

Service Level Agreement for Microsoft Azure operated by 21Vianet

Last updated: September 2019

1. Introduction

This Service Level Agreement for Azure (this “SLA”) is made by 21Vianet in connection with, and is a part of, the agreement under which Customer has purchased Azure Services from 21Vianet (the “Agreement”). This SLA applies to the following Azure Services:

- Application Gateway
- Automation
- Azure IoT Hub
- Backup
- CDN
- Cloud Services
- Event Hubs
- ExpressRoute
- HDInsight
- Media Services
- MySQL Database on Azure
- Multi-Factor Authentication
- Notification Hubs
- Power BI Embedded
- Redis Cache
- Managed Disks
- Stream Analytics
- Scheduler
- Service Bus
- Site Recovery
- SQL Database
- SQL Data Warehouse
- Storage
- StorSimple
- Service Fabric
- SQL Server Stretch Database
- Traffic Manager
- Key Vault
- Virtual Machines
- VPN Gateway
- App Service
- Virtual Machine Scale Sets
- Cognitive Services
- Azure Analysis Services
- Azure Cosmos DB
- Network Watcher
- Azure Monitor
- API Management
- Azure Database for MySQL
- Azure Database for PostgreSQL
- Functions
- Azure Active Directory B2C
- Azure Database for MariaDB
- Azure Kubernetes Service (AKS)
- Data Factory
- Security Center
- Azure Data Explorer
- Event Grid
- Azure Firewall

We provide financial backing to our commitment to achieve and maintain Service Levels for our Services. If we do not achieve and maintain the Service Levels for each Service as described in this SLA, then you may be eligible for a credit towards a portion of your monthly service fees. These terms will be fixed for term of your Agreement. If a subscription is renewed, the version of this SLA that is current at the time the renewal term commences will apply throughout the renewal term. We will provide at least 90 days' notice for adverse material changes to this SLA. You can review the most current version of this SLA at any time by visiting <http://www.windowsazure.cn/support/legal/sla>.

2. General Terms

a. Definitions

- i. "Claim" means a claim submitted by Customer to 21Vianet pursuant to this SLA that a Service Level has not been met and that a Service Credit may be due to Customer.
- ii. "Customer" refers to the organization that has entered into the Agreement.
- iii. "Customer Support" means the services by which 21Vianet may provide assistance to Customer to resolve issues with the Services.
- iv. "Error Code" means an indication that an operation has failed, such as an HTTP status code in the 5xx range.
- v. "External Connectivity" is bi-directional network traffic over supported protocols such as HTTP and HTTPS that can be sent and received from a public IP address.
- vi. "Incident" means any set of circumstances resulting in a failure to meet a Service Level.
- vii. "Management Portal" means the web interface, provided by 21Vianet, through which customers may manage the Service.
- viii. "21Vianet" means the 21Vianet entity that appears on Customer' s Agreement.
- ix. "Preview" refers to a preview, beta, or other pre-release version of a service or software offered to obtain customer feedback.

- x. "Service" or "Services" refers to a Azure service provided to Customer pursuant to the Agreement for which an SLA is provided below.
- xi. "Service Credit" is the percentage of the monthly service fees for the affected Service or Service Resource that is credited to Customer for a validated Claim.
- xii. "Service Level" means standards 21Vianet chooses to adhere to and by which it measures the level of service it provides for each Service as specifically set forth below.
- xiii. "Service Resource" means an individual resource available for use within a Service.
- xiv. "Success Code" means an indication that an operation has succeeded, such as an HTTP status code in the 2xx range.
- xv. "Support Window" refers to the period of time during which a Service feature or compatibility with a separate product or service is supported.
- xvi. "Virtual Network" refers to a virtual private network that includes a collection of user-defined IP addresses and subnets that form a network boundary within Azure.
- xvii. "Virtual Network Gateway" refers to a gateway that facilitates cross-premises connectivity between a Virtual Network and a customer on-premises network.

b. Service Credit Claims

- i. In order for 21Vianet to consider a Claim, Customer must submit the Claim to Customer Support within two months of the end of the billing month in which the Incident that is the subject of the Claim occurs. Customer must provide to Customer Support all information necessary for 21Vianet to validate the Claim, including but not limited to detailed descriptions of the Incident, the time and duration of the Incident, the affected resources or operations, and any attempts made by Customer to resolve the Incident
- ii. 21Vianet will use all information reasonably available to it to validate the Claim and to determine whether any Service Credits are due.

- iii. In the event that more than one Service Level for a particular Service is not met because of the same Incident, Customer must choose only one Service Level under which a Claim may be made based on the Incident.
 - iv. Service Credits apply only to fees paid for the particular Service, Service Resource, or Service tier for which a Service Level has not been met. In cases where Service Levels apply to individual Service Resources or to separate Service tiers, Service Credits apply only to fees paid for the affected Service Resource or Service tier, as applicable.
- c. SLA Exclusions.** This SLA and any applicable Service Levels do not apply to any performance or availability issues:
- i. Due to factors outside 21Vianet' s reasonable control (for example, a network or device failure external to 21Vianet' s data centers, including at Customer' s site or between Customer' s site and 21Vianet' s data center);
 - ii. That resulted from Customer' s use of hardware, software, or services not provided by 21Vianet as part of the Services (for example, third-party software or services purchased from the Azure Store or other non-Azure services provided by 21Vianet);
 - iii. Due to Customer' s use of the Service in a manner inconsistent with the features and functionality of the Service (for example, attempts to perform operations that are not supported) or inconsistent with published documentation or guidance;
 - iv. That resulted from faulty input, instructions, or arguments (for example, requests to access files that do not exist);
 - v. Caused by Customer' s use of the Service after 21Vianet advised Customer to modify its use of the Service, if Customer did not modify its use as advised;
 - vi. During or with respect to Previews or to purchases made using 21Vianet subscription credits;
 - vii. That resulted from Customer' s attempts to perform operations that exceed prescribed quotas or that resulted from throttling of suspected abusive behavior;

- viii. Due to Customer' s use Service features that are outside of associated Support Windows; or
- ix. Attributable to acts by persons gaining unauthorized access to 21Vianet' s Service by means of Customer' s passwords or equipment or otherwise resulting from Customer' s failure to follow appropriate security practices.

d. Service Credits

- i. The amount and method of calculation of Service Credits is described below in connection with each Service.
- ii. Service Credits are Customer' s sole and exclusive remedy for any failure to meet any Service Level.
- iii. The Service Credits awarded in any billing month for a particular Service or Service Resource will not, under any circumstance, exceed Customer' s monthly service fees that Service or Service Resource, as applicable, in the billing month.
- iv. For Services purchased as part of a suite, the Service Credit will be based on the pro-rata portion of the cost of the Service, as determined by 21Vianet in its reasonable discretion. In cases where Customer has purchased Services from a reseller, the Service Credit will be based on the estimated retail price for the applicable Service, as determined by 21Vianet in its reasonable discretion.

3. Azure Services Subject to SLA

Application Gateway

We guarantee that each Application Gateway Cloud Service having two or more medium or larger instances, or deployments capable of supporting autoscale or zone redundancy, will be available at least 99.95% of the time.

i. Additional Definitions

- A. **"Application Gateway Cloud Service"** refers to a collection of two or more medium or larger Application Gateway instances or deployments capable of supporting autoscale or zone redundancy, configured to perform HTTP load balancing services.

- B. **“Maximum Available Minutes”** is the total accumulated minutes during a billing month during which an Application Gateway Cloud Service has been deployed in a Azure subscription.
- C. **“Downtime”** is the total accumulated Maximum Available Minutes during a billing month for a given Application Gateway Cloud Service during which the Application Gateway Cloud Service is unavailable. A given minute is considered unavailable if all attempts to connect to the Application Gateway Cloud Service throughout the minute are unsuccessful.
- D. **“Monthly Uptime Percentage”** The Monthly Uptime Percentage is calculated using the following formula:

$$\frac{(\text{Maximum Available Minutes} - \text{Downtime})}{\text{Maximum Available Minutes}} * 100$$

- E. Service Credit

Monthly Uptime Percentage	Service Credit
<99.95%	10%
<99%	25%

Automation Service

- i. **Additional Definitions**

- A. “Job” means the execution of a Runbook.
- B. “Planned Start Time” is a time at which a Job is scheduled to begin executing.
- C. “Runbook” means a set of actions specified by Customer to execute within Azure.

- ii. **Monthly Uptime Calculation and Service Levels for the Automation Service**

- A. “Delayed Jobs” is the total number of Jobs that fail to start within thirty (30) minutes of their Planned Start Times.
- B. “Total Jobs” is the total number of Jobs scheduled for execution during a given billing month.

C. "Monthly Uptime Percentage" for the Automation Service is calculated as Total Jobs less Delayed Jobs divided by Total Jobs in a billing month for a given Azure subscription. Monthly Uptime Percentage is represented by the following formula:

D. The following Service Levels and Service Credits are applicable to Customer' s use of the Automation Service:

Monthly Uptime Percentage	Service Credit
<99.9%	10%
<99%	25%

Azure IoT Hub

- i. For IoT Hub, we promise that at least 99.9% of the time deployed IoT hubs will be able to send messages to and receive messages from registered devices and the Service will be able to perform create, read, update, and delete operations on IoT hubs.
- ii. No SLA is provided for the Free Tier of IoT Hub.

iii. Additional Definitions

- A. "Deployment Minutes" is the total number of minutes that a given IoT hub has been deployed in Azure during a billing month.
- B. "Maximum Available Minutes" is the sum of all Deployment Minutes across all IoT hubs deployed in a given Azure subscription during a billing month.
- C. "Message" refers to any content sent by a deployed IoT hub to a device registered to the IoT hub or received by the IoT hub from a registered device, using any protocol supported by the Service.
- D. "Device Identity Operations" refers to create, read, update, and delete operations performed on the device identity registry of an IoT hub.

- E. Downtime : The total accumulated Deployment Minutes, across all IoT hubs deployed in a given Azure subscription, during which the IoT hub is unavailable. A minute is considered unavailable for a given IoT hub if all continuous attempts to send or receive Messages or perform Device Identity Operations on the IoT hub throughout the minute either return an Error Code or do not result in a Success Code within five minutes.
- F. Monthly Uptime Percentage: The Monthly Uptime Percentage is calculated using the following formula:

$$\text{Monthly Uptime \%} = (\text{Maximum Available Minutes} - \text{Downtime}) / \text{Maximum Available Minutes}$$

- G. Service Credit:

Monthly Uptime Percentage	Service Credit
<99.9%	10%
<99%	25%

- H. Service Level Exceptions: The Free Tier of the IoT Hub Service is not covered by this SLA.

iv. Monthly Uptime Calculation and Service Levels for IoT Hub Device Provisioning Service

- A. "Maximum Available Minutes" is the total number of minutes for a given Device Provisioning Service deployed by the Customer in a Azure subscription during a billing month.
- B. "Downtime" is the total number of minutes within the Maximum Available Minutes during which Device Provisioning Service is unavailable. A minute is considered unavailable for a given Device Provisioning Service if all continuous attempts to register a device or perform enrollment/registration record operations on the Device Provisioning Service throughout the minute either return an Error Code or do not result in a Success Code within two minutes.

- C. Monthly Uptime Percentage: The Monthly Uptime Percentage is calculated using the following formula:
- D. Monthly Uptime % = (Maximum Available Minutes - Downtime) / Maximum Available Minutes X 100
- E. The following Service Levels and Service Credits are applicable to Customer's use of IoT Hub Device Provisioning Service
- F. Service Credit:

MONTHLY UPTIME PERCENTAGE	SERVICE CREDIT
<99.9%	10%
<99%	25%

Backup Service

i. **Additional Definitions**

- A. "Backup" or "Back Up" is the process of copying computer data from a registered server to a Backup Vault.
- B. "Backup Agent" refers to the software installed on a registered server that enables the registered server to Back Up or Restore one or more Protected Items.
- C. "Backup Vault" refers to a container in which Customer may register one or more Protected Items for Backup.
- D. "Failure" means that either the Backup Agent or the Service fails to fully complete a properly configured Backup or Recovery operation due to unavailability of the Backup Service.
- E. "Protected Item" refers to a collection of data, such as a volume, database, or virtual machine that has been scheduled for Backup to the Backup Service such that it is enumerated as a Protected Item in the Protected Items tab in the Recovery Services section of the Management Portal.

F. "Recovery" or "Restore" is the process of restoring computer data from a Backup Vault to a registered server.

ii. **Monthly Uptime Calculation and Service Levels for Backup Service**

A. "Deployment Minutes" is the total number of minutes during which a Protected Item has been scheduled for Backup to a Backup Vault.

B. "Maximum Available Minutes" is the sum of all Deployment Minutes across all Protected Items for a given Azure subscription during a billing month.

C. "Downtime" is the total accumulated Deployment Minutes across all Protected Items scheduled for Backup by Customer in a given Azure subscription during which the Backup Service is unavailable for the Protected Item. The Backup Service is considered unavailable for a given Protected Item from the first Failure to Back Up or Restore the Protected Item until the initiation of a successful Backup or Recovery of a Protected Item, provided that retries are continually attempted no less frequently than once every thirty minutes.

D. "Monthly Uptime Percentage" for the Backup Service is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in a billing month for a given Azure subscription. Monthly Uptime Percentage is represented by the following formula:

E. The following Service Levels and Service Credits are applicable to Customer' s use of the Backup Service:

Monthly Uptime Percentage	Service Credit
<99.9%	10%
<99%	25%

CDN Service

i. **Monthly Uptime Calculation and Service Levels for CDN Service**

- A. 21Vianet is not responsible for comprehensive monitoring of customer content; this responsibility lies with customer. 21Vianet will review data from any commercially reasonable independent measurement system used by Customer.
- B. Customer must select a set of agents from the measurement system's list of standard agents that are generally available and represent at least five major metropolitan areas in the People's Republic of China.
1. Measurement System tests (frequency of at least one test per hour per agent) will be configured to perform one HTTP(S) GET operation according to the model below:
 2. A test file will be placed on Customer's origin (e.g., Azure Storage account).
 3. The GET operation will retrieve the file through the CDN Service, by requesting the object from the appropriate Azure domain name hostname.
 4. The test file will meet the following criteria:
 - i. The test object will meet CDN caching specifications. For example, the response header should not include 'set-cookie' ; cache-control attributes cannot be set to 'private' or 'no-cache' ; the 'expires' value cannot be set at a past time value.
 - ii. The test object will be a file at least 50KB in size and no larger than 1MB.
 - iii. Raw data will be trimmed to eliminate any measurements that came from an agent experiencing technical problems during the measurement period.

Remarks:

1. In the event test results should reveal issues, please contact us as soon as possible so that we may verify said test results at first

opportunity in order to locate the causes for these issues.

2. If the chosen test source itself contains problematic links, this will cause anomalies in the test results; these anomalies should be excluded when calculating the final availability value.

- C. "Monthly Uptime Percentage" is the percentage of HTTP transactions in which the CDN responds to client requests and delivers the requested content without error. Monthly Uptime Percentage of the CDN Service is calculated as the number of times the object was delivered successfully divided by the total number of requests (after removing erroneous data).
- D. The following Service Levels and Service Credits are applicable to Customer' s use of the CDN Service:

Monthly Uptime Percentage	Service Credit
<99.9%	5%
<99%	20%

Cloud Services

i. **Additional Definitions**

- A. "Cloud Services" refers to a set of compute resources utilized for Web and Worker Roles.
- B. "Web Role" is a Cloud Services component run in the Azure execution environment that is customized for web application programming as supported by IIS and ASP.NET.
- C. "Worker Role" is a Cloud Services component run in the Azure execution environment that is useful for generalized development, and may perform background processing for a Web Role.

ii. **Monthly Uptime Calculation and Service Levels for Cloud Services**

- A. "Maximum Available Minutes" is the total accumulated minutes during a billing month for all Internet facing roles that have two or more instances deployed in different Update Domains. Maximum Available Minutes is measured from when the Tenant has been deployed and its associated roles

have been started resultant from action initiated by Customer to the time Customer has initiated an action that would result in stopping or deleting the Tenant.

- B. "Downtime" is the total accumulated minutes that are part of Maximum Available Minutes that have no External Connectivity.
- C. "Monthly Uptime Percentage" for Cloud Services is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in a billing month for a given Azure subscription. Monthly Uptime Percentage is represented by the following formula:
- D. The following Service Levels and Service Credits are applicable to Customer' s use of Cloud Services:

Monthly Uptime Percentage	Service Credit
<99.95%	10%
<99%	25%

Event Hubs

- i. **Monthly Uptime Calculation and Service Levels for Event Hubs**
 - A. "Deployment Minutes" is the total number of minutes that a given Event Hub has been deployed in Azure during a billing month.
 - B. "Maximum Available Minutes" is the sum of all Deployment Minutes across all Event Hubs deployed by Customer in a given Azure subscription under the Basic or Standard Event Hubs tiers during a billing month.
 - C. "Downtime" is the total accumulated Deployment Minutes, across all Event Hubs deployed by Customer in a given Azure subscription under the Basic or Standard Event Hubs tiers, during which the Event Hub is unavailable. A minute is considered unavailable for a given Event Hub if all

continuous attempts to send or receive Messages or perform other operations on the Event Hub throughout the minute either return an Error Code or do not result in a Success Code within five minutes.

D. "Monthly Uptime Percentage" for Event Hubs is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in a billing month for a given Azure subscription. Monthly Uptime Percentage is represented by the following formula:

E. The following Service Levels and Service Credits are applicable to Customer's use of the Basic and Standard Event Hubs tiers. The Free Event Hubs tier is not covered by this SLA.

Monthly Uptime Percentage	Service Credit
<99.9%	10%
<99%	25%

ExpressRoute

i. We guarantee a minimum of 99.95% ExpressRoute dedicated circuit availability.

ii. **Additional Definitions**

"Dedicated Circuit" means a logical representation of connectivity offered through the ExpressRoute Service between Customer's premises and Azure through an exchange provider or a network service provider, where such connectivity does not traverse the public Internet.

iii. **Monthly Uptime Calculation and Service Levels for ExpressRoute**

"Maximum Available Minutes" is the total number of minutes that a given Dedicated Circuit is linked to one or more Virtual Networks in Azure during a billing month in a given Azure subscription.

"Downtime" is the total accumulated minutes during a billing month for a given Azure subscription during which the Dedicated Circuit is unavailable. A minute is considered unavailable for a given Dedicated Circuit if all attempts by Customer within the minute to establish IP-level connectivity to the Virtual Network Gateway associated with the Virtual Network fail for longer than thirty seconds.

"Monthly Uptime Percentage" for a given Dedicated Circuit in a given billing month is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes. Monthly Uptime Percentage is represented by the following formula:

$$\text{Monthly Uptime \%} = (\text{Maximum Available Minutes} - \text{Downtime}) / \text{Maximum Available Minutes}$$

The following Service Levels and Service Credits are applicable to Customer' s use of each Dedicated Circuit within the ExpressRoute Service:

Monthly Uptime Percentage	Service Credit
<99.95%	10%
<99%	25%

HDInsight

i. **Additional Definitions**

- A. "Cluster Internet Gateway" means a set of virtual machines within an HDInsight Cluster that proxy all connectivity requests to the Cluster.
- B. "HDInsight Cluster" or "Cluster" means a collection of virtual machines running a single instance of the HDInsight Service.

ii. **Monthly Uptime Calculation**

- A. "Deployment Minutes" is the total number of minutes that a given HDInsight Cluster has been deployed in Azure.

- B. "Maximum Available Minutes" is the sum of all Deployment Minutes across all Clusters deployed by Customer in a given Azure subscription during a billing month.
- C. "Downtime" is the total accumulated Deployment Minutes when the HDInsight Service is unavailable. A minute is considered unavailable for a given Cluster if all continual attempts within the minute to establish a connection to the Cluster Internet Gateway fail.
- D. "Monthly Uptime Percentage" for the HDInsight Service is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in a billing month for a given Azure subscription. Monthly Uptime Percentage is represented by the following formula:
- E. The following Service Levels and Service Credits are applicable to Customer' s use of the HDInsight Service:

Monthly Uptime Percentage	Service Credit
<99.9%	10%
<99%	25%

Media Services

- For Media Services Encoding, we guarantee 99.9% availability of REST API transactions.
- For Streaming, we will successfully service requests with a 99.9% availability guarantee for existing media content when at least one Streaming Unit is purchased.
- For Live Channels, we guarantee that running Channels will have external connectivity at least 99.9% of the time.
- For Content Protection, we guarantee that we will successfully fulfill key requests at least 99.9% of the time.

- For Indexer, we will successfully service Indexer Task requests processed with an Encoding Reserved Unit 99.9% of the time.

Additional Definitions

"Allocated Egress Bandwidth" is the amount of bandwidth configured by Customer in the Management Portal for a Media Service. Allocated Egress Bandwidth may be labeled "Streaming Units" or a similar name in the Management Portal.

"Channel" means an end point within a Media Service that is configured to receive media data.

"Encoding" means the processing of media files per subscription as configured in the Media Services Tasks.

"Encoding Reserved Unit" means encoding reserved units purchased by the customer in an Azure Media Services account.

"Indexer Task" means a Media Services Task that is configured to index an MP3 input file with a minimum five-minute duration.

"Media Service" means an Azure Media Services account, created in the Management Portal, associated with Customer's Azure subscription. Each Azure subscription may have more than one associated Media Service.

"Media Service Request" means a request issued to Customer's Media Service.

"Media Services Task" means an individual operation of media processing work as configured by Customer. Media processing operations involve encoding, converting, or indexing media files.

"Streaming Unit" means a unit of reserved egress capacity purchased by Customer for a Media Service.

"Valid Key Requests" are all requests made to the Content Protection Service for existing content keys in a Customer's Media Service.

"Valid Media Services Requests" are all qualifying Media Service Requests for existing media content in a Customer's Azure Storage account associated with its Media Service when at least one Streaming Unit has been purchased and allocated to that Media Service. Valid Media Services Requests do not include Media Service Requests for which total throughput exceeds 80% of the Allocated Bandwidth.

Monthly Uptime Calculation and Service Levels for Encoding Service

"Total Transaction Attempts" is the total number of authenticated REST API requests with respect to a Media Service made by Customer during a billing month for a subscription. Total Transaction Attempts does not include REST API requests that return an Error Code that are continuously repeated within a five-minute window after the first Error Code is received.

"Failed Transactions" is the set of all requests within Total Transaction Attempts that do not return a Success Code within 30 seconds from 21Vianet's receipt of the request.

"Monthly Uptime Percentage" for the Azure Media Services Encoding Service is calculated as Total Transaction Attempts less Failed Transactions divided by Total Transaction Attempts in a billing month for a given Azure subscription. Monthly Uptime Percentage is represented by the following formula:

Monthly Uptime % = (Total Transaction Attempts - Failed Transactions) / Total Transaction Attempts

The following Service Levels and Service Credits are applicable to Customer's use of the Azure Media Services Encoding Service:

Monthly Uptime Percentage	Service Credit
<99.9%	10%
<99%	25%

Monthly Uptime Calculation and Service Levels for Indexer Service

"Total Transaction Attempts" is the total number of Indexer Tasks attempted to be executed using an available Encoding Reserved Unit by Customer during a billing month for a subscription.

"Failed Transactions" is the set of Indexer Tasks within Total Transaction Attempts that either

1. Do not complete within a time period that is 3 times the duration of the input file; or
2. Do not start processing within 5 minutes of the time that an Encoding Reserved Unit becomes available for use by the Indexer Task.

"Monthly Uptime Percentage" for the Azure Media Services Indexer Service is calculated as Total Transaction Attempts less Failed Transactions divided by Total Transaction Attempts in a billing month for a given Azure subscription. Monthly Uptime Percentage is represented by the following formula:

Monthly Uptime % = (Total Transaction Attempts - Failed Transactions) / Total Transaction Attempts

The following Service Levels and Service Credits are applicable to Customer' s use of the Azure Media Services Indexer Service:

Monthly Uptime Percentage	Service Credit
<99.9%	10%
<99%	25%

Monthly Uptime Calculation and Service Levels for Streaming Service

"Deployment Minutes" is the total number of minutes that a given Streaming Unit has been purchased and allocated to a Media Service during a billing month.

"Maximum Available Minutes" is the sum of all Deployment Minutes across all Streaming Units purchased and allocated to a Media Service during a billing month.

"Downtime" is the total accumulated Deployment Minutes when the Streaming Service is unavailable. A minute is considered unavailable for a given Streaming Unit if all continuous Valid Media Service Requests made to the Streaming Unit throughout the minute result in an Error Code.

"Monthly Uptime Percentage" for the Azure Media Services Streaming Service is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in a billing month for a given Azure subscription. Monthly Uptime Percentage is represented by the following formula:

Monthly Uptime % = (Maximum Available Minutes - Downtime) / Maximum Available Minutes

The following Service Levels and Service Credits are applicable to Customer' s use of the Azure Media Services On-Demand Streaming Service:

Monthly Uptime Percentage	Service Credit
<99.9%	10%
<99%	25%

Monthly Uptime Calculation and Service Levels for Live Channels

"Deployment Minutes" is the total number of minutes that a given Channel has been purchased and allocated to a Media Service and is in a running state during a billing month.

"Maximum Available Minutes" is the sum of all Deployment Minutes across all Channels purchased and allocated to a Media Service during a billing month.

"Downtime" is the total accumulated Deployment Minutes when the Live Channels Service is unavailable. A minute is considered unavailable for a given Channel if the Channel has no External Connectivity during the minute.

"Monthly Uptime Percentage" for the Live Channels Service is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in a billing month for a given Azure subscription. Monthly Uptime Percentage is represented by the following formula:

Monthly Uptime % = (Maximum Available Minutes - Downtime) / Maximum Available Minutes

The following Service Levels and Service Credits are applicable to Customer' s use of the Azure Media Services Live Channels Service:

Monthly Uptime Percentage	Service Credit
<99.9%	10%
<99%	25%

Monthly Uptime Calculation and Service Levels for Content Protection Service

"Total Transaction Attempts" are all Valid Key Requests made by Customer during a billing month for a given Azure subscription.

"Failed Transactions" are all Valid Key Requests included in Total Transaction Attempts that result in an Error Code or otherwise do not return a Success Code within 30 seconds after receipt by the Content Protection Service.

"Monthly Uptime Percentage" for Azure Media Services is calculated as Total Transaction Attempts less Failed Transactions divided by Total Transaction Attempts in a billing month for a given Azure subscription. Monthly Uptime Percentage is represented by the following formula:

Monthly Uptime % = (Total Transaction Attempts - Failed Transactions) / Total Transaction Attempts

The following Service Levels and Service Credits are applicable to Customer' s use of the Azure Media Services Content Protection Service:

Monthly Uptime Percentage	Service Credit
<99.9%	10%
<99%	25%

MySQL Database on Azure Service

i. **Additional Definitions**

- A. "Database Server" is a MySQL database server deployed by Customer within the MySQL Database on Azure Service.
- ii. **Monthly Uptime Calculation and Service Levels for the MySQL Database on Azure Service**
- A. "Deployment Minutes" is the total number of minutes that a given Database Server has been deployed in Azure during a billing month.
- B. "Maximum Available Minutes" is the sum of all Deployment Minutes across all Database Servers for a given Azure subscription during a billing month.
- C. "Downtime" is the total accumulated Deployment Minutes across all Database Servers deployed by Customer in a given Azure subscription during which the Database Server is unavailable. A minute is considered unavailable for a given Database Server if all continuous attempts by Customer to establish a connection to the Database Server within the minute fail.
- D. "Monthly Uptime Percentage" for the MySQL Database on Azure Service is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in a billing month for a given Azure subscription. Monthly Uptime Percentage is represented by the following formula:
- E. The following Service Levels and Service Credits are applicable to Customer' s use of the MySQL Database on Azure Service:

Monthly Uptime Percentage	Service Credit
<99.9%	10%
<99%	25%

Multi-Factor Authentication

- i. We guarantee 99.9% availability of Azure Multi-Factor Authentication.
- ii. No SLA is provided for the Free tier of Multi-Factor Authentication.
- iii. Monthly Uptime Calculation and Service Levels for Multi-Factor Authentication Service

"Deployment Minutes" is the total number of minutes that a given Multi-Factor Authentication provider has been deployed in Azure during a billing month.

"Downtime" is the total accumulated Deployment Minutes, across all Multi-Factor Authentication providers deployed by Customer in a given Azure subscription, during which the Multi-Factor Authentication Service is unable to receive or process authentication requests for the Multi-Factor Authentication provider.

"Maximum Available Minutes" is the sum of all Deployment Minutes across all Multi-Factor Authentication providers deployed by Customer in a given Azure subscription during a billing month.

"Monthly Uptime Percentage" for the Multi-Factor Authentication Service is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in a billing month for a given Azure subscription. Monthly Uptime Percentage is represented by the following formula:

$$\text{Monthly Uptime \%} = (\text{Maximum Available Minutes} - \text{Downtime}) / \text{Maximum Available Minutes}$$

The following Service Levels and Service Credits are applicable to Customer' s use of the Multi-Factor Authentication Service:

Monthly Uptime Percentage	Service Credit
<99.9%	10%
<99%	25%

Notification Hubs

i. **Monthly Uptime Calculation and Service Levels for Notification Hubs**

- A. "Deployment Minutes" is the total number of minutes that a given Notification Hub has been deployed in Azure during a billing month.
- B. "Maximum Available Minutes" is the sum of all Deployment Minutes across all Notification Hubs deployed by Customer in a given Azure subscription under the Basic or Standard Notification Hubs tiers during a billing month.
- C. "Downtime" is the total accumulated Deployment Minutes, across all Notification Hubs deployed by Customer in a given Azure subscription under the Basic or Standard Notification Hubs tiers, during which the Notification Hub is unavailable. A minute is considered unavailable for a given Notification Hub if all continuous attempts to send notifications or perform registration management operations with respect to the Notification Hub throughout the minute either return an Error Code or do not result in a Success Code within five minutes.
- D. "Monthly Uptime Percentage" for Notification Hubs is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in a billing month for a given Azure subscription. Monthly Uptime Percentage is represented by the following formula:
- E. The following Service Levels and Service Credits are applicable to Customer' s use of the Basic and Standard Notification Hubs tiers. The Free Notification Hubs tier is not covered by this SLA.

Monthly Uptime Percentage	Service Credit
<99.9%	10%
<99%	25%

Power BI Embedded

- ii. We guarantee at least 99.9% availability of Power BI Embedded for users to execute API calls and embed reports.

iii. Additional Definitions

"Deployment Minutes" is the total number of minutes for which a given Power BI Embedded has been provisioned during a billing month.

"Maximum Available Minutes" is the sum of all Deployment Minutes across all Power BI Embedded provisioned by customer in a given Azure subscription during a billing month.

"Downtime" is the total accumulated Deployment Minutes, during which the Power BI Embedded is unavailable. A minute is considered unavailable for a given Power BI Embedded if all continuous attempts within the minute to read or write any portion of Power BI Embedded data result in an Error Code or do not return a response within five minutes.

"Monthly Uptime Percentage" The Monthly Uptime Percentage is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in a billing month for the Azure subscription. Monthly Uptime Percentage is represented by the following formula:

Monthly Uptime % = (Maximum Available Minutes - Downtime) / Maximum Available Minutes

The following Service Levels and Service Credits are applicable to Customer' s use of the Power BI Embedded.

Monthly Uptime Percentage	Service Credit
<99.9%	10%
<99%	25%

Redis Cache Service

iv. We guarantee at least 99.9% of the time that customers will have connectivity between the Cache endpoints and our Internet gateway.

v. Additional Definitions

"Cache" refers to a deployment of the Cache Service created by Customer, such that its Cache Endpoints are enumerated in the Cache tab in the Management Portal.

"Cache Endpoints" refers to endpoints through which a Cache may be accessed.

vi. **Monthly Uptime Calculation and Service Levels for Cache Service**

"Deployment Minutes" is the total number of minutes that a given Cache has been deployed in Azure during a billing month.

"Maximum Available Minutes" is the sum of all Deployment Minutes across all Caches deployed by Customer in a given Azure subscription during a billing month.

"Downtime" is the total accumulated Deployment Minutes, across all Caches deployed by Customer in a given Azure subscription, during which the Cache is unavailable. A minute is considered unavailable for a given Cache when there is no connectivity throughout the minute between one or more Cache Endpoints associated with the Cache and 21Vianet's Internet gateway.

"

Monthly Uptime % = (Maximum Available Minutes - Downtime) / Maximum Available Minutes

The following Service Levels and Service Credits are applicable to Customer's use of the Cache Service, which includes the Standard tier of the Azure Redis Cache Service. The Basic tier of the Azure Redis Cache Service is not covered by this SLA.

Monthly Uptime Percentage	Service Credit
---------------------------	----------------

<99.9%	10%
<99%	25%

Stream Analytics

- i. We guarantee at least 99.9% availability of the Stream Analytics API.
- ii. We guarantee that 99.9% of the time, deployed Stream Analytics jobs will be either processing data or available to process data.
- iii. **Monthly Uptime Calculation for Stream Analytics API Calls**

"Total Transaction Attempts" is the total number of authenticated REST API requests to manage a streaming job within the Stream Analytics Service by Customer during a billing month for a given Azure subscription.

"Failed Transactions" is the set of all requests within Total Transaction Attempts that return an Error Code or otherwise do not return a Success Code within five minutes from 21Vianet' s receipt of the request.

"Monthly Uptime Percentage" for API calls within the Stream Analytics Service is calculated as Total Transaction Attempts less Failed Transactions divided by Total Transaction Attempts in a billing month for a given Azure subscription. Monthly Uptime Percentage is represented by the following formula:

$$\text{Monthly Uptime \%} = (\text{Total Transaction Attempts} - \text{Failed Transactions}) / \text{Total Transaction Attempts}$$

The following Service Credits are applicable to Customer' s use of API calls within the Stream Analytics Service.

Monthly Uptime Percentage	Service Credit
<99.9%	10%
<99%	25%

- iv. **Monthly Uptime Calculation for Stream Analytics Jobs**

"Deployment Minutes" is the total number of minutes that a given job has been deployed within the Stream Analytics Service during a billing month.

"Maximum Available Minutes" is the sum of all Deployment Minutes across all jobs deployed by Customer in a given Azure subscription during a billing month.

"Downtime" is the total accumulated Deployment Minutes, across all jobs deployed by Customer in a given Azure subscription, during which the job is unavailable. A minute is considered unavailable for a deployed job if the job is neither processing data nor available to process data throughout the minute.

"Monthly Uptime Percentage" for jobs within the Stream Analytics Service is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in a billing month for a given Azure subscription. Monthly Uptime Percentage is represented by the following formula:

$$\text{Monthly Uptime \%} = (\text{Maximum Available Minutes} - \text{Downtime}) / \text{Maximum Available Minutes}$$

The following Service Credits are applicable to Customer' s use of jobs within the Stream Analytics Service.

Monthly Uptime Percentage	Service Credit
<99.9%	10%
<99%	25%

Scheduler

i. **Additional Definitions**

A. "Planned Execution Time" is a time at which a Scheduled Job is scheduled to begin executing.

B. "Scheduled Job" means an action specified by Customer to execute within Azure according to a specified schedule.

ii. **Monthly Uptime C**

A. "Downtime" is the total accumulated minutes in a billing month during which one or more of Customer's Scheduled Jobs is in a state of delayed execution. A given Scheduled Job is in a state of delayed execution if it has not begun executing after a Planned Execution Time, provided that such delayed execution time shall not be considered Downtime if the Scheduled Job begins executing within thirty (30) minutes after a Planned Execution Time.

B. "Maximum Available Minutes" is the total number of minutes in a billing month.

C. "Monthly Uptime Percentage" for the Scheduler Service is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in a billing month for a given Azure subscription. Monthly Uptime Percentage is represented by the following formula:

D. The following Service Levels and Service Credits are applicable to Customer's use of the Scheduler Service:

Monthly Uptime Percentage	Service Credit
<99.9%	10%
<99%	25%

Service Bus Service

1. For Service Bus Relays, we guarantee that at least 99.9% of the time, properly configured applications will be able to establish a connection to a deployed Relay .

2. For Service Bus Queues and Topics, we guarantee that at least 99.9% of the time , properly configured applications will be able to send or receive messages or perform other operations on a deployed Queue or Topic.

- i. **Additional Definitions**

- A. "Message" refers to any user-defined content sent or received through Service Bus Relays, Queues, or Topics, using any protocol supported by Service Bus.

- ii. **Monthly Uptime Calculation and Service Levels for Relays**

- A. "Deployment Minutes" is the total number of minutes that a given Relay has been deployed in Azure during a billing month.
- B. "Maximum Available Minutes" is the sum of all Deployment Minutes across all Relays deployed by Customer in a given Azure subscription during a billing month.
- C. "Downtime" is the total accumulated Deployment Minutes, across all Relays deployed by Customer in a given Azure subscription, during which the Relay is unavailable. A minute is considered unavailable for a given Relay if all continuous attempts to establish a connection to the Relay throughout the minute either return an Error Code or do not result in a Success Code within five minutes.
- D. "Monthly Uptime Percentage" for Relays is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in a billing month for a given Azure subscription. Monthly Uptime Percentage is represented by the following formula:

$$\text{Monthly Uptime \%} = (\text{Maximum Available Minutes} - \text{Downtime}) / \text{Maximum Available Minutes}$$

- E. The following Service Levels and Service Credits are applicable to Customer' s use of Relays:

Monthly Uptime Percentage	Service Credit
<99.9%	10%

<99%	25%
------	-----

iii. **Monthly Uptime Calculation and Service Levels for Queues and Topics**

- A. "Deployment Minutes" is the total number of minutes that a given Queue or Topic has been deployed in Azure during a billing month.
- B. "Maximum Available Minutes" is the sum of all Deployment Minutes across all Queues and Topics deployed by Customer in a given Azure subscription during a billing month.
- C. "Downtime" is the total accumulated Deployment Minutes, across all Queues and Topics deployed by Customer in a given Azure subscription, during which the Queue or Topic is unavailable. A minute is considered unavailable for a given Queue or Topic if all continuous attempts to send or receive Messages or perform other operations on the Queue or Topic throughout the minute either return an Error Code or do not result in a Success Code within five minutes.
- D. "Monthly Uptime Percentage" for Queues and Topics is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in a billing month for a given Azure subscription. Monthly Uptime Percentage is represented by the following formula:
- E. The following Service Levels and Service Credits are applicable to Customer' s use of Queues and Topics:

Monthly Uptime Percentage	Service Credit
<99.9%	10%
<99%	25%

Site Recovery Service

i. **Additional Definitions**

- A. "Failover" is the process of transferring control, either simulated or actual, of a Protected Instance from a primary site to a secondary site.
 - B. "On-Premises-to-Azure Failover" is the Failover of a Protected Instance from a non-Azure primary site to an Azure secondary site. Customer may designate a particular Azure datacenter as a secondary site, provided that if Failover to the designated datacenter is not possible, 21Vianet may replicate to a different datacenter in the same region.
 - C. "On-Premises-to-On-Premises Failover" is the Failover of a Protected Instance from a non-Azure primary site to a non-Azure secondary site.
 - D. "Protected Instance" refers to a virtual or physical machine configured for replication by the Site Recovery Service from a primary site to a secondary site. Protected Instances are enumerated in the Protected Items tab in the Recovery Services section of the Management Portal.
- ii. **Monthly Uptime Calculation and Service Levels for On-Premises-to-On-Premises Failover**
- A. "Maximum Available Minutes" is the total number of minutes that a given Protected Instance has been configured for On-Premises-to-On-Premises replication by the Site Recovery Service during a billing month.
 - B. "Failover Minutes" is the total number of minutes in a billing month during which a Failover of a Protected Instance configured for On-Premises-to-On-Premises replication has been attempted but not completed.
 - C. "Downtime" is the total accumulated Failover Minutes in which the Failover of a Protected Instance is unsuccessful due to unavailability of the Site Recovery Service, provided that retries are continually attempted no less frequently than once every thirty minutes.
 - D. "Monthly Uptime Percentage" for On-Premises-to-On-Premises Failover of a specific Protected Instance in a given billing month is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes. Monthly Uptime Percentage is represented by the following formula:

- E. The following Service Levels and Service Credits are applicable to Customer' s use of each Protected Instance within the Site Recovery Service for On-Premises-to-On-Premises Failover:

Monthly Uptime Percentage	Service Credit
<99.9%	10%
<99%	25%

iii. **Monthly Recovery Time Objective and Service Levels for On-Premises-to-Azure Failover**

- A. "Recovery Time Objective (RTO)" means the period of time beginning when Customer initiates a Failover of a Protected Instance experiencing either a planned or unplanned outage for On-Premises-to-Azure replication to the time when the Protected Instance is running as a virtual machine in Azure, excluding any time associated with manual action or the execution of Customer scripts.
- B. "**Monthly Recovery Time Objective**" for a specific Protected Instance configured for On-Premises-to-Azure replication in a given billing month is 2 hours.
- C. The following Service Levels and Service Credits are applicable to Customer' s use of each Protected Instance within the Site Recovery Service for On-Premises-to-Azure Failover.

Monthly Recovery Time Objective	Service Credit
> 2 hours	100%

iv. **Monthly Recovery Time Objective and Service Levels for Azure-to-Azure Failover**

- A. "**Recovery Time Objective (RTO)**" means the period of time beginning when Customer initiates a Failover of a Protected Instance experiencing

either a planned or unplanned outage for On-Premises-to-Azure replication to the time when the Protected Instance is running as a virtual machine in Azure, excluding any time associated with manual action or the execution of Customer scripts.

- B. **“Monthly Recovery Time Objective”** for a specific Protected Instance configured for Azure-to-Azure replication in a given billing month is 2 hours.
- C. The following Service Levels and Service Credits are applicable to Customer’ s use of each Protected Instance within the Site Recovery Service for Azure-to-Azure Failover.

Monthly Recovery Time Objective	Service Credit
> 2 hours	100%

SQL Database

Azure SQL Database is a fully managed relational database with built-in regional high availability and turnkey geo-replication to any Azure region. It includes intelligence to support self-driving features such as performance tuning, threat monitoring, and vulnerability assessments and provides fully automated patching and updating of the code base.

- Azure SQL Database Business Critical or Premium tiers configured as Zone Redundant Deployments have an availability guarantee of at least 99.995%.
- Azure SQL Database Business Critical or Premium tiers not configured for Zone Redundant Deployments, General Purpose, Standard, or Basic tiers, or Hyperscale tier with two or more replicas have an availability guarantee of at least 99.99%.
- Azure SQL Database Hyperscale tier with one replica has an availability guarantee of at least 99.95% and 99.9% for zero replicas.

i. **Additional Definitions**

- A. "**Availability Zone**" is a fault-isolated area within an Azure region, providing redundant power, cooling, and networking.
- B. "**Database**" means any Azure SQL Database created in any of the Service tiers and deployed either as a single database or in an Elastic Pool or Managed Instance.
- C. "**Zone Redundant Deployment**" is a Database that includes multiple synchronized replicas provisioned in different Availability Zones.

ii. **Monthly Uptime Calculation and Service Levels for Azure SQL Database Service**

- A. "**Deployment Minutes**" is the total number of minutes that a given Database has been operational in Azure during a billing month.
- B. "**Maximum Available Minutes**" is the sum of all Deployment Minutes for a given Microsoft Azure subscription during a billing month.
- C. "**Downtime**" is the total accumulated Deployment Minutes across all Databases in a given Azure subscription during which the Database is unavailable. A minute is considered unavailable for a given Database if all continuous attempts by Customer to establish a connection to the Database within the minute fail.
- D. "**Monthly Uptime Percentage**" for a given Database is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in a billing month for a given Azure subscription. Monthly Uptime Percentage is represented by the following formula:

$$\text{Monthly Uptime \%} = 100 * (\text{Maximum Available Minutes} - \text{Downtime}) / \text{Maximum Available Minutes}$$

- E. The following Service Levels and Service Credits are applicable to Customer' s use of the Business critical or Premium tiers of the SQL Database Service configured for Zone Redundant Deployments:

Monthly Uptime Percentage	Service Credit
<99.995%	10%
<99%	25%

F. The following Service Levels and Service Credits are applicable to Customer' s use of the Business critical or Premium tiers of the SQL Database Service not configured for Zone Redundant Deployments:

Monthly Uptime Percentage	Service Credit
<99.99%	10%
<99%	25%

G. The following Service Levels and Service Credits are applicable to Customer' s use of the General purpose, Standard or Basic tiers of the SQL Database Service:

Monthly Uptime Percentage	Service Credit
<99.99%	10%
<99%	25%

H. The following Service Levels and Service Credits are applicable to Customer' s use of the Hyperscale tier of the SQL Database Service not configured for Zone Redundant Deployments:

Provisioned Replicas	Monthly Uptime Percentage	Service Credit
0	< 99.9%	10%
	< 99%	25%

Provisioned Replicas	Monthly Uptime Percentage	Service Credit
1	< 99.95%	10%
	< 99%	25%
2+	< 99.99%	10%
	< 99%	25%

SQL Data Warehouse Service

We guarantee that, at least 99.9% of the time client operations executed on a SQL Data Warehouse database will succeed.

Additional Definitions

“Database” means any SQL Data Warehouse Database.

“Maximum Available Minutes” is the total number of minutes that a given Database has been deployed in Azure during a billing month in a given Azure subscription.

“Client Operations” is the set of all documented operations supported by SQL Data Warehouse.

“Downtime” is the total accumulated minutes during a billing month for a given Azure subscription during which a given Database is unavailable. A minute is considered unavailable for a given Database if more than 1% of all Client Operations completed during the minute return an Error Code.

“Monthly Uptime Percentage” for a given Database is calculated as Maximum Available Minutes less Downtime, divided by Maximum Available Minutes in a billing month in a given Azure subscription. Monthly Uptime Percentage is represented by the following formula:

$$\text{Monthly Uptime \%} = (\text{Maximum Available Minutes} - \text{Downtime}) / \text{Maximum Available Minutes}$$

The following Service Levels and Service Credits are applicable to Customer’ s use of the Basic, Standard, and Premium Tiers of the Azure SQL Database Service:

MONTHLY UPTIME PERCENTAGE	SERVICE CREDIT
< 99.9%	10%
< 99%	25%

Storage Service

- We guarantee that at least 99.99% (99.9% for Cool Access Tier) of the time, we will successfully process requests to read data from Read Access-Geo Redundant Storage (RA-GRS) Accounts, provided that failed attempts to read data from the primary region are retried on the secondary region.
- We guarantee that at least 99.9% (99% for Cool Access Tier) of the time, we will successfully process requests to read data from Locally Redundant Storage (LRS) and Geo Redundant Storage (GRS) Accounts.
- We guarantee that at least 99.9% (99% for Cool Access Tier) of the time, we will successfully process requests to write data to Locally Redundant Storage (LRS) and Geo Redundant Storage (GRS) Accounts and Read Access-Geo Redundant Storage (RA-GRS) Accounts.

Additional Definitions

"Average Error Rate" for a billing month is the sum of Error Rates for each hour in the billing month divided by the total number of hours in the billing month.

"Blob Storage Account" is a storage account specialized for storing data as blobs and provides the ability to specify an access tier indicating how frequently the data in that account is accessed.

"Block Blob Storage Account" is a storage account specialized for storing data as block or append blobs on solid-state drives.

"Cool Access Tier" is an attribute of a Blob Storage Account indicating that the data in the account is infrequently accessed and has a lower availability service level than data in other access tiers.

"Excluded Transactions" are storage transactions that do not count toward either Total Storage Transactions or Failed Storage Transactions. Excluded Transactions include pre-authentication failures; authentication failures; attempted transactions for storage accounts over their prescribed quotas; creation or deletion of containers, file shares, tables, or queues; clearing of queues; and copying blobs between storage accounts.

"Error Rate" is the total number of Failed Storage Transactions divided by the Total Storage Transactions during a set time interval (currently set at one hour). If the Total Storage Transactions in a given one-hour interval is zero, the error rate for that interval is 0%.

"Failed Storage Transactions" is the set of all storage transactions within Total Storage Transactions that are not completed within the Maximum Processing Time associated with their respective transaction type, as specified in the table below. Maximum Processing Time includes only the time spent processing a transaction request within the Storage Service and does not include any time spent transferring the request to or from the Storage Service.

REQUEST TYPES	MAXIMUM PROCESSING TIME
PutBlob and GetBlob (includes blocks and pages) Get Valid Page Blob Ranges	Two (2) seconds multiplied by the number of MBs transferred in the course of processing the request

REQUEST TYPES	MAXIMUM PROCESSING TIME
PutBlob and GetBlob (includes blocks and pages) Get Valid Page Blob Ranges	Two (2) seconds multiplied by the number of MBs transferred in the course of processing the request
Copy Blob	Ninety (90) seconds (where the source and destination blobs are within the same storage account)
Copy File	Ninety (90) seconds (where the source and destination files are within the same storage account).
PutBlockList	Sixty (60) seconds
GetBlockList	
Table Query List Operations	Ten (10) seconds (to complete processing or return a continuation)
Batch Table Operations	Thirty (30) seconds
All Single Entity Table Operations All other Blob and Message Operations	Two (2) seconds

These figures represent maximum processing times. Actual and average times are expected to be much lower.

Failed Storage Transactions do not include:

- 1.Transaction requests that are throttled by the Storage Service due to a failure to obey appropriate back-off principles.
- 2.Transaction requests having timeouts set lower than the respective Maximum Processing Times specified above.
- 3.Read transactions requests to RA-GRS Accounts for which you did not attempt to execute the request against Secondary Region associated with the storage account if the request to the Primary Region was not successful.
- 4.Read transaction requests to RA-GRS Accounts that fail due to Geo-Replication Lag.

"Geo Replication Lag" for GRS and RA-GRS Accounts is the time it takes for data stored in the Primary Region of the storage account to replicate to the Secondary Region of the storage account. Because GRS and RA-GRS Accounts are replicated asynchronously to the Secondary Region, data written to the Primary Region of the storage account will not be immediately available in the Secondary Region. You can query the Geo Replication Lag for a storage account, but [Azure](#) does not provide any guarantees as to the length of any Geo Replication Lag under this SLA.

"Geographically Redundant Storage (GRS) Account" is a storage account for which data is replicated synchronously within a Primary Region and then replicated asynchronously to a Secondary Region. You cannot directly read data from or write data to the Secondary Region associated with GRS Accounts.

"Locally Redundant Storage (LRS) Account" is a storage account for which data is replicated synchronously only within a Primary Region.

"Primary Region" is a geographical region in which data within a storage account is located, as selected by you when creating the storage account. You may execute write requests only against data stored within the Primary Region associated with storage accounts.

"Read Access Geographically Redundant Storage (RA-GRS) Account" is a storage account for which data is replicated synchronously within a Primary Region and then replicated asynchronously to a Secondary Region. You can directly read data from, but cannot write data to, the Secondary Region associated with RA-GRS Accounts.

"Secondary Region" is a geographical region in which data within a GRS or RA-GRS Account is replicated and stored, as assigned by Azure based on the Primary Region associated with the storage account. You cannot specify the Secondary Region associated with storage accounts.

"Total Storage Transactions" is the set of all storage transactions, other than Excluded Transactions, attempted within a one-hour interval across all storage accounts in the Storage Service in a given subscription.

Monthly Uptime Percentage: Monthly Uptime Percentage is calculated using the following formula:

100% - Average Error Rate

Service Credit – hot blobs in LRS, GRS and RA-GRS (write requests) Accounts, and blobs in LRS Block Blob Storage Accounts:

MONTHLY UPTIME PERCENTAGE	SERVICE CREDIT
< 99.9%	10%
< 99%	25%

Service Credit – Hot Blobs in RA-GRS (read requests) Accounts:

MONTHLY UPTIME PERCENTAGE	SERVICE CREDIT
< 99.99%	10%
< 99%	25%

Service Credit – Cool Blobs in LRS, GRS, RA-GRS (write requests) Accounts:

MONTHLY UPTIME PERCENTAGE	SERVICE CREDIT
< 99%	10%
< 98%	25%

Service Credit – Cool Blobs in RA-GRS (read requests) Accounts:

MONTHLY UPTIME PERCENTAGE	SERVICE CREDIT
< 99.9%	10%
< 98%	25%

StorSimple Service

i. **Additional Definitions**

- A. "Backup" is the process of backing up data stored on a registered StorSimple device to one or more associated cloud storage accounts within Azure.
- B. "Cloud Tiering" is the process of transferring data from a registered StorSimple device to one or more associated cloud storage accounts within Azure.

- C. "Failure" means the inability to fully complete a properly configured Backup, Tiering, or Restoring operation due to unavailability of the StorSimple Service.
 - D. "Managed Item" refers to a volume that has been configured to Backup to the cloud storage accounts using the StorSimple Service.
 - E. "Restoring" is the process of copying data to a registered StorSimple device from its associated cloud storage account(s).
- ii. **Monthly Uptime Calculation and Service Levels for StorSimple Service**
- A. "Deployment Minutes" is the total number of minutes during which a Managed Item has been configured for Backup or Cloud Tiering to a StorSimple storage account in Azure.
 - B. "Maximum Available Minutes" is the sum of all Deployment Minutes across all Managed Items for a given Azure subscription during a billing month.
 - C. "Downtime" is the total accumulated Deployment Minutes across all Managed Items configured for Backup or Cloud Tiering by Customer in a given Azure subscription during which the StorSimple Service is unavailable for the Managed Item. The StorSimple Service is considered unavailable for a given Managed Item from the first Failure of a Backup, Cloud Tiering, or Restoring operation with respect to the Managed Item until the initiation of a successful Backup, Cloud Tiering, or Restoring operation of the Managed Item, provided that retries are continually attempted no less frequently than once every thirty minutes.
 - D. "Monthly Uptime Percentage" for the StorSimple Service is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in a billing month for a given Azure subscription. Monthly Uptime Percentage is represented by the following formula:
 - E. The following Service Levels and Service Credits are applicable to Customer' s use of the StorSimple Service:

Monthly Uptime Percentage	Service Credit
<99.9%	10%
<99%	25%

Service Fabric

- Service Fabric is a free service, therefore, it does not have a financially backed SLA itself. The availability of your Service Fabric cluster is based on the SLA of the underlying virtual machines and storage resources used. Please see the [Virtual Machine SLA](#) and [Storage SLA](#) for more details.

SQL Server Stretch Database

i. **SLA for SQL Server Stretch Database**

We guarantee at least 99.9% of the time customers will have connectivity between their SQL Server Stretch Database and our Internet gateway.

ii. **SLA details**

iii. **Additional Definitions**

- "Database" means one instance of SQL Server Stretch database.
- Monthly Uptime Calculation and Service Levels for SQL Server Stretch Database Service
- "Maximum Available Minutes" is the total number of minutes that a given Database has been deployed in Azure for a given Azure subscription during a billing month.
- "Downtime" is the total accumulated minutes across all Databases deployed by Customer in a given Azure subscription during which the Database is unavailable. A minute is considered unavailable for a given Database if all

continuous attempts by Customer to establish a connection to the Database within the minute fail.

- E. "Monthly Uptime Percentage" for the SQL Server Stretch Database Service is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in a billing month for a given Azure subscription. Monthly Uptime Percentage is represented by the following formula:

$$\text{Monthly Uptime \%} = (\text{Maximum Available Minutes} - \text{Downtime}) / \text{Maximum Available Minutes}$$

- F. The following Service Levels and Service Credits are applicable to Customer' s use of the Azure SQL Stretch Database Service:

Monthly Uptime Percentage	Service Credit
<99.9%	10%
<99%	25%

Traffic Manager Service

iv. **Additional Definitions**

- A. "Traffic Manager Profile" or "Profile" refers to a deployment of the Traffic Manager Service created by Customer containing a domain name, endpoints, and other configuration settings, as represented in the Management Portal.
- B. "Valid DNS Response" means a DNS response, received from at least one of the Traffic Manager Service name server clusters, to a DNS request for the domain name specified for a given Traffic Manager Profile.

v. **Monthly Uptime Calculation and Service Levels for Traffic Manager Service**

- A. "Deployment Minutes" is the total number of minutes that a given Traffic Manager Profile has been deployed in Azure during a billing month.

- B. "Maximum Available Minutes" is the sum of all Deployment Minutes across all Traffic Manager Profiles deployed by Customer in a given Azure subscription during a billing month.
- C. "Downtime" is the total accumulated Deployment Minutes, across all Profiles deployed by Customer in a given Azure subscription, during which the Profile is unavailable. A minute is considered unavailable for a given Profile if all continual DNS queries for the DNS name specified in the Profile that are made throughout the minute do not result in a Valid DNS Response within two seconds.
- D. "Monthly Uptime Percentage" for the Traffic Manager Service is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in a billing month for a given Azure subscription. Monthly Uptime Percentage is represented by the following formula:
- E. The following Service Levels and Service Credits are applicable to Customer' s use of the Traffic Manager Service:

Monthly Uptime Percentage	Service Credit
<99.99%	10%
<99%	25%

Key Vault

- i. We guarantee that we will process Key Vault transactions within 5 seconds at least 99.9% of the time.
- ii. Monthly Uptime Calculation and Service Levels for Key Vault

"**Deployment Minutes**" is the total number of minutes that a given key vault has been deployed in Azure during a billing month.

"**Maximum Available Minutes**" is the sum of all Deployment Minutes across all key vaults deployed by Customer in a given Azure subscription during a billing month.

"**Excluded Transactions**" are transactions for creating, updating, or deleting key vaults, keys, or secrets.

"**Downtime**" is the total accumulated Deployment Minutes, across all key vaults deployed by Customer in a given Azure subscription, during which the key vault is unavailable. A minute is considered unavailable for a given key vault if all continuous attempts to perform transactions, other than Excluded Transactions, on the key vault throughout the minute either return an Error Code or do not result in a Success Code within 5 seconds from 21Vianet's receipt of the request.

"**Monthly Uptime Percentage**" for the Key Vault Service is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in a billing month for a given Azure subscription. Monthly Uptime Percentage is represented by the following formula:

Monthly Uptime % = (Maximum Available Minutes - Downtime) / Maximum Available Minutes

The following Service Levels and Service Credits are applicable to Customer's use of the Key Vault Service:

Monthly Uptime Percentage	Service Credit
<99.9%	10%
<99%	25%

Virtual Machines

For all Virtual Machines that have two or more instances deployed in the same Availability Set, we guarantee you will have Virtual Machine Connectivity to at least one instance at least 99.95% of the time.

For any Single Instance Virtual Machine using premium storage for all disks, we guarantee you will have Virtual Machine Connectivity of at least 99.9%.

i. **Additional Definitions**

- A. "Availability Set" refers to two or more Virtual Machines deployed across different Fault Domains to avoid a single point of failure.
- B. "Fault Domain" is a collection of servers that share common resources such as power and network connectivity.
- C. "Virtual Machine" refers to persistent instance types that can be deployed individually or as part of an Availability Set.
- D. "Virtual Machine Connectivity" is bi-directional network traffic between the virtual machine and other IP addresses using TCP or UDP network protocols in which the virtual machine is configured for allowed traffic. The IP addresses can be IP addresses in the same Cloud Service as the virtual machine, IP addresses within the same virtual network as the virtual machine or public, routable IP addresses.
- E. "Announced Single Instance Maintenance" means periods of Downtime related to network, hardware, or Service maintenance or upgrades impacting Single Instances. We will publish notice or notify you at least five (5) days prior to the commencement of such Downtime.
- F. "Data Disk" is a persistent virtual hard disk, attached to a Virtual Machine, used to store application data.
- G. "Operating System Disk" is a persistent virtual hard disk, attached to a Virtual Machine, used to store the Virtual Machine's operating system.
- H. "Single Instance" is defined as any single Virtual Machine that either is not deployed in an Availability Set or has only one instance deployed in an Availability Set.

- ii. Monthly Uptime Calculation and Service Levels for Virtual Machines in an Availability Set

- A. "Maximum Available Minutes" is the total accumulated minutes during a billing month for all Virtual Machines that have two or more instances deployed in the same Availability Set. Maximum Available Minutes is measured from when at least two Virtual Machines in the same Availability Set have both been started resultant from action initiated by Customer to the time Customer has initiated an action that would result in stopping or deleting the Virtual Machines.
- B. "Downtime" is the total accumulated minutes that are part of Maximum Available Minutes that have no Virtual Machine Connectivity.
- C. "Monthly Uptime Percentage" for Virtual Machines is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in a billing month for a given Azure subscription. Monthly Uptime Percentage is represented by the following formula:

$$\text{Monthly Uptime \%} = (\text{Maximum Available Minutes} - \text{Downtime}) / \text{Maximum Available Minutes} \times 100$$
- D. The following Service Levels and Service Credits are applicable to Customer's use of Virtual Machines in an Availability Set:

Monthly Uptime Percentage	Service Credit
<99.95%	10%
<99%	25%
<95%	100%

iii. Monthly Uptime Calculation and Service Levels for Single-Instance Virtual Machines

A. "Minutes in the Month" is the total number of minutes in a given month.

B. "Downtime" is the total accumulated minutes that are part of Minutes in the Month that have no Virtual Machine Connectivity. Downtime excludes Announced Single Instance Maintenance.

C."Monthly Uptime Percentage" is calculated by subtracting from 100% the percentage of Minutes in the Month in which any Single Instance Virtual Machine using premium storage for all Operating System Disks and Data disks had Downtime.

$$\text{Monthly Uptime \%} = (\text{Minutes in the Month} - \text{Downtime}) / \text{Minutes in the Month} \times 100$$

E.The following Service Levels and Service Credits are applicable to Customer' s use of Single-Instance Virtual Machines:

Monthly Uptime Percentage	Service Credit
<99.9%	10%
<99%	25%
<95%	100%

VPN Gateway

We guarantee 99.9% availability for each Basic Gateway for VPN or Basic Gateway for ExpressRoute.

We guarantee 99.95% availability for all Gateway for VPN SKUs excluding Basic.

We guarantee 99.95% availability for all Gateway for ExpressRoute SKUs excluding Basic.

i. Monthly Uptime Calculation and Service Levels for VPN Gateway

- A. "Maximum Available Minutes" is the total accumulated minutes during a billing month during which a given VPN Gateway has been deployed in an Azure subscription
- B. "Downtime" is the total accumulated Maximum Available Minutes during which a VPN Gateway is unavailable. A minute is considered unavailable if all attempts to connect to the VPN Gateway within a thirty-second window within the minute are unsuccessful.

C. "Monthly Uptime Percentage" for a given VPN Gateway is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in a billing month for the VPN Gateway. Monthly Uptime Percentage is represented by the following formula:

$$\text{Monthly Uptime \%} = (\text{Maximum Available Minutes} - \text{Downtime}) / \text{Maximum Available Minute} \times 100$$

D. The following Service Levels and Service Credits are applicable to Customer's use of each VPN Gateway:

Basic Gateway for VPN or ExpressRoute

Monthly Uptime Percentage	Service Credit
<99.9%	10%
<99%	25%

Gateway for VPN and Gateway for ExpressRoute SKUs excluding Basic

Monthly Uptime Percentage	Service Credit
<99.95%	10%
<99%	25%

App Service

We guarantee that Apps running in a customer subscription will be available 99.95% of the time. No SLA is provided for Apps under either the Free or Shared tiers.

i. **Additional Definitions**

A. "**Deployment Minutes**" is the total number of minutes that a given App has been set to running in Azure during a billing month. Deployment Minutes is measured from when the App was created or Customer initiated an action that would result in running the App to the time Customer initiated an action that would result in stopping or deleting the App.

- B. **“Maximum Available Minutes”** is the sum of all Deployment Minutes across all Apps deployed by Customer in a given Azure subscription during a billing month.
- C. **“App”** is a Web App, Mobile App, API App or Logic App deployed by Customer within the App Service, excluding apps in the Free and Shared tiers.
- D. **“Downtime”** the total accumulated Deployment Minutes, across all Apps deployed by Customer in a given Azure subscription, during which the App is unavailable. A minute is considered unavailable for a given App when there is no connectivity between the App and Microsoft’s Internet gateway. The Monthly Uptime Percentage is represented by the following formula:
- E. The following Service Levels and Service Credits are applicable to Customer’s use of the Websites Service:

Monthly Uptime Percentage	Service Credit
<99.95%	10%
<99%	25%

Additional Terms: Service Credits are applicable only to fees attributable to your use of Web Apps, Mobile Apps, API apps or Logic Apps and not to fees attributable to other types of apps available through the App Service, which are not covered by this SLA.

Virtual Machine Scale Sets

Virtual Machine Scale Sets is a free service, therefore, it does not have a financially backed SLA itself. However, if the Virtual Machine Scale Sets includes Virtual Machines in at least 2 Fault Domains, the availability of the underlying Virtual Machines SLA applies. See the [Virtual Machines SLA](#) for more details.

Managed Disks

Managed Disks does not have a financially backed SLA itself. The availability of Managed Disks is based on the SLA of the underlying storage used and virtual machine to which it is attached. Please see the [Virtual Machines](#) and [storage SLA](#) for more details.

Cognitive Services

We guarantee that Cognitive Services will be available at least 99.9% of the time. No SLA is provided for the Free tier.

ii. **Additional Definitions**

- A. **"Total Transaction Attempts"** is the total number of authenticated API requests by Customer during a billing month for a given Cognitive Services API. Total Transaction Attempts do not include API requests that return an Error Code that are continuously repeated within a five-minute window after the first Error Code is received.
- B. **"Failed Transactions"** is the set of all requests to the Cognitive Services API within Total Transaction Attempts that return an Error Code. Failed Transaction Attempts do not include API requests that return an Error Code that are continuously repeated within a five-minute window after the first Error Code is received.

iii. **Monthly Uptime Calculation**

- A. **"Monthly Uptime Percentage"** for each API Service is calculated as Total Transaction Attempts less Failed Transactions divided by Total Transaction Attempts in a billing month for a given API subscription. Monthly Uptime Percentage is represented by the following formula:
- B. $\text{Monthly Uptime \%} = (\text{Total Transaction Attempts} - \text{Failed Transactions}) / \text{Total Transaction Attempts} * 100$

- C. T2. The following Service Levels and Service Credits are applicable to Cognitive Services APIs:

Monthly Uptime Percentage	Service Credit
<99.9%	10%
<99%	25%

iv. **Service Level Exceptions**

- A. No SLA is provided for the Free tier.

Azure Analysis Services

We guarantee that, at least 99.9% of the time Client Operations executed on an Azure Analysis Services server will succeed.

i. **Additional Definitions**

- A. "**Server**" means any Azure Analysis Services server.
- B. "**Maximum Available Minutes**" is the total number of minutes that a given Server has been deployed in Azure during a billing month in a given Azure subscription.
- C. "**Client Operations**" is the set of all documented operations supported by Azure Analysis Services.
- D. "**Downtime**" is the total accumulated minutes during a billing month for a given Azure subscription during which a given Server is unavailable. A minute is considered unavailable for a given Server if more than 1% of all Client Operations completed during the minute return an Error Code.
- E. "**Monthly Uptime Percentage**" for a given Server is calculated as Maximum Available Minutes less Downtime, divided by Maximum Available Minutes in a billing month in a given Azure subscription. Monthly Uptime Percentage is represented by the following formula:

Monthly Uptime % = (Maximum Available Minutes – Downtime) / Maximum Available Minutes

F. The following Service Levels and Service Credits are applicable to Customer's use of the Analysis Services Service:

Monthly Uptime Percentage	Service Credit
<99.9%	10%
<99%	25%

Azure Cosmos DB

Azure Cosmos DB, operated by 21Vianet in China, is a distributed multi-model database service. It offers turnkey distribution across any number of Azure regions in China by transparently scaling and replicating your data wherever your users are. The service offers comprehensive 99.99% SLAs which covers the guarantees for throughput, consistency, availability and latency for the Cosmos DB Database Accounts scoped to a single Azure region configured with any of the five Consistency Levels or Database Accounts spanning multiple Azure regions, configured with any of the four relaxed Consistency Levels. Furthermore, independent of the choice of a Consistency Level, Cosmos DB offers 99.999% SLA for read availability for Database Accounts spanning two or more Azure regions.

i. **Additional Definitions**

- A. "**Collection**" is a container of JSON documents, and a unit of scale for transactions and queries.
- B. "**Consumed RUs**" is the sum of the Request Units consumed by all the requests which are processed by the Azure Cosmos DB Collection in a given second.
- C. "**Database Account**" is the top-level resource of the Azure Cosmos DB resource model. An Azure Cosmos DB Database Account contains one or more databases.

D. **"Failed Requests"** are requests within Total Requests that either return an Error Code or fail to return a Success Code within the maximum upper bounds documented in the table below.

E. **"Failed Read Requests"** are requests within Total Read Requests that either return an Error Code or fail to return a Success Code within the maximum upper bounds documented in the table below.

OPERATION	MAXIMUM UPPER BOUND ON PROCESSING LATENCY
All Database Account configuration operations	2 Minutes
Add a new Region	60 Minutes
Manual Failover	5 Minutes
Resource Operations	5 Sec
Media Operations	60 Sec

F. **"Provisioned RUs"** is the total provisioned Request Units for a given Azure Cosmos DB Collection for a given second.

G. **"Rate Limited Requests"** are requests which are throttled by the Azure Cosmos DB Collection after Consumed RUs have exceeded the Provisioned RUs for a partition in the Collection for a given second.

- H. "**Request Unit (RU)**" is a measure of throughput in Azure Cosmos DB.
 - I. "**Resource**" is a set of URI addressable entities associated with a Database Account.
 - J. "**Successful Requests**" are Total Requests minus Failed Requests.
 - K. "**Total Requests**" is the set of all requests, including Rate Limited Requests and all Failed Requests, issued against Resources within a one-hour interval within a given Azure subscription during a billing month.
 - L. "**Total Read Requests**" is the set of all the read requests, including Rate Limited Requests and all the Failed Read Requests, issued against Resources within a one-hour interval within a given Azure subscription during a billing month.
- ii. **Availability SLA**
- A. "**Read Error Rate**" is the total number of Failed Read Requests divided by Total Read Requests, across all Resources in a given Azure subscription, during a given one-hour interval. If the Total Read Requests in a given one-hour interval is zero, the Read Error Rate for that interval is 0%.
 - B. "**Error Rate**" is the total number of Failed Requests divided by Total Requests, across all Resources in a given Azure subscription, during a given one-hour interval. If the Total Requests in a given one-hour interval is zero, the Error Rate for that interval is 0%.
 - C. "**Average Error Rate**" for a billing month is the sum of Error Rates for each hour in the billing month divided by the total number of hours in the billing month
 - D. "**Average Read Error Rate**" for a billing month is the sum of Read Error Rates for each hour in the billing month divided by the total number of hours in the billing month.
 - E. "**Monthly Availability Percentage**" for the Azure Cosmos DB Service deployed via Database Accounts scoped to a single Azure region configured with any of the five Consistency Levels or Database Accounts spanning multiple regions, configured with any of the four relaxed Consistency Levels

is calculated by subtracting from 100% the Average Error Rate for a given Azure subscription in a billing month. Monthly Availability Percentage is represented by the following formula:

$$\text{Monthly Uptime \%} = 100\% - \text{Average Error Rate}$$

Service Credit:

MONTHLY AVAILABILITY PERCENTAGE	SERVICE CREDIT
< 99.99%	10%
<99%	25%

- F. **"Monthly Availability Percentage"** for the Azure Cosmos DB Service deployed via Database Accounts scoped to a single Azure region configured with any of the five Consistency Levels or Database Accounts spanning multiple regions, configured with any of the four relaxed Consistency Levels is calculated by subtracting from 100% the Average Error Rate for a given Azure subscription in a billing month. Monthly Availability Percentage is represented by the following formula:

$$\text{Monthly Read Availability Uptime \%} = 100\% - \text{Average Read Error Rate}$$

Service Credit:

MONTHLY READ AVAILABILITY PERCENTAGE	SERVICE CREDIT
< 99.999%	10%
<99%	25%

- G. **"Monthly Multiple Write Locations Availability Percentage"** for the Azure Cosmos DB Service deployed via Database Accounts configured to span multiple Azure regions with multiple writable locations, is calculated by subtracting from 100% the Average Error Rate for a given Azure subscription

in a billing month. Monthly Availability Percentage is represented by the following formula:

$$\text{Monthly Uptime \%} = 100\% - \text{Average Error Rate}$$

Service Credit:

MONTHLY READ AVAILABILITY PERCENTAGE	SERVICE CREDIT
< 99.999%	10%
<99%	25%

iii. **Throughput SLA**

- A. **"Throughput Failed Requests"** are requests which are throttled by the Azure Cosmos DB Collection resulting in an Error Code, before Consumed RUs have exceeded the Provisioned RUs for a partition in the Collection for a given second.
- B. **"Error Rate"** is the total number of Throughput Failed Requests divided by Total Requests, across all Resources in a given Azure subscription, during a given one-hour interval. If the Total Requests in a given one-hour interval is zero, the Error Rate for that interval is 0%.
- C. **"Average Error Rate"** for a billing month is the sum of Error Rates for each hour in the billing month divided by the total number of hours in the billing month.
- D. **"Monthly Throughput Percentage"** for the Azure Cosmos DB Service is calculated by subtracting from 100% the Average Error Rate for a given Azure subscription in a billing month. Monthly Throughput Percentage is represented by the following formula:

$$\text{Monthly Throughput \%} = 100\% - \text{Average Error Rate}$$

E. Service Credit:

MONTHLY READ AVAILABILITY PERCENTAGE	SERVICE CREDIT
< 99.99%	10%
<99%	25%

iv. **Consistency SLA**

- A. " **K** " is the number of versions of a given document for which the reads lag behind the writes.
- B. " **T** " is a given time interval.
- C. " **Consistency Level** " is the setting for a particular read request that supports consistency guarantees. The following table captures the guarantees associated with the Consistency Levels. Note that Session, Bounded Staleness, Consistent Prefix and Eventual Consistency Levels are all referred to as "relaxed" .

CONSISTENCY LEVEL	CONSISTENCY GUARANTEES
Strong	Linearizability
Session	Read Your Own Write (Within Write Region)
	Monotonic Read
	Consistent Prefix

CONSISTENCY LEVEL	CONSISTENCY GUARANTEES
Bounded Staleness	Read Your Own Write (Within Write Region)
	Monotonic Read (Within a Region)
	Consistent Prefix
	Staleness Bound < K,T >
Consistent Prefix	Consistent Prefix
Eventual	Eventual

- D. "**Consistency Violation Rate**" is Successful Requests that could not be delivered when performing the consistency guarantees specified for the chosen Consistency Level divided by Total Requests, across all Resources in a given Azure subscription, during a given one-hour interval. If the Total Requests in a given one-hour interval is zero, the Consistency Violation Rate for that interval is 0%.
- E. "**Average Consistency Violation Rate**" for a billing month is the sum of Consistency Violation Rates for each hour in the billing month divided by the total number of hours in the billing month.

- F. **"Monthly Consistency Attainment Percentage"** for the Azure Cosmos DB service is calculated by subtracting from 100% the Average Consistency Violation Rate for a given Azure Subscription in a billing month.

$$\text{Monthly Consistency \%} = 100\% - \text{Average Consistency Violation Rate}$$

- G. Service Credit:

MONTHLY CONSISTENCY ATTAINMENT PERCENTAGE	SERVICE CREDIT
< 99.99%	10%
<99%	25%

- v. **Latency SLA**

- A. **"Application"** is an Azure Cosmos DB application deployed within a local Azure region with accelerated networking enabled and using the Azure Cosmos DB client SDK configured with TCP direct connectivity for a given Azure subscription in a billing month.
- B. **"N"** is the number of Successful Requests for a given Application performing either document read or document write operations with a payload size less than or equal to 1 KB in a given hour.
- C. **"S"** is the latency-sorted set of Successful Request response times in ascending order for a given Application performing document read or document write operations with a payload size less than or equal to 1 KB in a given hour.
- D. **"Ordinal Rank"** is the 99th percentile using the nearest rank method represented by the following formula:

$$\text{Ordinal Rank} = (99 / 100) * N$$
- E. **"P99 Latency"** is the value at the Ordinal Rank of S.
- F. **"Excessive Latency Hours"** is the total number of one-hour intervals during which Successful Requests submitted by an Application resulted in a P99 Latency greater than or equal to 10ms for document read or 10ms for

document write operations. If the number of Successful Requests in a given one-hour interval is zero, the Excessive Latency Hours for that interval is 0.

G. "**Average Excessive Latency Rate**" for a billing month is the sum of Excessive Latency Hours divided by the total number of hours in the billing month.

H. "**Monthly P99 Latency Attainment Percentage**" for a given Azure Cosmos DB Application deployed via Database Accounts scoped to a single Azure region configured with any of the five Consistency Levels or Database Accounts spanning multiple regions, configured with any of the four relaxed Consistency Levels is calculated by subtracting from 100% the Average Excessive Latency Rate for a given Azure subscription in a billing month. Monthly P99 Latency Attainment Percentage is represented by the following formula:

$$\text{Monthly P99 Latency Attainment \%} = 100\% - \text{Average Excessive Latency}$$

Rate

I. Service Credit:

MONTHLY P99 LATENCY ATTAINMENT PERCENTAGE	SERVICE CREDIT
< 99.99%	10%
<99%	25%

Network Watcher

We guarantee that 99.9% of the time Network Diagnostic Tools will successfully execute and return a response.

i. **Additional Definitions**

A. "**Network Diagnostic Tools**" is a collection of network diagnostic and topology tools.

ii. **Monthly Uptime Calculation and Service Levels for Network Diagnostic Tools**

- A. **“Maximum Diagnostic Checks”** is the total number of diagnostic actions performed by the Network Diagnostic Tool as configured by Customer in a billing month for a given Azure subscription.
- B. **“Failed Diagnostic Checks”** is the total number of diagnostic actions within Maximum Diagnostic Checks that returns an Error Code or does not return a response within the Maximum Processing Time documented in the table below.

DIAGNOSTIC TOOL	MAXIMUM PROCESSING TIME
IPFlow Verify NextHop Packet Capture Security Group View Topology	2 minutes
VPN Troubleshoot	10 minutes

- C. **“Monthly Uptime Percentage”** is calculated as Maximum Diagnostic Checks less Failed Diagnostic Checks divided by Maximum Diagnostic Checks.
- D. Monthly Uptime Percentage is represented by the following formula:
$$\text{Monthly Uptime \%} = (\text{Maximum Diagnostic Checks} - \text{Failed Diagnostic Checks}) / \text{Maximum Diagnostic Checks} \times 100$$
- E. The following Service Levels and Service Credits are applicable to Customer’ s use of the Network Diagnostic Tool

Monthly Uptime Percentage	Service Credit
<99.9%	10%
<99%	25%

Azure Monitor

Azure Monitor allows you to collect granular performance and utilization data, activity and diagnostics logs, and define alerts and notifications from your Azure resources in a consistent manner. We guarantee that 99.9% of the time notifications will be successfully delivered.

i. **Monthly Uptime Calculation and Service Levels for the Azure Monitor Alerts**

- A. **"Alert Rule"** is a collection of signal criteria used to generate alerts using monitoring event data already available to Alert Service for analysis.
- B. **"Maximum Available Minutes"** is the total number of minutes which Alert Rule(s) are deployed by Customer in a given Azure subscription during a billing month.
- C. **"Downtime"** is the total number of minutes within Maximum Available Minutes during which the Alert Rule is unavailable. A minute is considered unavailable for a given Alert Rule if all continuous attempts to analyze telemetry signals for resources defined within the Alert Rule throughout the minute either return an Error Code or do not result in a Success Code within five minutes from scheduled Alert Rule start time.
- D. **"Monthly Uptime Percentage"** is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes multiplied by 100. Monthly Uptime Percentage is represented by the following formula:

$$\text{Monthly Uptime \%} = (\text{Maximum Available Minutes} - \text{Downtime}) / \text{Maximum Available Minutes} \times 100$$
- E. The following Service Levels and Service Credits are applicable to Customer' s use of Azure Monitor Alerts:

Monthly Uptime Percentage	Service Credit
<99.9%	10%
<99%	25%

ii. **Monthly Uptime Calculation and Service Levels for the Azure Monitor Notification Delivery**

- A. **"Action Group"** is a collection of actions deployed by Customer in a given Azure subscription which defines preferred notification delivery methods.
- B. **"Deployment Minutes"** is the total number of minutes that a given Action Group has been deployed by Customer in Azure subscription during a billing month.
- C. **"Maximum Available Minutes"** is the sum of all Deployment Minutes across all Action Groups deployed by Customer in a given Azure subscription during a billing month.
- D. **"Downtime"** is the total accumulated Deployment Minutes, across all Action Groups, during which the Action Group is unavailable. A minute is considered unavailable for a given Action Group if all continuous attempts to send alerts or perform registration management operations with respect to the Action Group throughout the minute either return an Error Code or do not result in a Success Code within five minutes.
- E. **"Monthly Uptime Percentage"** is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in a billing month for a given Azure subscription. Monthly Uptime Percentage is represented by the following formula:

$$\text{Monthly Uptime \%} = (\text{Maximum Available Minutes} - \text{Downtime}) / \text{Maximum Available Minutes} \times 100$$
- F. The following Service Levels and Service Credits are applicable to Customer' s use of the Azure Monitor Notification Delivery:

Monthly Uptime Percentage	Service Credit
<99.9%	10%
<99%	25%

API Management

We guarantee that API Management Service instances running in the Basic, Standard and Premium tiers will respond to requests to perform operations at least 99.9% of the time.

No SLA is provided for the Developer tier of the API Management Service.

- A. **"Deployment Minutes"** is the total number of minutes that a given API Management instance has been deployed in Azure during a billing month.
- B. **"Maximum Available Minutes"** is the sum of all Deployment Minutes across all API Management instances deployed by Customer in a given Azure subscription during a billing month.
- C. **"Proxy"** is the component of the API Management Service responsible for receiving API requests and forwarding them to the configured dependent API.
- D. **"Downtime"** is the total number of minutes within Maximum Available Minutes during which the API Management Service is unavailable. A minute is considered unavailable for a given API Management instance if all continuous attempts to perform operations through the Proxy throughout the minute result in either an Error Code or do not return a Success Code within five minutes.
- E. **"Monthly Uptime Percentage"** is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes multiplied by 100.

Monthly Uptime % = (Maximum Available Minutes - Downtime) / Maximum Available Minutes x 100

i. **Service Credit for Basic, Standard and Premium Tier**

Monthly Uptime Percentage	Service Credit
<99.9%	10%
<99%	25%

Azure Database for MySQL

We guarantee at least 99.99% of the time customers will have connectivity between their Azure Database for MySQL Server and our Internet gateway.

- A. **"Server"** is any given Azure Database for MySQL server
- i. **Monthly Uptime Calculation and Service Levels for Azure Database for MySQL**
 - A. **"Maximum Available Minutes"** is the total number of minutes for a given Server deployed by Customer in a Azure subscription during a billing month.
 - B. **"Downtime"** is the total number of minutes within Maximum Available Minutes during which a Server is unavailable. A minute is considered unavailable if all continuous attempts by Customer to establish a connection to the Server returned an Error Code.
 - C. **"Monthly Uptime Percentage"** for the Azure Database for MySQL is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes.

Monthly Uptime Percentage is represented by the following formula:

$$\text{Monthly Uptime \%} = (\text{Maximum Available Minutes} - \text{Downtime}) / \text{Maximum Available Minutes} \times 100$$

- ii. **The following Service Levels and Service Credits are applicable to Customer' s use of the Azure Database for MySQL:**

Monthly Uptime Percentage	Service Credit
<99.99%	10%
<99%	25%

Azure Database for PostgreSQL

We guarantee at least 99.99% of the time customers will have connectivity between their Azure Database for PostgreSQL Server and our Internet gateway.

- A. **"Server" is any given Azure Database for PostgreSQL server**
- i. **Monthly Uptime Calculation and Service Levels for Azure Database for PostgreSQL**

- A. **"Maximum Available Minutes"** is the total number of minutes for a given Server deployed by Customer in a Azure subscription during a billing month.
- B. **"Downtime"** is the total number of minutes within Maximum Available Minutes during which a Server is unavailable. A minute is considered unavailable if all continuous attempts by Customer to establish a connection to the Server returned an Error Code.
- C. **"Monthly Uptime Percentage"** for the Azure Database for PostgreSQL is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes.

Monthly Uptime Percentage is represented by the following formula:

$$\text{Monthly Uptime \%} = (\text{Maximum Available Minutes} - \text{Downtime}) / \text{Maximum Available Minutes} \times 100$$

- ii. **The following Service Levels and Service Credits are applicable to Customer' s use of the Azure Database for PostgreSQL:**

Monthly Uptime Percentage	Service Credit
<99.99%	10%
<99%	25%

Functions

For Function Apps running on Consumption Plans we guarantee that the associated functions will run 99.95% of the time after trigger is successfully fired.

For Function Apps running on App Service Plans we guarantee that the associated Functions compute will be available 99.95% of the time.

i. **Additional Definitions**

A. **"Function App"** is a collection of one or more functions deployed with an associated trigger.

ii. **Monthly Uptime Calculation and Service Levels for Function App on Consumption Plan**

A. **"Total Triggered Executions"** is the total number of all Function App executions triggered by Customer in a given Azure subscription during a billing month.

B. **"Unavailable Executions"** is the total number of executions within Total Triggered Executions which failed to run. An execution failed to run when the given Function App history log did not capture any output five (5) minutes after the trigger is successfully fired.

C. **"Monthly Uptime Percentage"** for Function Apps on Consumption Plan is calculated as Total Triggered Executions less Unavailable Executions divided by Total Triggered Executions multiplied by 100.

$$\text{Monthly Uptime \%} = (\text{Total Triggered Executions} - \text{Unavailable Executions}) / (\text{Total Triggered Executions}) \times 100$$

The following Service Levels and Service Credits are applicable to Customer's use of Function App on Consumption Plan.

Monthly Uptime Percentage	Service Credit
<99.95%	10%
<99%	25%

iii. **Monthly Uptime Calculation and Service Levels for Function App on App Service Plan**

A. **"Deployment Minutes"** is the total number of minutes that a given Function App is available to be triggered during a billing month. Deployment Minutes are measured based on the total time that the service

is available to trigger a function execution and not based on the potential number of function executions that might be triggered during a given month.

- B. "**Maximum Available Minutes**" is the sum of all Deployment Minutes for a given Function App deployed by Customer in a given Azure subscription during a billing month.
- C. "**Downtime**" is the total number of minutes within Maximum Available Minutes, during which the Function App is unavailable to be triggered. A minute is considered unavailable for a given Function App when there is no connectivity between the App Service Plan on which the Function App is hosted and Microsoft's Internet gateway.
- D. "**Monthly Uptime Percentage**" for Functions on App Service Plan is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes multiplied by 100.

Monthly Uptime % = (Maximum Available Minutes-Downtime) / (Maximum Available Minutes) x 100

Monthly Uptime Percentage	Service Credit
<99.95%	10%
<99%	25%

Azure Active Directory B2C

The SLA will be effective from 7/1/2019.

We guarantee at least 99.9% availability of the Azure Active Directory B2C service. The service is considered available for a directory in the following scenarios:

The service is able to process user sign-up, sign-in, profile editing, password reset and multi-factor authentication requests.

Developers are able to create, read, write and delete entries in the directory.

No SLA is provided for the Free tier of Azure Active Directory B2C.

iv. **Additional Definitions**

- A. "**Deployment Minutes**" is the total number of minutes for which an Azure AD B2C directory has been deployed during a billing month.
- B. "**Maximum Available Minutes**" is the sum of all Deployment Minutes across all Azure AD B2C directories in a given Azure subscription during a billing month.
- C. "**Downtime**" is the total accumulated minutes across all Azure AD B2C directories deployed by Customer in a given Azure subscription during which the Azure AD B2C service is unavailable. A minute is considered unavailable if either all attempts to process user sign-up, sign-in, profile editing, password reset and multi-factor authentication requests, or all attempts by developers to create, read, write and delete entries in a directory, fails to return tokens or valid Error Codes, or do not return responses within two minutes.
- D. "**Monthly Uptime Percentage**" is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes in a billing month for the Azure subscription.

Monthly Uptime Percentage is represented by the following formula:

$$\text{Monthly Uptime \%} = (\text{Maximum Available Minutes} - \text{Downtime}) / \text{Maximum Available Minutes}$$

Monthly Uptime Percentage	Service Credit
<99.9%	10%
<99%	25%

Azure Database for MariaDB

We guarantee at least 99.99% of the time customers will have connectivity between their Azure Database for MariaDB Server and our Internet gateway.

i. **Additional Definitions**

- A. "**Server**" is any given Azure Database for MariaDB server.

ii. **Monthly Uptime Calculation and Service Levels for Azure Database for MariaDB**

- A. "**Maximum Available Minutes**" is the total number of minutes for a given Server deployed by Customer in a Azure subscription during a billing month.
- B. "**Downtime**" is the total number of minutes within Maximum Available Minutes during which a Server is unavailable. A minute is considered unavailable if all continuous attempts by Customer to establish a connection to the Server returned an Error Code.
- C. "**Monthly Uptime Percentage**" for the Azure Database for MariaDB is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes.

Monthly Uptime Percentage is represented by the following formula:

$$\text{Monthly Uptime \%} = (\text{Maximum Available Minutes} - \text{Downtime}) / \text{Maximum Available Minutes} * 100$$

The following Service Levels and Service Credits are applicable to Customer' s use of the Azure Database for MariaDB:

Monthly Uptime Percentage	Service Credit
<99.99%	10%
<99%	25%

Azure Kubernetes Service (AKS)

As a free service, AKS does not offer a financially-backed service level agreement. We will strive to attain at least 99.5% availability for the Kubernetes API server. The availability of the agent nodes in your cluster is covered by the Virtual Machines SLA. Please see the [Virtual Machines SLA](#) for more details.

Data Factory

We guarantee at least 99.9% of the time we will successfully process requests to perform operations against Data Factory resources.

We guarantee that at least 99.9% of the time, all activity runs will initiate within 4 minutes of their scheduled execution times.

i. **Additional Definitions**

A. "**Resources**" means integration runtimes (including Azure, SSIS and self-hosted Integration Runtimes), triggers, pipelines, data sets, and linked services created within a Data Factory.

B. "**Activity Run**" means the execution or attempted execution of an activity.

ii. **Monthly Uptime Calculation for Data Factory API Calls**

A. "**Total Requests**" is the set of all requests, other than Excluded Requests, to perform operations against Resources during a billing month for a given Azure subscription.

B. "**Excluded Requests**" is the set of requests that result in an HTTP 4xx status code, other than an HTTP 408 status code.

C. "**Failed Requests**" is the set of all requests within Total Requests that either return an Error Code or an HTTP 408 status code or otherwise fail to return a Success Code within two minutes.

D. "**Monthly Uptime Percentage**" for the API calls made to the Data Factory Service is calculated as Total Requests less Failed Requests divided by Total Requests in a billing month for a given Azure subscription.

Monthly Uptime Percentage is represented by the following formula:

$$\text{Monthly Uptime \%} = (\text{Total Requests} - \text{Failed Requests}) / \text{Total Requests} * 100$$

The following Service Credits are applicable to Customer' s use of API calls within the Data Factory Service:

Monthly Uptime Percentage	Service Credit
<99.9%	10%
<99%	25%

iii. **Monthly Uptime Calculation for Data Factory Activity Runs**

- A. **"Total Activity Runs"** is the total number of Activity Runs attempted during a given billing month for a given Azure subscription.
- B. **"Delayed Activity Runs"** is the total number of attempted Activity Runs in which an activity fails to begin executing within four (4) minutes after the time at which it is scheduled for execution and all dependencies that are prerequisite to execution have been satisfied.
- C. **"Monthly Uptime Percentage"** for the Data Factory Service is calculated as Total Activity Runs less Delayed Activity Runs divided by Total Activity Runs in a billing month for a given Azure subscription.

Monthly Uptime Percentage is represented by the following formula:

$$\text{Monthly Uptime \%} = (\text{Total Activity Runs} - \text{Delayed Activity Runs}) / \text{Total Activity Runs} * 100$$

The following Service Levels and Service Credits are applicable to Customer's Activity Runs within the Data Factory Service:

Monthly Uptime Percentage	Service Credit
<99.9%	10%
<99%	25%

Security Center

We guarantee 99.9% availability of a Protected Node's Security Monitoring information for the Standard tier of Azure Security Center. No SLA is provided for the Free tier of Azure Security Center.

i. **Additional Definitions**

- A. **"Protected Node"** is a Azure resource, counted as a node for billing purposes that is configured for the Azure Security Center Standard Tier.
- B. **"Security Monitoring"** is the assessment of a Protected Node resulting in potential findings such as security health status, recommendations, and security alerts, exposed in Azure Security Center.

- C. **"Maximum Available Minutes"** is the total number of minutes during a billing month that a given Protected Node has been deployed and configured for Security Monitoring.
- D. **"Downtime"** is the total accumulated minutes during a billing month for which Security Monitoring information of a given Protected Node is unavailable. A minute is considered unavailable for a given Protected Node if all continuous attempts to retrieve Security Monitoring information throughout the minute result in either an Error Code or do not return a Success Code within two minutes.
- E. **"Monthly Uptime Percentage"** for Azure Security Center of a given Protected Node in a given billing month is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes.

Monthly Uptime Percentage is represented by the following formula:

$$\text{Monthly Uptime \%} = (\text{Maximum Available Minutes} - \text{Downtime}) / \text{Maximum Available Minutes} * 100$$

The following Service Levels and Service Credits are applicable to Customer's use of each Protected Node:

Monthly Uptime Percentage	Service Credit
<99.9%	10%
<99%	25%

Service Level Exceptions: The Free Tier of the Azure Security Center is not covered by this SLA.

Azure Data Explorer

We guarantee at least 99.9% availability of the Azure Data Explorer Service. Formerly known as Kusto.

- i. **Additional Definitions**
 - A. **"Cluster"** means Azure Data Explorer (ADX) operated cluster.
- ii. **Monthly Uptime Calculation and Service Levels for Azure Data Explorer**

- A. "**Maximum Available Minutes**" is the total number of minutes for a given Cluster deployed by Customer in a Azure subscription during a billing month.
- B. "**Downtime**" is the total number of minutes within Maximum Available Minutes during which a Cluster is unavailable. A minute is considered unavailable for a given Cluster if all continuous attempts within the minute to establish a connection to the Cluster returned an Error Code.
- C. "**Monthly Uptime Percentage**" for the Azure Data Explorer is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes.

Monthly Uptime Percentage is represented by the following formula:

$$\text{Monthly Uptime \%} = (\text{Maximum Available Minutes} - \text{Downtime}) / \text{Maximum Available Minutes} * 100$$

The following Service Levels and Service Credits are applicable to Customer's use of the Service:

Monthly Uptime Percentage	Service Credit
<99.9%	10%
<99%	25%

Event Grid

We guarantee that Customers will be able to publish messages to Event Grid 99.99% of the time.

i. **Monthly Uptime Calculation and Service Levels for Event Grid**

- A. "**Maximum Available Minutes**" is the total number of minutes that an Event Grid has been deployed by Customer in a Azure subscription during a billing month.
- B. "**Downtime**" is the total number of minutes within Maximum Available Minutes across all Event Grids deployed by Customer in a given Azure subscription during which Event Grid is unavailable. A minute is considered

unavailable for a given Event Grid if all requests to publish a message either return an Error Code or do not result in a Success Code within one minute.

- C. **"Monthly Uptime Percentage"** for Event Grid is calculated as Maximum Available Minutes less Downtime divided by Maximum Available Minutes multiplied by 100.

Monthly Uptime Percentage is represented by the following formula:

$$\text{Monthly Uptime \%} = (\text{Maximum Available Minutes} - \text{Downtime}) / \text{Maximum Available Minutes} * 100$$

- D. The following Service Levels and Service Credit are applicable to Customer' s use of the Event Grid.

Monthly Uptime Percentage	Service Credit
<99.99%	10%
<99%	25%

Azure Firewall

Azure Firewall offers fully stateful native firewall capabilities for Virtual Network resources, with built-in high availability and the ability to scale automatically.

We guarantee that Azure Firewall will be available at least 99.95% of the time, when deployed within a single Availability Zone.

We guarantee that Azure Firewall will be available at least 99.99% of the time, when deployed within two or more Availability Zones in the same Azure region.

- i. **Additional Definitions**

- A. **"Azure Firewall Service" refers to a logical firewall instance deployed in a customer Virtual Network.**

- ii. **Monthly Uptime Calculation and Service Levels for the Azure Firewall Service deployed within a single Availability Zone**

- A. **"Maximum Available Minutes"** is the total accumulated minutes in a billing month during which the Azure Firewall Service has been deployed in an Azure subscription.

B. **"Downtime"** is the total accumulated Maximum Available Minutes in a billing month for a given Azure Firewall Service during which the Azure Firewall Service is unavailable. A given minute is considered unavailable if all attempts to connect to the Azure Firewall Service throughout the minute are unsuccessful.

C. **"Monthly Uptime Percentage"** The Monthly Uptime Percentage is calculated using the following formula:

$$\frac{(\text{Maximum Available Minutes} - \text{Downtime})}{\text{Maximum Available Minutes}} \times 100$$

D. The following Service Levels and Service Credits are applicable to Customer' s use of the Azure Firewall Service, when deployed within a single Availability Zone:

Monthly Uptime Percentage	Service Credit
<99.95%	10%
<99%	25%

iii. **Monthly Uptime Calculation and Service Levels for the Azure Firewall Service deployed within two or more Availability Zones**

A. **"Maximum Available Minutes"** is the total accumulated minutes in a billing month during which the Azure Firewall Service has been deployed within two or more Availability Zones in the same region in a subscription.

B. **"Downtime"** is the total accumulated Maximum Available Minutes in a billing month for a given Azure Firewall Service, deployed within two or more Availability Zones, during which the Azure Firewall Service is unavailable. A given minute is considered unavailable if all attempts to connect to the Azure Firewall Service throughout the minute are unsuccessful.

C. **"Monthly Uptime Percentage"** for Azure Firewalls deployed within two or more Availability Zones is calculated using the following formula:

$$\frac{(\text{Maximum Available Minutes} - \text{Downtime})}{(\text{Maximum Available Minutes})} \times 100$$

D. The following Service Levels and Service Credits are applicable to Customer' s use of the Azure Firewall Service, when deployed within two or more Availability Zones in the same region:

Monthly Uptime Percentage	Service Credit
<99.99%	10%
<99%	25%