy state, engineers applied following mitigation steps to reduce the resource utilization levels:

- Load balancing configuration changes
- Reducing internal background processes in the scale unit

Next steps: We sincerely 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):

- Enhance existing load balancing and monitoring mechanism to maintain sustainable resource utilization at the storage scale unit level
- Improve service response to high resource utilization levels
- Improve automatic load balancing efficiency to respond more quickly to an unusual increase in traffic

Provide feedback: Please help us improve the Azure customer communications experience by taking our survey https://aka.ms/HMD0-LP0

7/3 Azure Monitoring (Diagnostic logs, Autoscale, Classic Alerts (v2))

Summary of impact: Between 02:00 and 08:00 UTC on 03 Jul 2019, a subset of Azure monitor customers may have received failure errors while performing service management operations - such as create, update, delete - for Autoscale settings, Classic alerts and Diagnostics settings. Existing autoscale, classic alerts & diagnostics were not impacted.

Root cause: As part of service engineering improvements, a configuration change was incorrectly applied to the underlying KeyVault resources which are used by Azure Monitor's management services (Autoscale, Classic Alerts, Diagnostic Settings) during start up. Subsequently when the management services recycled, the services were not able to correctly access the KeyVault resources due to the misconfiguration and hence could not startup correctly.

Mitigation: Engineers performed a roll back of misconfiguration to mitigate the issue.

Next steps: We sincerely 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:

- Review and improve configuration change, testing & rollout process to prevent a reoccurrence.

7/2 Azure Services - Intermittent Service Availability Issues

Summary of Impact: Between 19:20 and 22:20 UTC on 02 Jul 2019, a subset of customers using Microsoft Azure Services may have intermittently experienced degraded performance, latency, network drops or time outs when accessing Azure resources due to a network event. This impact would have potentially spanned multiple Azure services. During the impact window, traffic peering through San Jose route would have been impacted.

Root Cause: One of the network devices in San Jose had a hardware grey failure at 19:20 UTC, causing traffic going to one of Microsoft peer networks to experience intermittent failures. This partial failure caused intermittent connectivity issues to services peered through San Jose for a subset of customers connecting through this faulty peer.

Mitigation: Engineering localized and isolated the faulty router and took the device out of service at 20:15 UTC. The traffic engineering systems immediately redirected the traffic and services started to recover. The additional load caused by the redirection out of the device was spread across multiple devices to ensure optimal latency and throughput to customers after the failure. At 22:20 UTC all traffic optimizations were completed by the automation systems, mitigating the impact.

Next Steps: We sincerely 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:

- Replaced the faulty route. [COMPLETED]
- Enhanced monitoring and alerting to detect hardware grey failure. [COMPLETED]
- Improve auto-mitigation logic to quickly redirect traffic across multiple devices to avoid future instances of issue. [COMPLETED]
- Continuing the investigation into the network router with the vendor.

June 2019

6/14 RCA - Virtual Machines and Virtual Networks (Management) - North Europe

Summary of impact: Between 15:45 and 20:58 UTC on 14 Jun 2019, a subset of customers using Virtual Machines, Virtual Machine Scale Sets, and/or Virtual Networks in the North Europe region may have experienced failures when attempting service management operations including create, update, delete, and scale. Azure Cloud Shell customers may have seen the error message "Failed to provision a Cloud Shell."

Root cause: The Fabric Controllers are the services that control the allocation of resources in an Azure scale unit. The Fabric Controllers are reached via a Virtual IP Address (VIP) implemented by a Software Loading Balancing (SLB) system. In order to function properly, the Software Load Balancing servers must be configured with the identities of the network switches to which they are attached so that they can create connections to the network switches and send the switches information on how to reach the VIPs.

During a configuration change to decommission old equipment in one Availability Zone, the switches attached to the active SLB servers were incorrectly marked as no longer part of their active scale unit, and the SLBs ceased to connect with the network switches as designed. This resulted in loss of connectivity to the Fabric Controllers in that Availability Zone, which blocked the ability to allocate, deallocate, or change resources in that Availability Zone.

There was no impact to VIPs or Load Balancing services for Azure tenants during this incident, as the SLB system that provides connectivity to the Fabric Controller VIPs is separate from the one that provides VIPs and Load Balancing to Azure tenants.

The rollout of the bad configuration information to the SLB servers was not automatically stopped and rolled back by the Azure Deployment system as the SLB monitoring and alerting components did not determine this was an incorrect and undesirable configuration.

Mitigation: Azure Engineers identified and rolled back the incorrect configuration.

It should be noted that resilience options existed that would have made this issue non-impactful for customers. Specifically, this issue affected one Availability Zone in the region. Services that leverage multiple Availability Zones as part of their service availability design would have been unaffected during this incident. Customers interested in learning more regarding resilience options should visit:

https://azure.microsoft.com/en-us/features/resiliency/

Next steps: We apologize for the impact this incident may have caused you and your services. Among the repair actions being taken to prevent recurrence are the following:

- Verify that no other changes have been made to network metadata that incorrectly removed a switch from an active scale unit.
- Improve the resilience of the VIP in front of the Fabric Controllers by distributing responsibility for it across multiple SLB instances.
- Improvement to the SLB system monitoring components so that any configuration that causes SLB components to no longer connect to network switches is considered an error and rolled back.
- Implement validations on the network-metadata so that switches cannot be removed from membership in an active scale unit without additional review and approval steps.

Please help us improve the Azure customer communications experience by taking our survey https://aka.ms/F5_P-TVG