NVIDIA Virtual GPU Software Packaging, Pricing, and Licensing Guide

Application Note

DA-09924-001_v12 | August 2024

Document History

Version	Date	Authors	Description of Change
01	April 22, 2020	CH, SM	Initial Release
02	May 12, 2020	CH, SM	Updates to “Perpetual Concurrent User License” and “NVIDIA Education Pricing Program” sections
03	June 25, 2020	CH, RR, MR	Update NVIDIA Virtual Compute Server subscription price
04	July 20, 2020	NN	Update NVIDIA Virtual Compute Server abbreviated name
05	September 30, 2020	MR, RR	Add Ampere GPU support
06	February 17, 2021	MR, VK, NS, SM	Updated new Ampere GPU support, vGPU rebranding changes
07	April 14, 2021	NS, SM	Added A10 and A16
08	June 25, 2021	NN, SM	Updated CCU description in Table 1 Updated maximum concurrent VMs
09	August 3, 2021	NN, SM	Added subscription pricing to Table 9 Updated annual license subscription in Table 10 Updated education pricing in Table 12
10	Feb 1, 2022	MR	Added A2, A100X and A30X support
11	Mar 30, 2023	MR	Added L4, L40 support
12	August 1, 2024	MR, PL	Updated software editions and supported GPUs

Table of Contents

Terminology 4
Overview 5
    General Purchasing Information 6
Virtual GPU Software Product Details 8
    Virtual GPU Software Products 8
        NVIDIA vApps 8
        NVIDIA vPC 8
        NVIDIA RTX vWS 9
    vGPU Software Products and Entitlement 9
vGPU Software Licensing and Pricing 11
    Enterprise vGPU Software Pricing 11
    Subscription Concurrent User License 11
    Perpetual Concurrent User License 12
    NVIDIA Education Pricing Program 12
Choosing the Right License 14
NVIDIA License System 15

Terminology

The following table lists the terms that will be used throughout this guide.

Table 1. Terminology

Term	Meaning
SUMS	Support, Upgrade and Maintenance program
Perpetual License	A non-expiring, permanent software license that can be used on a perpetual basis without the need to renew. SUMS is required and is available in five-year increments. One-year SUMS available only for renewals.
Annual Subscription	A software license that is active for a fixed period as defined by the terms of the subscription. An annual subscription includes SUMS for the duration of the license term.
License Server	An application that manages license allocation, installed on a physical or virtual server.
Concurrent User (CCU) Licensing	A method of allocating licenses based on the number of VMs that are concurrently being used. A CCU license allows only one concurrent VM to be hosted per license. For application streaming or session-based solutions, each concurrent user requires a CCU license.
GPU	Graphics processing unit
Per GPU Licensing	A method of allocating licenses based on the number of GPUs that are being used. A per GPU license allows a certain number of concurrent VMs to be hosted per license.

Overview

NVIDIA virtual GPU (vGPU) solutions bring the power of NVIDIA GPUs to virtual desktops, applications, and workstations, accelerating graphics and compute to make virtualized workspaces accessible to creative and technical professionals working from home offices or anywhere.

This guide covers the entitlement, packaging, and licensing of the NVIDIA virtual GPU (vGPU) software family of products. It is intended to be a quick reference to understand the product portfolio at a high level, with the corresponding SKU information. It does not contain detailed product information, which can be obtained from the NVIDIA vGPU website at http://www.nvidia.com/virtualgpu. This guide is not intended to replace or contradict the End User License Agreement (EULA). Refer to the EULA for more detailed information.

Table 2. NVIDIA Virtual GPU Software Licensed Products

Product	Description
NVIDIA Virtual Applications (vApps)	For organizations deploying Citrix Virtual Apps and Desktops, RDSH or other app streaming or session-based solutions. Designed for PC level applications and server-based desktops.
NVIDIA Virtual PC (vPC)	For users who want a virtual desktop but need great user experience leveraging PC Windows applications, browsers, and high-definition video.
NVIDIA RTX™ Virtual Workstation (vWS)	For users who want to be able to use remote professional graphics applications with full performance on any device, anywhere.

Note: For compute-intensive server workloads, such as AI, deep learning, or HPC running VMware vSphere, refer to the *NVIDIA AI Enterprise Packaging, Pricing and Licensing Guide*.

NVIDIA vGPU software brings graphics and virtualization capabilities to NVIDIA data center deployments and is currently supported on the NVIDIA graphics cards listed in Table 3 through Table 6. Find certified servers that are supported by NVIDIA GPUs and vGPU software at https://www.nvidia.com/en-us/data-center/resources/vgpu-certified-servers/.

Table 3. Supported NVIDIA GPUs

	Framebuffer (GB)	Recommended vGPU Software
NVIDIA M10	4 x 8	NVIDIA vPC
NVIDIA M60	2 x 16	NVIDIA vWS, NVIDIA vPC
NVIDIA M6*	1 x 8	NVIDIA vWS, NVIDIA vPC
NVIDIA P4	1 x 8	NVIDIA vWS (Entry to Mid)
NVIDIA P6*	1 x 16	NVIDIA vWS, NVIDIA vPC
NVIDIA P40	1 x 24	NVIDIA vWS (Mid to High-end)
NVIDIA P100	1 x 12 / 1 x 16	NVIDIA vWS (Mid to High-end)
NVIDIA T4	1 x 16	NVIDIA vWS (Entry) NVIDIA vPC
NVIDIA RTX 6000	1 x 24	NVIDIA vWS (High-end)
NVIDIA RTX 8000	1 x 48	NVIDIA vWS (High end)
NVIDIA V100S/V100 SXM2	1 x 32	NVIDIA vWS
NVIDIA A40	1 x 48	NVIDIA vWS (High-end)
NVIDIA A10	1 x 24	NVIDIA vWS (Mid)
NVIDIA A16	4 x 16	NVIDIA vWS (Entry) NVIDIA vPC
NVIDIA A2	1 x 16	NVIDIA vWS, NVIDIA vPC
NVIDIA A6000	1 x 48	NVIDIA vWS (High-end)
NVIDIA A5000	1 x 24	NVIDIA vWS (Mid)
NVIDIA L40/L40S	1 x 48	NVIDIA vWS (High-end)
NVIDIA L4	1 x 24	NVIDIA vWS (Entry-Mid) NVIDIA vPC
NVIDIA RTX 6000 Ada	1 x 48	NVIDIA vWS (High-end)
NVIDIA RTX 5880 Ada	1 x 48	NVIDIA vWS (High-end)
NVIDIA RTX 5000 Ada	1 x 32	NVIDIA vWS (Mid)

* Designed for blade servers

Note: NVIDIA L2 and L20 are also supported with NVIDIA virtual GPU software.

General Purchasing Information

NVIDIA vGPU software products can be purchased through NVIDIA Preferred Partners and select server OEMs. A list of these Preferred Partners and OEMs can be obtained from: http://www.nvidia.com/buygrid.

NVIDIA vGPU software products can be purchased as either a perpetual license with a Support Updates and Maintenance Subscription (SUMS), or as an annual subscription. The perpetual license gives the user the right to use the software indefinitely, with no expiration. All NVIDIA vGPU software products with perpetual licenses must be purchased in conjunction with a five-year SUMS. A one-year SUMS is available only for renewals.

The annual subscription offering is a more affordable option to allow IT departments to better manage the flexibility of license volumes. NVIDIA vGPU software products with an annual subscription include SUMS for the duration of the software's subscription license.

Table 4. General Purchasing Information

Entitlement	NVIDIA vGPU Production SUMS
Maintenance	Access to all maintenance releases, defect resolutions, and security patches for flexibility in upgrading as per the NVIDIA Virtual GPU Software Lifecycle Policy
Upgrades	Access to all new major version releases including feature enhancements and new hardware support
Long-term branch maintenance	Available for up to 3 years from general availability as per the NVIDIA Virtual GPU Software Lifecycle Policy
Direct support	Direct access to NVIDIA support engineering for timely resolution of customer-specific issues
Support availability	Customer support available during standard business hours Cases accepted 24 × 7
Knowledgebase access	✓
Web support	✓
E-mail support	✓
Phone support	✓

Virtual GPU Software Product Details

NVIDIA vGPU software is the industry's most advanced technology for sharing true virtual GPU hardware acceleration between multiple users—without compromising the graphics experience. This virtualization technology ensures complete application compatibility, which means features and experience are the same as they would be on a physical device.

Virtual GPU Software Products

NVIDIA vGPU desktop and application virtualization solutions are designed to bring the power of virtualization to the users who need to be their most productive. vGPU technology ensures application compatibility, meaning any application that can run in a physical desktop environment can run in a virtual environment. Organizations can now expand their virtualization footprint without compromise.

The following NVIDIA vGPU software products are available:

- NVIDIA Virtual Applications (vApps)
- NVIDIA Virtual PC (vPC)
- NVIDIA RTX Virtual Workstation (vWS)

NVIDIA vApps

This product is for organizations deploying Citrix Virtual Apps and Desktops, RDSH or other app streaming or session-based solutions. Designed to deliver PC Windows applications at full performance, NVIDIA vApps allows users to access any Windows application at full performance on any device, anywhere.

Windows Server hosted RDSH desktops are also supported by NVIDIA vApps.

NVIDIA vPC

This product is ideal for users who want a virtual desktop but need great user experience leveraging PC Windows applications, browsers, and high-definition video. NVIDIA vPC delivers a native experience to users in a virtual environment, allowing them to run all their PC applications at full performance.

NVIDIA RTX vWS

This product is ideal for compute intensive workloads, such as accelerated graphics, NVIDIA Omniverse, and AI development. It is suitable for high-end designers who use powerful 3D content creation applications such as Dassault CATIA, SOLIDWORKS, 3DExcite, Siemens NX, PTC Creo, Schlumberger Petrel, or Autodesk Maya. NVIDIA vWS allows users to access their applications with full features and performance, anywhere, on any device.

vGPU Software Products and Entitlement

Each NVIDIA vGPU software product includes the feature entitlements listed in Table 5.

Table 5. Feature Entitlements

Feature	NVIDIA vApps	NVIDIA vPC	NVIDIA vWS
License Entitlement			
Concurrent User (CCU)	✓	✓	✓
Per GPU			
Capability Entitlement			
Desktop Virtualization		✓	✓
RDSH App Hosting	✓	✓	✓[2]
RDSH Desktop Hosting	✓	✓	✓[2]
Compute Virtualization			✓
Windows Guest OS Support	✓	✓	✓
Linux Guest OS Support	✓	✓	✓
Maximum Displays	1[3]	4	4
Maximum Resolution[4]	1280 × 1024	5120 × 2880 (5K)	7680 × 4320 (8K)
NVIDIA RTX Enterprise Software Features			✓
OpenGL, DirectX and Vulkan	✓	✓	✓
CUDA and OpenCL Support			✓[5]
ECC and Page Retirement			✓

Multi-vGPU			✓
NVLink			✓
GPU Pass Through Support[7]	✓		✓
Bare Metal Support[8]	✓		✓
vGPU Profile Sizes Supported[9]			
512 MB		✓	✓
1 GB	✓	✓	✓
2 GB	✓	✓	✓
3 GB	✓		✓
4 GB	✓		✓
5 GB			
6 GB	✓		✓
8 GB	✓		✓
10 GB			
12 GB	✓		✓
16 GB	✓		✓
20 GB			
24 GB	✓		✓
32 GB	✓		✓
40 GB			
48 GB	✓		✓

Notes:

[1]Maximum 10 concurrent VMs per GPU

[2]With packaged NVIDIA vApps license

[3]Applies only to the console display in remote application environments. For details, see Supported GPUs

[4]Review the Virtual GPU Software User Guide for the supported display configurations for each profile

[5]Supported on 8GB 1:1 profile on NVIDIA Maxwell and all profiles on Pascal

[6]ECC support begins with Pascal

[7]Only supported on 1:1 profiles

[8]Only NVIDIA M6 hardware supported as primary display device

[9]Review the Virtual GPU Software User Guide for the vGPU profiles supported on your GPU

vGPU Software Licensing and Pricing

Enterprise vGPU Software Pricing

NVIDIA vApps, NVIDIA vPC, and NVIDIA RTX vWS are available on a per Concurrent User (CCU) model. A CCU license is required for every user who is accessing or using the software at any given time, whether an active connection to the virtualized desktop or session is maintained.

NVIDIA vGPU software products can be purchased by enterprises as either perpetual licenses with annual Support Updates and Maintenance Subscription (SUMS), or as an annual subscription. All NVIDIA vGPU software products with perpetual licenses must be purchased in conjunction with a five-year SUMS. A one-year SUMS is available only for renewals. For annual licenses, SUMS is bundled into the annual license cost.

Enterprise vGPU software licensing and pricing are described in the following subsections. Pricing is suggested pricing only. Contact your authorized NVIDIA partner for final pricing. If you are looking to run or host a service by using NVIDIA vGPU software, you need to join the NPN CSP partner program.

Subscription Concurrent User License

An annual Enterprise subscription is active for a fixed period as defined by the terms of the subscription license. To be kept active, the license will need to be renewed at the end of the subscription period. The subscription license includes the software license and production level SUMS for the duration of the license subscription period.

Table 6. Concurrent User License Annual Subscription Pricing

License	Annual Subscription Pricing
NVIDIA Virtual Applications	$10 per CCU subscription
NVIDIA Virtual PC	$50 per CCU subscription

	4-year subscription - $43.75 per CCU 5-year subscription - $40 per CCU
NVIDIA RTX Virtual Workstation	$250 per CCU subscription
	4-year subscription - $206.25 per CCU 5-year subscription - $180 per CCU

Perpetual Concurrent User License

A perpetual enterprise license allows for use of the licensed software indefinitely. Users that opt to license using this model are required to subscribe to SUMS for five years. The SUMS subscription can be renewed on a yearly basis after the expiration of the initial subscription.

Table 7. Enterprise Perpetual Licensing plus SUMS Pricing

License	Annual Licensing plus SUMS Pricing
NVIDIA Virtual Applications	$20 per CCU perpetual license
	$5 SUMS per year
NVIDIA Virtual PC	$100 per CCU perpetual license
	$25 SUMS per year
NVIDIA RTX Virtual Workstation	$450 per CCU perpetual license
	$100 SUMS per year

NVIDIA Education Pricing Program

The NVIDIA Education Pricing Program supports the use of visual computing in teaching and research institutions. The program makes it easy to procure and administer NVIDIA solutions, software licensing, and services for qualified educational institutions and helps

reduce their total cost. For more information on eligibility, review the NVIDIA Education Pricing Program documentation.

The Annual Subscription Model is available for NVIDIA vWS . The Perpetual Licensing Model is available only for NVIDIA vWS. Users that opt for the Perpetual Licensing Model are required to subscribe to SUMS for five years. The SUMS subscription can be renewed on a yearly basis after the expiration of the initial subscription.

Table 8. NVIDIA Education Pricing

License	Pricing Model	Price
NVIDIA vWS	Perpetual Licensing plus SUMS	$99 per CCU perpetual license
		$25 SUMS per year
	Annual Subscription	$50 per CCU subscription

Choosing the Right License

The following table provides a guideline to help decide which license is required based on the desired entitlement and capability. This is not an all-inclusive list of possible solutions. If you have further questions or are unable to decide based on the decision tree, contact NVIDIA vGPU Support at https://www.nvidia.com/en-us/support/enterprise/.

Table 9. Common Use Cases and Solutions

I am using...	I need this license...
Citrix Virtual Desktops	NVIDIA vPC – for PC level applications NVIDIA vWS – for workstation/professional 3D use cases
Omnissa Horizon 8	NVIDIA vPC – for PC level applications NVIDIA vWS – for each session utilizing a workstation/professional 3D use case
Citrix Virtual Apps and Desktops	NVIDIA vApps NVIDIA vWS – for each session utilizing a workstation/professional 3D use case
Omnissa Horizon 8 RDSH	NVIDIA vApps NVIDIA vWS – for each session utilizing a workstation/professional 3D use case
Other RDSH / Session-based	NVIDIA vApps NVIDIA vWS – for each session utilizing a workstation/professional 3D use case
Microsoft RemoteFX	NVIDIA vPC – for PC level applications
VMware vSGA	NVIDIA vPC – for PC level applications
Microsoft Hyper-V (DDA)	NVIDIA vWS
Microsoft AzureStack GPU-P	NVIDIA vPC – for PC level applications NVIDIA vWS – for each session utilizing a workstation/professional 3D use case
Red Hat Enterprise Linux with KVM	NVIDIA vWS – for workstation/professional 3D use cases

NVIDIA License System

The NVIDIA License System provides monitoring and reporting on license usage for capacity planning. It is available for download with the vGPU software packages from the NVIDIA Licensing Portal. This license server can be installed on either a physical server or, more likely, a dedicated virtual machine.

To ensure that the end user experience is not impaired by license overages or connection issues, the vGPU software runs with or without a valid license server connection. The license server allows the IT administrator to track license usage and determine correct sizing for their environments. In compliance with the EULA, IT administrators may also use any other method that reliably tracks the software usage to ensure they have enough licenses for their CCU or per GPU usage.

Notice

This document is provided for information purposes only and shall not be regarded as a warranty of a certain functionality, condition, or quality of a product. NVIDIA Corporation (“NVIDIA”) makes no representations or warranties, expressed or implied, as to the accuracy or completeness of the information contained in this document and assumes no responsibility for any errors contained herein. NVIDIA shall have no liability for the consequences or use of such information or for any infringement of patents or other rights of third parties that may result from its use. This document is not a commitment to develop, release, or deliver any Material (defined below), code, or functionality.

NVIDIA reserves the right to make corrections, modifications, enhancements, improvements, and any other changes to this document, at any time without notice.

Customer should obtain the latest relevant information before placing orders and should verify that such information is current and complete.

NVIDIA products are sold subject to the NVIDIA standard terms and conditions of sale supplied at the time of order acknowledgement, unless otherwise agreed in an individual sales agreement signed by authorized representatives of NVIDIA and customer (“Terms of Sale”). NVIDIA hereby expressly objects to applying any customer general terms and conditions with regards to the purchase of the NVIDIA product referenced in this document. No contractual obligations are formed either directly or indirectly by this document.

NVIDIA products are not designed, authorized, or warranted to be suitable for use in medical, military, aircraft, space, or life support equipment, nor in applications where failure or malfunction of the NVIDIA product can reasonably be expected to result in personal injury, death, or property or environmental damage. NVIDIA accepts no liability for inclusion and/or use of NVIDIA products in such equipment or applications and therefore such inclusion and/or use is at customer’s own risk.

NVIDIA makes no representation or warranty that products based on this document will be suitable for any specified use. Testing of all parameters of each product is not necessarily performed by NVIDIA. It is customer’s sole responsibility to evaluate and determine the applicability of any information contained in this document, ensure the product is suitable and fit for the application planned by customer, and perform the necessary testing for the application in order to avoid a default of the application or the product. Weaknesses in customer’s product designs may affect the quality and reliability of the NVIDIA product and may result in additional or different conditions and/or requirements beyond those contained in this document. NVIDIA accepts no liability related to any default, damage, costs, or problem which may be based on or attributable to: (i) the use of the NVIDIA product in any manner that is contrary to this document or (ii) customer product designs.

No license, either expressed or implied, is granted under any NVIDIA patent right, copyright, or other NVIDIA intellectual property right under this document. Information published by NVIDIA regarding third-party products or services does not constitute a license from NVIDIA to use such products or services or a warranty or endorsement thereof. Use of such information may require a license from a third party under the patents or other intellectual property rights of the third party, or a license from NVIDIA under the patents or other intellectual property rights of NVIDIA.

Reproduction of information in this document is permissible only if approved in advance by NVIDIA in writing, reproduced without alteration and in full compliance with all applicable export laws and regulations, and accompanied by all associated conditions, limitations, and notices.

THIS DOCUMENT AND ALL NVIDIA DESIGN SPECIFICATIONS, REFERENCE BOARDS, FILES, DRAWINGS, DIAGNOSTICS, LISTS, AND OTHER DOCUMENTS (TOGETHER AND SEPARATELY, “MATERIALS”) ARE BEING PROVIDED “AS IS.” NVIDIA MAKES NO WARRANTIES, EXPRESSED, IMPLIED, STATUTORY, OR OTHERWISE WITH RESPECT TO THE MATERIALS, AND EXPRESSLY DISCLAIMS ALL IMPLIED WARRANTIES OF NONINFRINGEMENT, MERCHANTABILITY, AND FITNESS FOR A PARTICULAR PURPOSE. TO THE EXTENT NOT PROHIBITED BY LAW, IN NO EVENT WILL NVIDIA BE LIABLE FOR ANY DAMAGES, INCLUDING WITHOUT LIMITATION ANY DIRECT, INDIRECT, SPECIAL, INCIDENTAL, PUNITIVE, OR CONSEQUENTIAL DAMAGES, HOWEVER CAUSED AND REGARDLESS OF THE THEORY OF LIABILITY, ARISING OUT OF ANY USE OF THIS DOCUMENT, EVEN IF NVIDIA HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. Notwithstanding any damages that customer might incur for any reason whatsoever, NVIDIA’s aggregate and cumulative liability towards customer for the products described herein shall be limited in accordance with the Terms of Sale for the product.

Trademarks

NVIDIA, the NVIDIA logo, CUDA, NVIDIA HGX, NVIDIA Maxwell, NVIDIA Turing, NVIDIA RTX, NVIDIA Volta, NVIDIA Ampere, NVIDIA Ada NVLink, and TensorRT are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. Other company and product names may be trademarks of the respective companies with which they are associated.

OpenCL

OpenCL is a trademark of Apple Inc. used under license to the Khronos Group Inc.

Copyright

© 2020, 2021, 2022, 2023, 2024 NVIDIA Corporation. All rights reserved.