



# NVIDIA VIRTUAL GPU PACKAGING, PRICING AND LICENSING

June 2019



# TABLE OF CONTENTS

## OVERVIEW

### [1.1 GENERAL PURCHASING INFORMATION](#)

## VGPU PRODUCT DETAILS

### [1.2 VIRTUAL GPU SOFTWARE EDITIONS](#)

### [1.3 VIRTUAL GPU SOFTWARE EDITIONS AND ENTITLEMENT](#)

## VGPU SOFTWARE LICENSING MODEL

### [1.4 CHOOSING THE RIGHT LICENSE](#)

### [1.5 NVIDIA LICENSE MANAGER](#)

## 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 a need to renew. SUMS is required and is available in three, four, or 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 license, typically yearly. This includes SUMS for the duration of the license term.
License Manager	An application that manages license allocation, installed on a physical or virtual server.
Concurrent User (CCU)	A method of counting licenses based on the number of VMs that are concurrently being used.
Per GPU Licensing	A method of providing licenses based on the number of GPUs that will be utilized. A per GPU license allows a certain number of concurrent VMs to be hosted per license.

# OVERVIEW

The NVIDIA virtual GPU software delivers accelerated virtual desktops and applications from the data center to any user, on any device, anywhere.

This guide covers the entitlement, packaging and licensing of the NVIDIA virtual GPU (vGPU) 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 accessed from the NVIDIA vGPU website at <http://www.nvidia.com/virtualgpu>. This document is not intended to replace or contradict the End User License Agreement (EULA). Please refer to the EULA ([here](#)) for more detailed information.

NVIDIA virtual GPU Software Editions	
<b>NVIDIA GRID® Virtual Applications</b>	For organizations deploying XenApp, RDSH or other app streaming or session-based solutions. Designed for PC level applications and server based desktops.
<b>NVIDIA GRID® Virtual PC</b>	For users who want a virtual desktop but need great user experience leveraging PC Windows applications, browsers, and high definition video.
<b>NVIDIA Quadro® vDWS</b>	For users who want to be able to use remote professional graphics applications with full performance on any device, anywhere.
<b>NVIDIA Virtual Compute Server</b>	For compute intensive, server workload such as artificial intelligence (AI), deep learning, or high-performance computing (HPC).

NVIDIA vGPU brings graphics and virtualization capabilities to NVIDIA data-center deployments, and is currently supported on the following NVIDIA GPUs. Find certified servers with NVIDIA GPUs that are supported by vGPU at <http://www.nvidia.com/grid-certified-servers>.

	Tesla M10	Tesla M60	Tesla M6
<b>Recommended Use case</b>	GRID vPC - User Density-Optimized	Performance-Optimized	Blade-Optimized
<b>Number of GPUs</b>	4 NVIDIA Maxwell GPUs	2 NVIDIA Maxwell GPUs	1 NVIDIA Maxwell GPU
<b>Total NVIDIA CUDA® Cores</b>	2,560 (640 per GPU)	4,096 (2,048 per GPU)	1,536
<b>Total Memory Size</b>	32 GB GDDR5 (8 GB per GPU)	16 GB GDDR5 (8 GB per GPU)	8 GB GDDR5
<b>Max Power</b>	225 W	300 W	100 W
<b>Form Factor</b>	PCIe 3.0 Dual Slot	PCIe 3.0 Dual Slot	MXM
<b>Board Dimensions</b>	10.5" x 4.4"	10.5" x 4.4"	3.2" x 4.1"
<b>Cooling Solution</b>	Passive	Passive / Active	Bare Board

	Tesla P4	Tesla P6	Tesla P40	Tesla P100
<b>Recommended Use case</b>	Quadro vDWS (Entry to mid)	Quadro vDWS or GRID vPC - Blade-Optimized	Quadro vDWS - Performance Optimized (Mid to high)	Quadro vDWS (Mid to high)
<b>Number of GPUs</b>	1 NVIDIA Pascal GP104	1 NVIDIA Pascal GP104	1 NVIDIA Pascal GP102	1 NVIDIA Pascal GP100
<b>Total NVIDIA CUDA® Cores</b>	2,560	2,048	3,840	3,584
<b>Total Memory Size</b>	8 GB GDDR5	16 GB GDDR5	24 GB GDDR5	12 GB HBM2/16 GB HBM2
<b>Max Power</b>	75 W	90 W	250 W	250 W/300 W
<b>Form Factor</b>	PCIe 3.0 Single Slot	MXM	PCIe 3.0 Dual Slot	SXM2/PCIe 3.0
<b>Board Dimensions</b>	2.7" x 6.6"	3.2" x 4.1"	10.5" x 4.4"	10.5" x 4.4"
<b>Cooling Solution</b>	Passive	Bare Board	Passive	Passive

	NVIDIA T4	Quadro RTX 6000	Quadro RTX 8000	Tesla V100
<b>Recommended Use case</b>	Quadro vDWS - Performance Optimized (Entry to mid); GRID vPC - Density Optimized; NVIDIA vCS - Compute Optimized	Quadro vDWS - Performance Optimized (High end)	Quadro vDWS - Performance Optimized (High end)	Quadro vDWS - Performance-Optimized; NVIDIA vCS - Compute Optimized
<b>Number of GPUs</b>	1 NVIDIA Turing TU104	1 NVIDIA Turing TU102	1 NVIDIA Turing TU102	1 NVIDIA Volta GPU
<b>Total NVIDIA CUDA® Cores</b>	2,560	4,608	4,608	5120
<b>Tensor Cores</b>	320	576	576	--
<b>RT Cores</b>	40	72	72	--
<b>Total Memory Size</b>	16 GB GDDR6	24 GB GDDR6	48 GB GDDR6	16 GB HBM2
<b>Max Power</b>	70 W	295 W	295 W	300 W
<b>Form Factor</b>	PCIe 3.0 Single Slot	PCIe 3.0 Dual Slot	PCIe 3.0 Dual Slot	PCIe, SXM2, Full Height/Length
<b>Board Dimensions</b>	2.7" x 6.6"	10.5" x 4.4"	10.5" x 4.4"	10.5" x 4.4" 5.5" x 3.1" x .5" 6.9" x 4.4"
<b>Cooling Solution</b>	Passive	Active	Active	Passive

## 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 perpetual licenses with yearly Support Updates and Maintenance agreement (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 three, four, or five year of SUMS. 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 annual subscription are bundled with SUMS for the duration of the software’s subscription license.

Entitlement	NVIDIA vGPU Production SUMS
Maintenance	Access to all maintenance releases, defect resolutions, and security patches for flexibility in upgrading for up to 3 years
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
Direct support	Direct access to NVIDIA support engineering for timely resolution of customer-specific issues
Support availability	24 x 7
Knowledgebase access	✓
Web support	✓
E-mail support	✓
Phone support	✓

## VIRTUAL GPU PRODUCT DETAILS

NVIDIA vGPU 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.

# 1.1 VIRTUAL GPU SOFTWARE EDITIONS

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 can run in a virtual environment. Organizations can now expand their virtualization footprint without compromise.

NVIDIA vGPU software is available in four editions: NVIDIA GRID Virtual PC (GRID vPC), NVIDIA Quadro® Virtual Data Center Workstation (Quadro vDWS), GRID Virtual Applications (GRID vApps), and NVIDIA Virtual ComputeServer (NVIDIA vCS).

## GRID 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 GRID Virtual PC delivers a native experience to users in a virtual environment, allowing them to run all of their PC applications at full performance.

## Quadro vDWS

This edition is ideal for mainstream and high-end designers who use powerful 3D content creation applications like Dassault CATIA, SOLIDWORKS, and 3DExcite, Siemens NX, PTC Creo, Schlumberger Petrel, or Autodesk Maya. NVIDIA Quadro® Virtual Data Center Workstation allows users to access their professional graphics applications with full features and performance, anywhere, on any device.

## GRID vApps

For organizations deploying XenApp, RDSH or other app streaming or session-based solutions. Designed to deliver PC Windows applications at full performance. NVIDIA GRID Virtual Applications allows users to access any Windows application at full performance on any device, anywhere.

This edition is suited for users who would like to virtualize applications using XenApp, RDSH or other app streaming or session-based solutions. Windows Server hosted RDSH desktops are also supported by GRID vApps.

## NVIDIA vCS

For organizations running compute-intensive server workloads such as Artificial Intelligence (AI), Deep Learning (DL), and High Performance Computing (HPC). NVIDIA vComputeServer is software that enables the NVIDIA GPU to be virtualized to accelerate compute-intensive server workloads with features such as ECC, page retirement, peer-to-peer over NVLink, Multi-vGPU.

## 1.2 vGPU SOFTWARE EDITIONS AND ENTITLEMENT

Each NVIDIA vGPU software product includes the following feature entitlement:

Feature	GRID vApps	GRID vPC	Quadro vDWS	vCompute Server
<b>License Entitlement</b>				
Concurrent User (CCU)	✓	✓	✓	
Per GPU				✓
<b>Capability Entitlement</b>				
Desktop Virtualization		✓	✓	
RDSH App Hosting	✓	✓	✓ <sub>1</sub>	
RDSH Desktop Hosting	✓	✓	✓ <sub>1</sub>	
Compute Virtualization			✓	✓
Windows Guest OS	✓	✓	✓	
Linux Guest OS		✓	✓	✓
Maximum Displays	1 <sup>2</sup>	4	4	1
Maximum Resolution	1280*1024	4096*2160 (4K) <sup>3</sup>	4096*2160 (4K)	4096*2160 (4K)
NVIDIA Quadro Software Features			✓	
CUDA & OpenCL Supported			✓ <sub>4</sub>	✓
ECC & Page Retirement			✓	✓
Multi-GPU			✓	✓
NVLINK			✓	✓
GPU Pass-through Supported <sup>5</sup>	✓		✓	✓
Bare Metal Supported <sup>6</sup>	✓		✓	
<b>vGPU Profiles Supported<sup>7</sup></b>				
512 MB		✓	✓	
1 GB	✓	✓	✓	
2 GB	✓	✓	✓	
3 GB	✓		✓	
4 GB	✓		✓	✓
6 GB	✓		✓	✓
8 GB	✓		✓	✓
12 GB	✓		✓	✓
16 GB	✓		✓	✓
24 GB	✓		✓	✓
32 GB	✓		✓	✓
48 GB			✓	

<sup>1</sup> With packaged GRID vApps license

<sup>2</sup> Applies only to the console display in remote application environments. For details, see [Supported GPUs](#)



<sup>3</sup> Supports up to two 4K displays or four 2560x1600 displays on 2B profile. Supports single 4K display or up to four 2560x1600 displays on 1B profile.

<sup>4</sup> Supported on 8GB 1:1 profile on Maxwell and all profiles on Pascal

<sup>5</sup> Only supported on 1:1 profiles

<sup>6</sup> Only NVIDIA Tesla M6 Hardware supported as primary display device

<sup>7</sup> Please review the [Virtual GPU Software User Guide](#) for the vGPU profiles supported on your GPU

## VGPU SOFTWARE LICENSING AND PRICING

GRID vPC, Quadro vDWS and GRID vApps 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 or not an active connection to the virtualized desktop or session is maintained.

NVIDIA vCS is available on a per GPU model. A GPU license is required for every GPU that will host vCS-enabled VMs. A single vCS license enables a maximum of 8 concurrent VMs.

NVIDIA vGPU editions 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 three, four, or five year of SUMS. One year SUMS is available only for renewals. For annual licenses, SUMS is bundled into the annual license cost.

Enterprise vGPU Software Pricing is listed in the tables below, [find the full SKU list here](#). Pricing is suggested pricing only, contact your authorized NVIDIA partner for final pricing. If you are looking to run or host a service using NVIDIA vGPU Software, you need to join the [NPN CSP partner program](#).

### 1.1 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.

Enterprise Annual Subscription Pricing	
GRID Virtual Applications	\$10 per CCU subscription
GRID Virtual PC	\$50 per CCU subscription
	4 year subscription - \$43.75 per CCU 5 year subscription - \$40 per CCU
Quadro Virtual Data Center Workstation	\$250 per CCU subscription

## 1.2 SUBSCRIPTION PER GPU 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.

Annual Subscription Pricing	
NVIDIA Virtual Compute Server	\$50 per GPU subscription

## 1.3 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 three years. The SUMS subscription can be renewed on a yearly basis after the expiration of the initial subscription.

Enterprise Perpetual Licensing + SUMS Pricing	
GRID Virtual Applications	\$20 per CCU perpetual license
	\$5 SUMS per year
GRID Virtual PC	\$100 per CCU perpetual license
	\$25 SUMS per year
Quadro Virtual Data Center Workstation	\$450 per CCU perpetual license
	\$100 SUMS per year

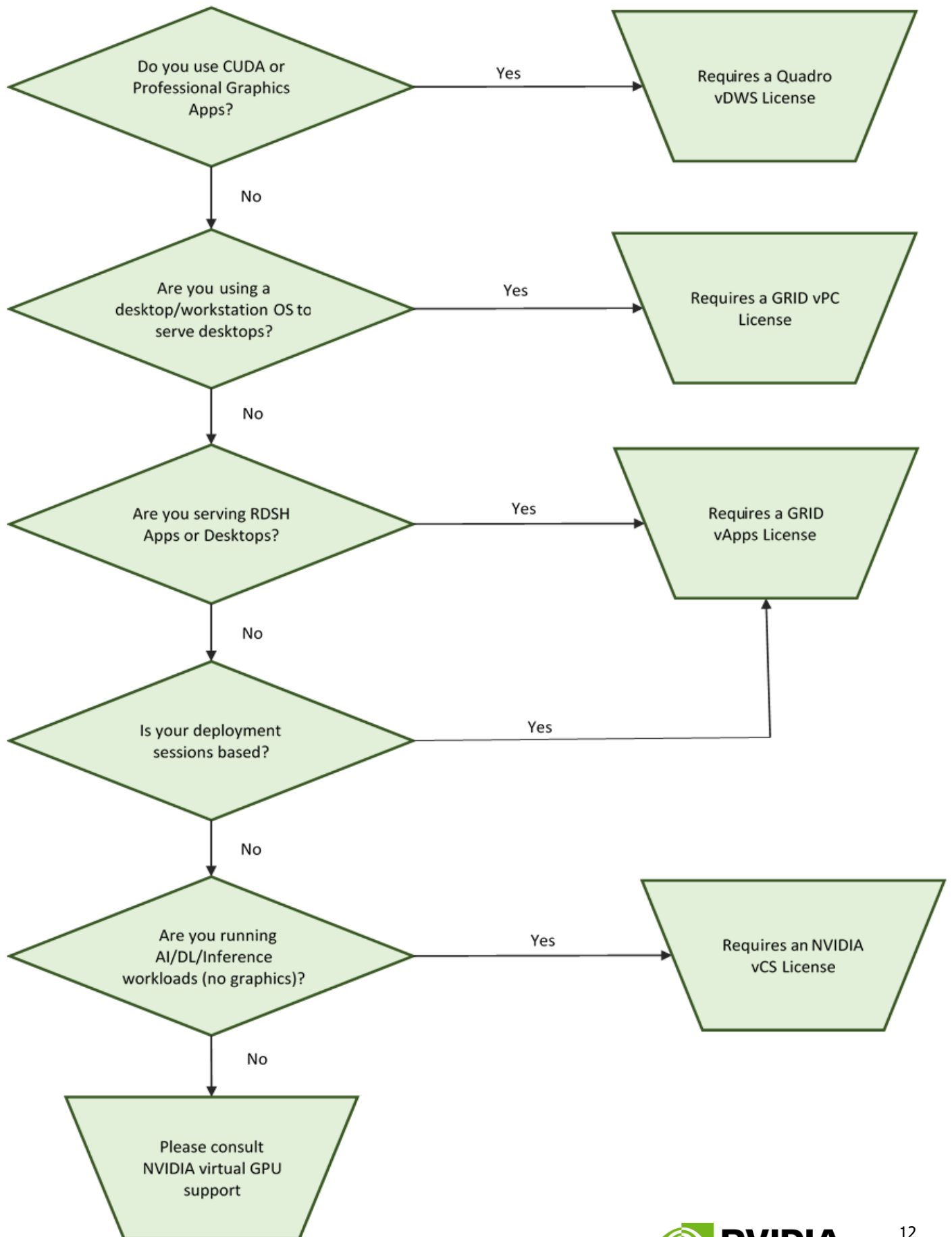
## 1.4 NVIDIA EDUCATION PRICING PROGRAM

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 and helps reduce the total cost of NVIDIA Solutions, software licensing and services for qualified educational institutions. The program includes NVIDIA solutions, software and services. For more information on eligibility, please review the [NVIDIA Education Pricing Program](#) documentation.

Annual Subscription Pricing	
Quadro Virtual Data Center Workstation	\$50 per CCU subscription
Perpetual Licensing + SUMS Pricing	
Quadro Virtual Data Center Workstation	\$99 per CCU perpetual license
	\$25 SUMS per year

## 1.3 DECIDING THE RIGHT LICENSE BASED ON CAPABILITY AND ENTITLEMENT

The following flowchart provides a simple decision tree to help decide which license is required based on the desired entitlement and capability. If you have further questions or are unable to decide based on the decision tree, please contact NVIDIA vGPU Support at <http://www.nvidia.com/gridsupport>.



The table below summarizes some common use cases of different solutions. This is not an all-inclusive list of possible solutions. If you have questions, please contact NVIDIA vGPU Support.

I am using...	I need this license...
Citrix XenDesktop	GRID vPC - for PC level applications Quadro vDWS - for workstation/professional 3D use cases
VMware Horizon (View)	GRID vPC - for PC level applications Quadro vDWS - for each session utilizing a workstation/professional 3D use case
Citrix XenApp	GRID vApps Quadro vDWS - for each session utilizing a workstation/professional 3D use case
VMware Horizon RDSH	GRID vApps Quadro vDWS - for each session utilizing a workstation/professional 3D use case
Other RDSH / Session-based	GRID vApps Quadro vDWS - for each session utilizing a workstation/professional 3D use case
Microsoft RemoteFX	GRID vPC - for PC level applications
VMware Horizon vSGA	GRID vPC - for PC level applications
Microsoft Hyper-V (DDA)	Quadro vDWS

## 1.4 NVIDIA LICENSE MANAGER

The NVIDIA License Manager provides monitoring and reporting on license usage for capacity planning and is included with the vGPU software packages. This License Manager can be installed on either a physical server or, more likely, a dedicated virtual machine.

To improve end user experience, the vGPU software will run with or without a valid license server connection. This is done to ensure that user experience is not impacted by license overages or connection issues. The license server will allow 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

The information provided in this specification is believed to be accurate and reliable as of the date provided. However, NVIDIA Corporation (“NVIDIA”) does not give any representations or warranties, expressed or implied, as to the accuracy or completeness of such information. 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 publication supersedes and replaces all other specifications for the product that may have been previously supplied.

NVIDIA reserves the right to make corrections, modifications, enhancements, improvements, and other changes to this specification, at any time and/or to discontinue any product or service without notice. Customer should obtain the latest relevant specification 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. NVIDIA hereby expressly objects to applying any customer general terms and conditions with regard to the purchase of the NVIDIA product referenced in this specification.

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 these specifications will be suitable for any specified use without further testing or modification. Testing of all parameters of each product is not necessarily performed by NVIDIA. It is customer’s sole responsibility to ensure the product is suitable and fit for the application planned by customer and to do 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 specification. NVIDIA does not accept any 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 specification, 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 specification. 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 specification is permissible only if reproduction is approved by NVIDIA in writing, is reproduced without alteration, and is accompanied by all associated conditions, limitations, and notices.

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. 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 NVIDIA terms and conditions of sale for the product.

## OpenCL

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

## Trademarks

NVIDIA and the NVIDIA logo 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.

## Copyright

© 2015-2016 NVIDIA Corporation. All rights reserved