NVIDIA CLOUD SERVICES - SERVICE LEVEL AGREEMENT (SLA)

INTRODUCTION

This Service Level Agreement ("SLA") is an integral part of the NVIDIA Cloud Agreement ("Agreement"). Capitalized terms used but not defined below have the meaning assigned to them in the Agreement.

The General Terms below apply to all Cloud Offerings covered under this SLA. For each NVIDIA Cloud Offering, service specific terms, such as availability levels, capacity levels (if applicable) and remedies are defined in the following sections of this SLA:

1. Page 3 - NVIDIA DGX Cloud
2. Page 4 - NVIDIA Omniverse Cloud
3. Page 5 - NVIDIA cuOpt Service
4. Page 6 - NVIDIA Cloud Functions (NVCF) Service

If NVIDIA does not achieve and maintain the Service Levels for a Cloud Offering as described in this SLA, then Customer may be eligible for a remedy as specified in each service specific section of this SLA. NVIDIA may modify the terms of the SLA at any time, provided that the SLA version in place at the commencement of Customer's Subscription will remain in effect with Customer for the duration of the Subscription.

GENERAL TERMS

Calculation of Service Availability

To calculate Service Availability, as available for a Cloud Offering, NVIDIA uses a combination of methods, including information from NVIDIA's event monitoring systems and logs and observing external causes that may impact the services (to the extent known to NVIDIA) and matching these findings with Customer reports to determine the actual loss time frame.

NVIDIA determines Service Availability over one calendar month and tracks status every 15 minutes. NVIDIA will not include any gaps shorter than 15 minutes in duration in its calculations. NVIDIA will not include in its Service Availability measurement an outage that has no impact to the Customer.

Service Availability Target Time to Restore Service.

For each Critical Service Availability issue reported by Customer via the designated report channels, NVIDIA will use commercially reasonable efforts to find a resolution within 24 hours from the time NVIDIA determines that the incident is Critical, with the understanding that this is a target timeframe and may not always be achieved. "Critical" means a production situation where the Cloud Offering (a) is totally inoperable or fails catastrophically and there is no workaround, (b) performance degrades substantially under reasonable loads causing a severe impact on use, (c) is operable but functionality is materially incomplete, (d) has one or more key functions or commands inoperable, or (e) experiences a material security issue as confirmed by NVIDIA that requires resolution. NVIDIA may provide a resolution with hardware or software, workarounds, other mitigations, or decision to close an open issue. The Exclusions listed below may impact the resolution.

Calculation of Capacity Availability

To calculate Capacity Availability, as available for a Cloud Offering, NVIDIA uses a combination of methods, including information from NVIDIA's event monitoring systems and logs and observing external causes that may impact the services (to the extent known to NVIDIA) and matching these findings with Customer reports to determine the actual loss time frame.

NVIDIA measures Capacity Availability over one calendar month and tracks every 60 minutes. NVIDIA will not include any gaps shorter than 60 minutes in duration in its calculations. NVIDIA will not include in its Capacity Availability measurement a capacity shortfall that does not cause a reported customer impact.

Service Availability or Capacity Availability: Customer Claims and Exclusions

In order for NVIDIA to consider Customer's claim regarding loss of a Cloud Offering's Service Availability or Capacity Availability ("Claim"), as applicable, Customer must submit the Claim to NVIDIA's customer support and include all necessary information, including: (a) a detailed description of the incident, (b) information regarding when the incident began and the duration of the incident, (c) details regarding the applicable Cloud Offering, Customer's attempted use of the applicable Cloud Offering, and affected users (if applicable), and (d) information regarding Customer's attempts to resolve the incident. NVIDIA must receive the Customer's Claim related to the Cloud Offering within two months from the date the incident began.

NVIDIA will exclude loss of Cloud Offerings related to Service Availability or Capacity Availability, as applicable, from its calculations, and this SLA will not apply, to the following (each, an "Exclusion"):

(a) Issues due to factors outside NVIDIA's reasonable control, including (without limitation)incidents that result from: (i) Customer's devices, software, telecommunications, firewalls, service providers, or use of the public internet, (ii) the type of device Customer has, Customer's location, the available bandwidth, or the speed of Customer's internet connection, (iii) other Customer technical issues, errors, or bandwidth constraints, (iv) cyberattacks, denial of service or similar attacks, mail bombs, DNS resolution, domain name expiration, IP transit provider or internet issues, SYN attacks, and other similar events or loss, or (v) Customer's use of third-party software, hardware, or services not provided by NVIDIA.

(b) Delay or non-performance by NVIDIA due to Force Majeure.

(c) Issues that result from Customer's actions or inactions, or the actions or inactions of Customer's Authorized Users, agents and suppliers, or any Authorized User issues, including incidents that result after NVIDIA advises Customer to modify its use of a Service if Customer did not modify its use as advised.

(d) Suspension of Service Availability for NVIDIA to perform maintenance, whether planned or unplanned.

(e) Cloud Offerings designated as "pre-release" or pre-general availability, unless otherwise agreed by NVIDIA in writing.

(f) NVIDIA's intentional suspension of Cloud Offerings as permitted by the Agreement.

(g) If any payment is overdue from Customer.

1. NVIDIA DGX CLOUD

The following terms supplement the SLA for the Agreement. NVIDIA DGX Cloud will be delivered with the following target Service Availability and Capacity Availability levels per each Accelerate Computing Environment (ACE) measured over a calendar month:

Service Availability:	99%
Capacity Availability:	95%

Service Availability Definition

The NVIDIA DGX Cloud comprises two major capabilities, each of which will meet the above Service Availability measured per calendar month. The capabilities are:

- Private registry: List, download and/or upload artifacts.

- Data management access and transfer data.

Workload Management: Logging onto the system and launch jobs. Service Availability assumes that there is sufficient available contracted capacity that has not been consumed by other Authorized Users or jobs initiated by Customer.

Capacity Availability Definition

Capacity Availability is measured in system hours, which is the number of DGX systems dedicated to Customer times the number of hours that Customer is committed in a given month. Capacity Availability is the delivered system hours divided by the committed system hours for a calendar month.

Remedy

If NVIDIA does not meet one or both of the Service Availability and Capacity Availability levels specified in a given month, Customer may submit a Claim. If NVIDIA validates Customer's Claim and provides Customer with written confirmation of acceptance, as Customer's sole remedy NVIDIA will provide a credit equal to the scope of the service interruption as defined above rounded up to the nearest day. Customer may only use a credit that NVIDIA grants towards a future NVIDIA DGX Cloud order term.

2. NVIDIA OMNIVERSE CLOUD

The following terms supplement the SLA for the Agreement. NVIDIA Omniverse Cloud will be delivered with the following target Service Availability level measured over a calendar month:

Service Availability:	99%

Service Availability Definition

Service Availability is measured as percentage of time during a month when services are up enabling at least one Authorized User to perform at least one of the following functions listed below on available capacity in the NVIDIA Omniverse Cloud:

- Logging in to NVIDIA Omniverse Cloud,

- Opening a dummy USD file that is accessible via Customer-managed storage and Nucleus instance(s), or
- Starting a streaming job in the NVIDIA USD Explorer application.

Service Availability assumes that there is sufficient available contracted capacity that has not been consumed by other Authorized Users or batch jobs initiated by Customer.

Remedy

If NVIDIA does not meet the Service Availability level specified above in a given month, Customer may submit a Claim. If NVIDIA validates Customer's Claim and provides Customer with written confirmation of acceptance, as Customer's sole remedy NVIDIA will provide a credit for a percentage of the NVIDIA Omniverse Cloud Subscription net price paid in that month as specified in the table below. Customer may only apply the credit towards a future NVIDIA Omniverse Cloud order, as examples for additional capacity via the cloud marketplace through which the original order was purchased.

Service Availability	Credit Percentage
Less than 99% but greater than or equal to 95%	10%
Less than 95%	25%

3. NVIDIA CUOPT SERVICE

The following terms supplement the SLA for the Agreement. NVIDIA cuOpt Service will be delivered with the following target Service Availability level measured over a calendar month:

Service Availability:	99%

Service Availability Definition

Service Availability is measured as a "non-500 response" from the cuOpt Service API health endpoint.

Remedy

If NVIDIA does not meet the Service Availability levels specified above in a given month, Customer may submit a Claim. If NVIDIA validates Customer's Claim and provides Customer with written confirmation of acceptance, NVIDIA will provide as Customer's sole remedy a credit equal to the scope of the service interruption as defined above rounded up to the nearest hour. Customer may only apply the credit to extend the current NVIDIA cuOpt Service order term.

4. NVIDIA CLOUD FUNCTIONS (NVCF) SERVICE

The following terms supplement the SLA for the Agreement. NVIDIA Cloud Functions (NVCF) Service will be delivered with the following target Service Availability level measured over a calendar month:

Service Availability:	99%

Service Availability Definition

Service Availability is measured as a "non-500 response" from the NVIDIA Cloud Functions API health endpoint.

Remedy

If NVIDIA does not meet the Service Availability levels specified above in a given month, Customer may submit a Claim. If NVIDIA validates Customer's Claim and provides Customer with written confirmation of acceptance, as Customer's sole remedy NVIDIA will provide a credit equal to the scope of the service interruption as defined above rounded up to the nearest hour. Customer may only apply the credit to extend the current NVIDIA Cloud Functions Service order term.