Azure status history

Product:

All

Region:

All

Date:

Most recent

September 2018

9/4

South Central US - Preliminary RCA

The following material is intended to be an overview and act as a preliminary root cause analysis (RCA) for the South Central US incident beginning 4 September 2018.

The Azure engineering teams are continuing with a detailed investigation over the coming weeks to identify resiliency gaps and areas of improvement for Azure services. A detailed analysis will be made available in the weeks ahead.

Summary of impact: In the early morning of September 4, 2018, high energy storms hit southern Texas in the vicinity of Microsoft Azure's South Central US region. Multiple Azure datacenters in the region saw voltage sags and swells across the utility feeds. At 08:42 UTC, lightning caused electrical activity on the utility supply, which caused significant voltage swells. These swells triggered a portion of one Azure datacenter to transfer from utility power to generator power. Additionally, these power swells shutdown the datacenter's mechanical cooling systems despite having surge suppressors in place. Initially, the datacenter was able to maintain its operational temperatures through a load dependent thermal buffer that was designed within the cooling system. However, once this thermal buffer was depleted the datacenter temperature exceeded safe operational thresholds, and an automated shutdown of devices was initiated. This shutdown mechanism is intended to preserve infrastructure and data integrity, but in this instance, temperatures increased so quickly in parts of the datacenter that some hardware was damaged before it could shut down. A significant number of storage servers were damaged, as well as a small number of network devices and power units.
While storms were still active in the area, onsite teams took a series of actions to prevent further damage – including transferring the rest of the datacenter to generators thereby stabilizing the power supply. To initiate the recovery of infrastructure, the first step was to recover the Azure Software Load Balancers (SLBs) for storage scale units. SLB services are critical in the Azure networking stack, managing the routing of both customer and platform service traffic. The second step was to recover the storage servers and the data on these servers. This involved replacing failed infrastructure components, migrating customer data from the damaged servers to healthy servers, and validating that none of the recovered data was corrupted. This process took time due to the number of servers damaged, and the need to work carefully to maintain customer data integrity above all else. The decision was made to work towards recovery of data and not fail over to another datacenter, since a fail over would have resulted in limited data loss due to the asynchronous nature of geo replication.
Despite onsite redundancies, there are scenarios in which a datacenter cooling failure can impact customer workloads in the affected datacenter. Unfortunately, this particular set of issues also caused a cascading impact to services outside of the region, as described below.

Customer impact:
[1] Impact to resources in South Central US
Customer impact began at approximately 09:29 UTC when storage servers in the datacenter began shutting down as a result of unsafe temperatures. This impacted multiple Azure services that depended on these storage servers, specifically: Storage, Virtual Machines, Application Insights, Cognitive Services & Custom Vision API, Backup, App Service (and App Services for Linux and Web App for Containers), Azure Database for MySQL, SQL Database, Azure Automation, Site Recovery, Redis Cache, Cosmos DB, Stream Analytics, Media Services, Azure Resource Manager, Azure VPN gateways, PostgreSQL, Application Insights, Azure Machine Learning Studio, Azure Search, Data Factory, HDInsight, IoT Hub, Analysis Services, Key Vault, Log Analytics, Azure Monitor, Azure Scheduler, Logic Apps, Databricks, ExpressRoute, Container Registry, Application Gateway, Service Bus, Event Hub, Azure Portal IaaS Experiences- Bot

Service, Azure Batch, Service Fabric and Visual Studio Team Services (VSTS). Although the vast majority of these services were mitigated by 11:00 UTC on 5 September, full mitigation was not until approximately 08:40 UTC on 7 September.

[2] Impact to Azure Service Manager (ASM)
Insufficient resiliency for Azure Service Manager (ASM) led to the widest impact for customers outside of South Central US. ASM performs management operations for all 'classic' resource types. This is often used by customers who have not yet adopted Azure Resource Manager (ARM) APIs which have been made available in the past several years. ARM provides reliable, global resiliency by storing data in every Azure region. ASM stores metadata about these resources in multiple locations, but the primary site is South Central US. Although ASM is a global service, it does not support automatic failover. As a result of the impact in South Central US, ASM requests experienced higher call latencies, timeouts or failures when performing service management operations. To alleviate failures, engineers routed traffic to a secondary site. This helped to provide temporary relief, but the impact was fully mitigated once the associated South Central US storage servers were brought back online at 01:10 UTC on 5 September

[3] Impact to Azure Active Directory (AAD)
Customer impact began at 4 September at 11:00 UTC. One of the AAD sites for North America is based in the South Central US region, specifically in the affected datacenter. The design for AAD includes globally distributed sites for high availability so, when infrastructure began to shutdown in the impacted datacenter, authentication traffic began automatically routing to other sites. Shortly thereafter, there was a significantly increased rate in authentication requests. Our automatic throttling mechanisms engaged, so some customers continued to experience high latencies and timeouts. Adjustments to traffic routing, limited IP address blocking, and increased capacity in alternate sites improved overall system responsiveness – restoring service levels. Customer impact was fully mitigated at 14:40 UTC on 4 September

[4] Impact to Visual Studio Team Services (VSTS)
VSTS organizations hosted in South Central US were down. Some VSTS services in this region provide capabilities used by services in other regions, which led to broader impact – slowdowns, errors in the VSTS Dashboard functionality, and inability to access user profiles stored in South Central US to name a few. Customers with organizations hosted in the US were unable to use Release Management and Package Management services. Build and release pipelines using the Hosted macOS queue failed. To avoid data loss, VSTS services did not failover and waited for the recovery of the Storage services. After VSTS services had recovered, additional issues for customers occurred in Git, Release Management, and some Package Management feeds due to Azure Storage accounts that had an extended recovery. This VSTS impact was fully mitigated by 00:05 UTC on 6 September. More information can be found at https://aka.ms/VSTS-CPD1PLG (https://aka.ms/VSTS-CPD1PLG).

[5] Impact to Azure Application Insights
Application Insights resources across multiple regions experienced impact. This was caused by a dependency on Azure Active Directory and platform services that provide data routing. This impacted the

ability to query data, to update/manage some types of resources such as Availability Tests, and significantly delayed ingestion. Engineers scaled out the services to more quickly process the backlog of data and recover the service. During recovery, customers experienced gaps in data, as seen in the Azure portal; Log Search alerts firing, based on latent ingested data; latency in reporting billing data to Azure commerce; and delays in results of Availability Tests. Although 90% of customers were mitigated by 16:16 UTC on 6 September, full mitigation was not until 22:12 UTC on 7 September.

[6] Impact to the Azure status page
The Azure status page (status.azure.com) is a web app that uses multiple Azure services in multiple Azure regions. During the incident, the status page received a significant increase in traffic, which should have caused a scale out of resources as needed. The combination of the increased traffic and incorrect auto-scale configuration settings prevented the web app from scaling properly to handle the increased load. This resulted in intermittent 500 errors for a subset of page visitors, starting at approximately 12:30 UTC. Once these configuration issues were identified, our engineers adjusted the auto-scale settings, so the web app could expand organically to support the traffic. The impact to the status page was fully mitigated by 23:05 UTC on 4 September.

[7] Impact to Azure subscription management
From 09:30 UTC on 4 September, Azure subscription management experienced five separate issues related to the South Central US datacenter impact. First, provisioning of new subscriptions experienced long delays as requests were queued for processing. Second, the properties of existing subscriptions could not be updated. Third, subscription metadata could not be accessed, and as a result billing portal requests failed. Fourth, customers may have received errors in the Azure management and/or billing portals when attempting to create support tickets or contacting support. Fifth, a subset of customers were incorrectly shown a banner asking that they contact support to discuss a billing system issue. Impact to subscription management was fully mitigated by 08:30 UTC on 5 September.

Next steps: We sincerely apologize for the impact to affected customers. Investigations have started to assess the best ways to improve architectural resiliency and mitigate the issues that contributed to this incident and its widespread impact. While this is our preliminary analysis and more details will be made available in the weeks ahead, there are several workstreams that are immediately underway. These include (but are not limited to) the following themes which we understand to be the biggest contributing factors to this incident:
1. A detailed forensic analysis of the impacted datacenter hardware and systems, in addition to a thorough review of the datacenter recovery procedures.
2. A review with every internal service to identify dependencies on the Azure Service Manager (ASM) API. We are exploring migration options to move these services from ASM to the newer ARM architecture.
3. An evaluation of the future hardware design of storage scale units to increase resiliency to environmental factors. In addition, for scenarios in which impact is unavoidable, we are determining software changes to automate and accelerate recovery.
4. Impacted customers will receive a credit pursuant to the Microsoft Azure Service Level Agreement, in their October billing statement.

Provide feedback: Please help us improve the Azure customer communications experience by taking our survey https://aka.ms/CPD-1PLG (https://aka.ms/CPD-1PLG)

9/6

Azure Active Directory - Multiple Regions

Summary of impact: Between 16:13 UTC on 06 Sep 2018 and 01:35 UTC on 07 Sep 2018, a subset of Azure Active Directory customers using multiple services may have experienced intermittent authentication request failures when connecting to resources.

Preliminary root cause: Engineers determined that a recent deployment task impacted instances of a backend service which became unhealthy, preventing requests from completing.

Mitigation: Engineers deployed a platform hotfix in order to mitigate the issue.

Next steps: Engineers will continue to investigate to establish the full root cause and prevent future occurrences.

August 2018

8/20

Iot Hub - West Europe and North Europe

Summary of impact: Between 17:30 UTC on 20 Aug 2018 and 16:42 UTC on 21 Aug 2018, a subset of IoT Hub customers in West Europe may experience errors, such as 'InternalServerError', when trying to access IoT Hub devices hosted in these regions. Customers may also experience issues when attempting to delete resources/devices.

Preliminary root cause: Engineers determined that multiple backend service nodes were stuck in a unhealthy state, preventing failover from occurring.

Mitigation: Engineers performed a manual reboot of backend service nodes to mitigate the issue.

Next steps: Engineers will continue to investigate and monitor to identify why the failover was unsuccessful.

8/20

RCA - Storage - West Europe

Summary of impact:
Between 08:40 and 13:43 UTC on 20 Aug, 2018, a subset of customers using Azure Storage may have been unable to create ARM storage accounts in West Europe. Customers may have received 400 HTTP Bad Request responses with an error code of "RequestIsNotValidJson" even for valid, well-formed requests. Customers attempting to deploy a new VM in the affected regions may have also encountered failures due to this issue. Other services and operations which rely on the creation of storage accounts would have similarly been impacted. Other storage service management operations were not impacted. Reading and writing data within existing storage accounts was also not affected.

Customer impact workaround:
Customers may have chosen to use an existing account or create a new account in an unaffected region or create classic storage accounts.

Root cause and mitigation:
The incident was caused by a bug in the Storage Resource Provider (SRP). The bug was triggered in the create account path when we were evaluating the scale unit to create the account in. The bug only manifests under specific configuration settings which was only present in West Europe. Due to the bug, customers would have experienced account creation failures. The bug was not found during testing, and the service had been deployed to other production regions without issue before encountering the incident in West Europe.
Engineers mitigated the issue by making a configuration update.

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.
1) We are working on a fix to make the account creation more robust in the event of a similar bug should they re-occur in the future [in progress].
2) We are also working on completing implementation of per-region flighting of new deployments for Storage with automatic rollback whenever core scenarios are impacted [in progress].

Provide feedback: Please help us improve the Azure customer communications experience by taking our survey https://aka.ms/G4XN-Y_Z (https://aka.ms/G4XN-Y_Z)

8/13

RCA - Latency and Slow I/O issues in East US

Summary of impact: Between 18:45 and 21:35 UTC on 13 Aug 2018, a subset of customers in East US would have experienced some of the below issues with their Azure services:
• Latency and potentially connectivity loss to Virtual Machines and Storage accounts
• Latency or failed requests for traffic going over an Express Route circuit
• Latency or failures when attempting to Remote Desktop Protocol (RDP) into their Virtual Machines.

Customer impact: This incident impacted a single region. Services deployed with geographic redundancy may have been able to operate out of other regions.
For services located in East US, retrying requests or failed operations on a new TCP or UDP connection would have had a 93% chance of success.

Root cause and mitigation: This incident resulted from one network router that carries traffic between data centers in the East US region being put back into service without sufficient uplink capacity available to handle its share of the load. There are multiple routers filling this role in the East US region, and the correct behavior would have been to leave this network router out of service until it had the amount of uplink capacity required. Once engineers discovered the congested link, the network router was removed from service and network operations returned to normal.

Next steps: Azure Networking understands that this issue caused significant impact to our customers, and we are taking the following steps to prevent a recurrence:
1. Reinstate alerting for high utilization links. [Completed]
2. Create a system that continually verifies that critical alerts, like high utilization alarms, are functioning properly. [In Planning]
3. Create a separate alert to warn on-call engineers about links with high rates of packet discard. [In Progress]
4. Extend the faulty link repair system to perform the health checks for the engineer, removing human action and judgement from the process. [In Progress]
5. Update the troubleshooting guide to unambiguously specify how to check that there is sufficient capacity available. [Completed]

Provide feedback: Please help us improve the Azure customer communications experience by taking our survey https://aka.ms/SNBD-MT0 (https://aka.ms/SNBD-MT0)

8/1

Azure Service Management Failures

Customer Impact: Between 21:14 UTC on 1st Aug and 00:00 UTC 2nd Aug 2018, a subset of customers may have experienced timeouts or failures when attempting to perform service management operations for Azure Resources dependent on management.core.windows.net, this is Azure's v1 API endpoint. In addition, some customers may have experienced intermittent latency and connection failures to their Classic Cloud Service resources.

Prelimary Root Cause and Mitigation: Underlying Storage experienced larger than normal loads by our internal services, engineers manually reduced the load to mitigate the issue.

8/1

RCA - Networking Connectivity and Latency - East US

Summary of impact: Between 13:18 UTC and 22:40 UTC on 01 Aug 2018, a subset of customers in the East US region may have experienced difficulties or increased latency when connecting to some resources.

Customer impact: This incident affected a single region, East US, and services deployed with geo-redundancy in multiple regions would not have been impacted.

Root cause and mitigation: A fiber cut caused by construction approximately 5 km from Microsoft data centers resulted in multiple line breaks, impacting separate splicing enclosures that reduced capacity between 2 Azure data centers in the East US region. As traffic ramped up over the course of the day, congestion occurred and caused packet loss, particularly to services hosted in or dependent on one data center in the region.

Microsoft engineers reduced the impact to Azure customers by reducing internal and discretionary traffic to free up network capacity. Microsoft had been in preparations to add capacity between data centers in the region, and it completed the mitigation of the issue by pulling in the activation of the additional capacity. Repair of the damaged fiber continued after the mitigation, and it will be returned to service when fully ready.

Next steps:
Microsoft regrets the impact to our customers and their services. We design our networks to survive the

failure of any single component, however in this instance there was not enough capacity left after the failure to handle all offered traffic. To prevent recurrence, we are taking the following steps:

1) Reducing the utilization threshold for capacity augmentation - in progress.
2) Improving our systems for automatically reducing discretionary traffic during congestion events to preserve performance for customer traffic - in progress.
3) Improving our ability to identify high-rate flows that can cause impact to other traffic during congestion events and create appropriate ways to handle this traffic - in progress.

Provide feedback: Please help us improve the Azure customer communications experience by taking our survey https://aka.ms/CL78-G88 (https://aka.ms/CL78-G88)

July 2018

7/31

RCA - Storage - North Central US

Summary of impact: Between 17:52 UTC and 18:40 UTC on 31 Jul 2018 a subset of customers using Storage in North Central US may have experienced difficulties connecting to resources hosted in this region. Other services that leverage Storage in this region may also have been experiencing impact related to this.

Root cause and mitigation: During a planned power maintenance activity, a breaker tripped transferring the IT load to its second source. A subset of the devices saw a second breaker trip resulting in the restart of a subset of Storage nodes in one of the availability zones in North Central region. Maintenance was completed with no further issues or impact to services. A majority of Storage resources self-healed. Engineers manually recovered the remaining unhealthy Storage nodes. Additional manual mitigating actions were performed by Storage engineers to fully mitigate the incident.

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):

1. Engineers have sent failed components to the OEM manufacturer for further testing and analysis [in progress]

Provide feedback: Please help us improve the Azure customer communications experience by taking our survey https://aka.ms/JKZZ-41Z (https://aka.ms/JKZZ-41Z)

7/28

App Service - East US - Mitigated

Summary of impact: Between 08:00 and 12:30 UTC on 28 Jul 2018, a subset of customers using App Services in East US may have received HTTP 500-level response codes, have experienced timeouts or high latency when accessing App Service (Web, Mobile and API Apps) deployments hosted in this region.

Preliminary root cause: Engineers identified an internal configuration conflict as part of an ongoing deployment as the potential root cause.

Mitigation: Engineers made a manual adjustment of the deployment configuration to mitigate the issue.

Next steps: Engineers will continue to investigate to establish the full root cause and prevent future occurrences.

7/26

RCA - Azure Service Management Failures

Summary of impact: Between 22:15 on 26 Jul 2018 and 06:20 UTC on 27 Jul 2018, a subset of customers may have experienced timeouts or failures when attempting to perform service management operations for Azure Resources dependent on management.core.windows.net, this is Azure's v1 API endpoint. In addition, some customers may have experienced intermittent latency and connection failures to the Azure Management Portal. Some services with a reliance on triggers from service management calls may have seen failures for running instances.

Root cause and mitigation: The Azure Service Management layer (ASM) is composed of multiple services and components. During this event, a platform service update introduced a dependency conflict between two backend components.

Engineers established that one of the underlying components was incompatible with the newer version of a dependent component. When this incompatibility occurs, internal communications slow down and may fail. The impact of this is that above a certain load threshold, newer connections to ASM would have a higher fault rate. Azure Service management has a dependency on the impacted components and consequently, was partially impacted.

The ASM update went through the extensive standard testing, and the fault did not manifest itself when the update transited through first, the pre-production environments, and later through the first Production slice where it baked for three days. There was no indication of such a failure in those environments during this time.

When the update rolled out to the main Production environment, there was no immediately visible impact. Hours later, engineers started seeing alerts on failed connections. This was challenging to diagnose as ASM appeared internally healthy and the intermittent connection failures could have been due to multiple reasons related to network or other parts of the infrastructure. The offending components were identified based on a review of detailed logging of the impacted components, and the incident was mitigated by removing the dependency on the faulted code path.

A secondary issue occurred coincidentally which prevented notifications from reaching the Azure Service Health Dashboard or the Azure Portal Service Health blade. Engineers have determined that as part of a deployment earlier, a bug was introduced in the communications tooling. The issue was discovered during the event, and a rollback was performed immediately. While it was not related to the incident above, it delayed our notifications to our 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):

1. Refine validation and service upgrade processes to prevent future occurrences with customer impact [In Progress]
2. Enhance telemetry for detecting configuration anomalies of the backend components post service upgrade to mitigate customer impact scenarios faster [In Progress]

Provide feedback: Please help us improve the Azure customer communications experience by taking our survey: https://aka.ms/J0KF-78G (https://aka.ms/J0KF-78G)

7/24

Delays in ARM generated Activity Logs

Summary of impact: Between 23:00 UTC on 24 Jul 2018 and 07:00 UTC on 27 Jul 2018, customers may have not received activity logs for ARM resources. All service management operations for ARM resources would have been successful, however the corresponding activity logs would not have been generated. As a result, any dependent actions or workflows that are dependent on ARM generated activity logs would have not been initiated. Engineers have confirmed that customers can receive ARM generated activity logs as of 07:00 on the 27 July 2018. However, at this time customers may not able to access the activity logs that were generated during this impact time frame.

Preliminary root cause: Engineers determined that a recent Azure Resource Manager deployment task impacted instances of a backend service which became unhealthy, preventing requests from completing.

Mitigation: Engineers deployed a platform hotfix to mitigate the issue.

Next steps: This incident is remaining active as engineers access options for the missing logs that would have been generated during the impact window. Any customers who remains impacted will receive communications via their management portal. This is being tracked under the tracking id: FLLH-MRG.

7/23

Error notifications in Microsoft Azure Portal

Summary of impact: Between approximately 20:00 and 22:56 UTC on 23 Jul 2018, a subset of customers may have received timeout errors or failure notifications when attempting to load multiple blades in the Microsoft Azure Portal. Customers may also have experienced slowness or difficulties logging into the Portal.

Preliminary root cause: Engineers determined that instances of a backend service became unhealthy, preventing these requests from completing.

Mitigation: Engineers performed a change to the service configuration to return the backend service to a healthy state and mitigate the issue.

Next steps: Engineers will continue to investigate to establish the full root cause and prevent future

occurrences.

7/16

RCA - Networking - Service Availability

Summary of impact: Between 14:55 and 16:15 UTC on 16 July 2018, a subset of Azure customers in West US, West Central US, India South, India Central, and Australia East may have experienced difficulties connecting to Azure endpoints, which in-turn may have caused errors when accessing Microsoft Services in the impacted regions.

Root cause and mitigation: Azure utilizes the industry standard Border Gateway Protocol (BGP) to advertise network paths (routes) for all Azure services. As part of the monthly route prefix update process for ExpressRoute, a misconfigured ExpressRoute device began advertising a subset of Microsoft routes which conflicted with the Microsoft global network. All traffic destined for the public endpoints via these routes either externally or from within an impacted region could not reach their destination. This included Azure Virtual Machines and other services with a dependence on impacted Azure storage accounts. Engineers responded to the alerts, identified the cause of the routing anomaly, and isolated the portion of the network 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):

1. Enhanced configuration validation at the time of deployment and automatic detection of configuration anomalies post deployment – in progress

2. Monitoring and alerting of incorrect route advertisement into the Microsoft global network for faster identification and mitigation – in progress

3. Enforce network route protection to ensure that route advertisements are more limited in scope and radius – in progress

Provide feedback: Please help us improve the Azure customer communications experience by taking our survey: https://aka.ms/YY0R-NS8 (https://aka.ms/YY0R-NS8)

7/13

RCA - Virtual Machines - UK South

Summary of impact: Between 22:13 UTC on 12 July 2018 and 16:20 UTC on 13 July 2018, a subset of customers using Virtual Machines in UK South may have experienced difficulties connecting to some of their resources hosted in the region. Customers may have been unable to establish RDP or SSH session into their resources. This issue impacted customers with resources hosted on one of the compute scale units hosted in the region.

Root cause and mitigation: The impacted scale unit was introduced into rotation a few days before the impact window. Prior to its release to production and during the build out process, an incorrect network parameter was set which impacted virtual network (VNET) connectivity on this scale unit. This connectivity issue triggered the expected monitoring alerts which were temporally muted from the network monitoring tool as the build out process progressed. Unfortunately, these alerts weren't unmuted prior to the unit accepting workloads. Engineers partially mitigated the issue by asking impacted customers to deallocate, restart impacted workloads and taking the scale unit out of rotation. The issue was fully mitigated by adjusting the incorrect parameter to the expected value.

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):

1. Scan of recently built scale units to validate that the correct network parameter is set [Done]
2. Scan of recently built scale units to validate that monitoring alerts are unmuted [Done]
3. Roll out automation to remove the manual setting of the network parameter [Done]
4. Add a check to ensure that no network alerts are supressed before scale units are introduced into rotation [In Progress]

Provide feedback: Please help us improve the Azure customer communications experience by taking our survey: https://aka.ms/RW6S-DL8 (https://aka.ms/RW6S-DL8)

7/11

RCA - SQL Database - Intermittent Database Login Failures - South Central US

Summary of impact: Between 18:38 on 11 Jul 2018 and 01:13 UTC on 12 Jul 2018, a subset of customers using SQL Database in South Central US may have experienced intermittent issues accessing services.

New connections to existing databases in this region may have resulted in an error or timeout, and existing connections may have been terminated. Engineers determined that the initial issue was mitigated at 20:54 UTC but login failure recurred again at 22:25 UTC. The overall impact to the region would have been less than 10% of traffic being served to SQL Databases, however the impact to customers in the region should have been substantially lower as retries would have been successful in most cases.

Root cause and mitigation: The South Central US region has two clusters of gateway machines (or connectivity routers) serving login traffic. Engineers determined that a set of network devices serving the gateway machines malfunctioned, causing packet drops leading to intermittent connectivity to Azure SQL DB in the region. One out of two gateway clusters serving the region was impacted. A total of 37 out of 100 gateway nodes in the region experienced packet drops. Login success rate in the region dropped to 90%. Retries would have been successful in most cases. The triggering cause was an increased load in the region. The issue was mitigated by load balancing across the two clusters in the region and thus reducing load on gateway machines in the impacted region.

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 work is underway to increase gateway capacity in the region as well as upgrading network devices.

Provide feedback: Please help us improve the Azure customer communications experience by taking our survey https://aka.ms/V2M7-5TZ (https://aka.ms/V2M7-5TZ)

7/1

RCA - IoT Hub - Connectivity Issues

Summary of impact: Between 01:30 UTC on 01 Jul 2018 and 18:00 UTC on 02 Jul 2018, a subset of customers using Azure IoT Hub in West Europe, North Europe, East US, and West US may have experienced difficulties connecting to resources hosted in these regions.

Customer impact: During the impact, symptoms may have included failures when creating IoT Hub, connection latencies, and disconnects in the impacted regions.

Root cause and mitigation: A subset of IoT Hub scale units experienced a significant increase in load due to high volume of operational activity and recurring import jobs across multiple instances, which ultimately throttled the backend system. The issue was exacerbated as the loads were unevenly distributed across the clusters.

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):
1. Optimize throttling of queries to prevent system load – In Progress
2. Fine-tune load balancing algorithm to provide quicker detection and alerting – In Progress
3. Improve failover strategy to quickly mitigate – In Progress

Provide feedback: Please help us improve the Azure customer communications experience by taking our survey https://aka.ms/V8BX-PW8 (https://aka.ms/V8BX-PW8)

June 2018

6/27

RCA - App Service - West Europe

Summary of impact: Between 16:00 UTC on 27 Jun 2018 and 13:00 UTC on 28 Jun 2018, a subset of customers using App Service in West Europe may have received HTTP 500-level response codes, timeouts or high latency when accessing App Service (Web, Mobile and API Apps) deployments hosted in this region.

Root cause and mitigation: During a recent platform deployment several App Service scale units in West Europe encountered a backend performance regression due to a modification in the telemetry collection systems. Due to this regression, customer with .NET applications running large workloads may have encountered application slowness. The root cause of this was an inefficiency in the telemetry collection pipeline which caused overall virtual machine performance degradation and slowdown. The issue was detected automatically, and the engineering team was engaged. A mitigation to remove the inefficiency causing the issue was applied at 10:00 UTC on June 28. After further review, a secondary mitigation was applied to a subset of VMs at 22:00 UTC on June 28. More than 90% of the impacted customers saw mitigation at this time. After additional monitoring, a final mitigation was applied to a single remaining scale unit at 15:00 UTC on June 29. All customers were mitigated at this 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):
• Removing the changes that caused the regression
• Reviewing and if necessary adjusting the performance regression detection and alerting

Provide feedback: Please help us improve the Azure customer communications experience by taking our survey: https://aka.ms/7NTS-N70 (https://aka.ms/7NTS-N70)

6/25

RCA - Multiple Services - South Central US

Summary of impact: Between 19:40 and 20:52 UTC on 25 Jun 2018, a subset of customers in South Central US may have experienced difficulties connecting to resources and/or 500-level errors hosted in this region. Virtual Machines may have rebooted unexpectedly. Impacted services included: Storage, Virtual Machines, Key Vault, Site Recovery, Machine Learning, Cloud Shell, Logic Apps, Redis Cache, Visual Studio Team Services, Service Bus, ExpressRoute, Application Insights, Backup, Networking, API Management, App Service (Linux) and App Service.

Root cause and mitigation: One storage scale unit in South Central US experienced a load imbalance. This load imbalance caused network congestion on the backend servers. Due to the congestion, many of the read requests started to timeout as the servers which were hosting the data couldn't complete the request in time. Azure Storage uses erasure coding to more efficiently and durably store data. Since the servers, which were hosting the primary copy of the data, didn't respond in time, it triggered an alternative read mode (aka reconstruct read) which puts more load on the network. The network load on the backend system became high enough that many of the customer requests including those from VMs started timing out. Our load balancing system takes care of balancing the load across the servers to prevent hot spots from being created but in this case, it was not sufficient. The system recovered on its own. The Storage engineering team applied additional configuration changes to spread the load more evenly and removed some of the load on 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):
- Load balancing system changes to handle such load imbalances better [in progress]
- Controlling the rate and resource usage of the reconstruct read [in progress]

Provide feedback: Please help us improve the Azure customer communications experience by taking our survey: https://aka.ms/CD4_-VF0 (https://aka.ms/CD4_-VF0)

6/20

RCA - Azure Data Factory Services - Multiple Regions

Summary of impact: Between 12:25 and 15:00 UTC on 20 Jun 2018, customers using Data Factory V1, Data Factory V2, Data Movement, SSIS Integration Runtime, and Data Movement & Dispatch may have experienced errors, including, but not limited to, pipeline execution errors, copy activity and store procedure activity errors, manual and scheduled SSIS activity failures, and data movement and dispatch failures.
As a workaround, customers had to manually restart their self-hosted Integration Runtime environments.

Workaround: Self Hosted Integration Runtime customers had to manually restart their self-hosted Integration Runtime environments.

Root cause and mitigation: The underlying cause of this incident was a faulty service deployment, which resulted in a configuration error being deployed alongside the front end micro service. This configuration is leveraged for all communications to the underlying data movement micro services. As a result of the configuration error, no API calls succeeded. While the deployment did follow all documented safe deployment processes, an earlier update in the configuration rollout caused a configuration to be deployed unexpectedly. The service deployment will automatically load the latest configuration, thus the configuration error became a blocker for the micro service API path. Engineers rolled back the faulty deployment, which reverted to the correct configuration, and thus allowed the API calls to succeed for the underlying micro-services.

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): Engineers will modify the configuration issuing and authorization process, so that faulty configurations cannot be deployed in future. This will ensure that any deployment activity with these configurations will not cause an interruption to downstream services.

Provide feedback: Please help us improve the Azure customer communications experience by taking our survey https://aka.ms/Z4MD-6Y0 (https://aka.ms/Z4MD-6Y0)

6/19

RCA - Service Availability Issue in North Europe

Summary of impact: From 17:44 on 19 Jun 2018 to 04:30 UTC on 20 Jun 2018, customers using Azure services in North Europe may have experienced connection failures when attempting to access resources hosted in the region. Customers leveraging a subset of Azure services may have experienced residual

impact for a sustained period post-mitigation of the underlying issue.

Final root cause and mitigation: On 19 Jun 2018, Datacenter Critical Environments systems in one of our datacenters in the North Europe region experienced an increase in outside air temperature which triggered a confluence of events that contributed to this outage.

The humidity increase resulted in a datacenter control system requesting for evaporative cooling across the data center campus. When the call for cooling occurred, it uncovered issues within the controls system which led to failures within the mechanical system in the data centers. Manual intervention was required to help remediate the escalating temperature and humidity issues within the data center.

The team's manual intervention in conjunction with the control system issues did not immediately remediate the environmental issues. However the team was eventually able to clear and reset the control systems, bringing the data center cooling systems back in control lowering the temperature and humidity.

This unexpected rise in humidity levels in the operational areas caused multiple Top of Rack (TORs) network devices and hard disk drives supporting two Storage scale units in the region to experience hardware component failures. These failures caused significant latency and/or intra scale unit communications issues between the servers which led to availability issues for customers with data hosted on the affected scale units.

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):

1. Correct issues for datacenter evaporative cooling systems in the affected data center [complete]
2. Audit controls system and test functionalities [in progress]
3. Update standard and emergency operational procedures for environmental excursion [complete]
4. Engineers continue to analyze detailed event data to determine if additional environmental systems modifications are required [in progress]

Provide feedback: Please help us improve the Azure customer communications experience by taking our survey: https://aka.ms/7THS-LY8 (https://aka.ms/7THS-LY8)

6/14

Azure Bot Services - Mitigated

Summary of impact: Between 07:15 and 10:06 UTC on 14 Jun 2018, a subset of customers using Azure Bot Service may have experienced difficulties while trying to connect to bot resources.

Preliminary root cause: Engineers identified a code defect with a recent deployment task as the potential root cause.

Mitigation: Engineers performed a rollback of the recent deployment task to mitigate the issue.

Next steps: Engineers will continue to investigate to establish the full root cause and prevent future occurrences.

6/13

RCA - Multiple Azure Services availability issues and Service Management issues for a subset of Classic Azure resources - South Central US

Summary of impact: Between 15:57 UTC and 19:00 UTC on 13 Jun 2018 a subset of customers in South Central US may have experienced difficulties connecting to resources hosted in this region. Engineers have determined that this was caused by an underlying Storage availability issue. Other services that leverage Storage in this region may also be experienced impact related to this. These services may include: Virtual Machines, App Service, Visual Studio Team Services, Logic Apps, Azure Backup, Application Insights, Service Bus, Event Hub, Site Recovery, Azure Search, and Media Services. In addition, customers may have experienced failures when performing Service Management operations on their resources. Communications for Service Management operations issue is published to the Azure Portal.

Root cause and mitigation: One storage scale unit in the South Central US region experienced a significant increase in load, which caused increased resource utilization on the backend servers in the scale unit. The increased resource utilization caused several backend roles to become temporary unresponsive resulting in timeouts and other errors, which caused VMs and other storage dependent services to be impacted. The Storage service utilizes automatic load balancing to help automatically mitigate this type of incident, however in this case automatic load balancing was not sufficient and engineer intervention was required to mitigate the incident. To stabilize the scale unit and backend roles, engineers rebalanced the load. Impacted services recovered shortly thereafter. Engineers are continuing to investigate the cause for the increase in load and implement additional steps to help prevent a reoccurrence.

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):

1. Improve load balancing system to better manage the load across servers -pending
2. Improve resource usage management to better handle low resource situations – pending
3. Improve server restart time after an incident that results in degraded Storage availability – in progress
4. Improve fail over strategy to help ensure that impacted services can more quickly recover - in progress

Provide feedback: Please help us improve the Azure customer communications experience by taking our survey https://aka.ms/3TP7-TM8 (https://aka.ms/3TP7-TM8)