

Virtual GPU Software

Quick Start Guide

Table of Contents

About this Guide	iii
Chapter 1. Getting Your NVIDIA vGPU Software	1
1.1. Before You Begin	1
1.2. Your Order Confirmation Message	1
1.3. NVIDIA Enterprise Account Requirements	3
1.4. Creating your NVIDIA Enterprise Account	4
1.5. Linking an Evaluation Account to an NVIDIA Enterprise Account for Purchased Licenses	
1.6. Downloading your NVIDIA vGPU Software	8
Chapter 2. Installing Your NVIDIA vGPU Software License Server and License Files 2.1. Overview of NVIDIA vGPU software Licensing	
2.2. Platform Configuration Requirements	
2.3. Installing the Java Runtime Environment	
2.4. Installing the License Server Software	
2.5. Obtaining the License Server's MAC Address	
2.6. Managing your License Servers and Getting your License Files	
2.6.1. Creating a License Server on the NVIDIA Licensing Portal	
2.6.2. Downloading a License File	
2.7. Installing a License	
Chapter 3. Installing and Configuring NVIDIA vGPU Manager and the Guest Driver	24
3.1. Switching the Mode of a GPU that Supports Multiple Display Modes	
3.2. Switching the Mode of a Tesla M60 or M6 GPU	
3.3. Installing the NVIDIA Virtual GPU Manager	
3.3.1. Installing the NVIDIA Virtual GPU Manager on VMware vSphere	
3.3.2. Installing the NVIDIA Virtual GPU Manager on Citrix Hypervisor	26
3.4. Disabling and Enabling ECC Memory	27
3.4.1. Disabling ECC Memory	27
3.4.2. Enabling ECC Memory	29
3.5. Attaching an NVIDIA vGPU Profile to a VM	30
3.5.1. Changing the Default Graphics Type in VMware vSphere 6.5 and Later	30
3.5.2. Configuring a vSphere VM with NVIDIA vGPU	34
3.5.3. Configuring a Citrix Hypervisor VM with Virtual GPU	36
3.6. Installing the NVIDIA vGPU Software Graphics Driver	
3.7. Licensing an NVIDIA vGPU	38

About this Guide

Virtual GPU Software Quick Start Guide provides minimal instructions for installing and configuring NVIDIA[®] virtual GPU software on the Citrix Hypervisor or VMware vSphere hypervisor and for installing and configuring a single-instance NVIDIA virtual GPU License Server on Windows. If you need complete instructions, are using other platforms, or require high availability for the license server, refer to <u>Virtual GPU Software User Guide</u> and <u>Virtual GPU License Server User Guide</u>.

Chapter 1. Getting Your NVIDIA vGPU Software

After your order for NVIDIA vGPU software is processed, you will receive an order confirmation message from NVIDIA. This message contains information that you need for getting your NVIDIA vGPU software from the NVIDIA Licensing Portal. To log in to the NVIDIA Licensing Portal, you must have an NVIDIA Enterprise Account.

1.1. Before You Begin

Before following the procedures in this guide, ensure that the following prerequisites are met:

- You have a server platform that is capable of hosting your chosen hypervisor and NVIDIA GPUs that support NVIDIA vGPU software. For a list of validated server platforms, refer to <u>NVIDIA GRID Certified Servers</u>.
- One or more NVIDIA GPUs that support NVIDIA vGPU software is installed in your server platform.
- A supported virtualization software stack is installed according to the instructions in the software vendor's documentation.
- A virtual machine (VM) running a supported Windows guest operating system (OS) is configured in your chosen hypervisor.

For information about supported hardware and software, and any known issues for this release of NVIDIA vGPU software, refer to the *Release Notes* for your chosen hypervisor:

- Virtual GPU Software for Citrix Hypervisor Release Notes
- Virtual GPU Software for VMware vSphere Release Notes

1.2. Your Order Confirmation Message

After your order for NVIDIA vGPU software is processed, you will receive an order confirmation message to which your NVIDIA Entitlement Certificate is attached.

entitlement pdf 13 KB

Thank you for your software and/or services order!

Please find enclosed your Entitlement Certificate for the Software and/or Services products you ordered.

Please refer to the attached Entitlement Certificate to register for your software and services.

The following is your order information:

PO Number	NVIDIA Sales Order	NVIDIA Delivery Number	

Questions?

NVIDIA Enterprise Support contact information can be found here https://www.NVIDIA.com/en-us/support/enterprise/

Your NVIDIA Entitlement Certificate contains your product activation keys.



NVIDIA Corporation 2788 San Tomas Expressway SANTA CLARA CA 95051 USA

NVIDIA® Entitlement Certificate

This certificate serves as evidence that NVIDIA has entitled you for the following product(s).

End Customer (NVIDIA Delivery	
	Entitlement Date	11 APR 2019
	PO Number	
	NVIDIA Sales Order	

No	Entitlement Description	Quantity	Sales Type	Term	Start Date	End Date
1	Quadro vDWS Bundle - Subscription, CCU					
	Quadro vDWS Subscription, CCU	48 EA	Initial	1 Year	11 APR 2019	10 APR 2020
	GRID vApps Subscription, CCU	48 EA	Initial	1 Year	11 APR 2019	10 APR 2020
	PAK ID					

Please follow the instructions provided in the following section to register your entitlements.

Thank you for your order!

Your NVIDIA Entitlement Certificate also provides instructions for using the certificate.

NOTICE

HOW TO USE THIS CERTIFICATE

Registration Instructions

Please refer to the NVIDIA Virtual GPU Software Quick Start Guide for information on how to get started, including instructions on how to register for your entitlement.

Sales Type: Initial

- 1. Already have vGPU entitlements? Please Login
- 2. New to vGPU entitlements? Please register and follow instructions on the registration page.

Questions?

NVIDIA Enterprise Support contact information can be found here

Rights and restrictions on the use, transfer and copying of the Software are set forth in NVIDIA's End User License Agreement.

1.3. NVIDIA Enterprise Account Requirements

To get your NVIDIA vGPU software, you must have a suitable NVIDIA Enterprise Account for accessing your licenses.

- **Note:** For a Support, Upgrade, and Maintenance Subscription (SUMS) renewal, you should already have a suitable NVIDIA Enterprise Account and this requirement should already be met. However, if you have an account that was created for an evaluation license and you want to access licenses that you purchased, you must repeat the registration process.
- If you do not have an account, follow the **Register** link in the instructions for using the certificate to create your account. For details, refer to <u>Creating your NVIDIA Enterprise</u> <u>Account</u>.
- If you have an account that was created for an evaluation license and you want to access licenses that you purchased, follow the **Register** link in the instructions for using the certificate to create an account for your **purchased** licenses. You can choose to create a separate account for your purchased licenses or link your existing account for an evaluation license to the account for your purchased licenses.
 - To create a separate account for your purchased licenses, follow the instructions in <u>Creating your NVIDIA Enterprise Account</u>, specifying a different e-mail address than the address with which you created your existing account.
 - To link your existing account for an evaluation license to the account for your purchased licenses, follow the instructions in <u>Linking an Evaluation Account to an</u>

<u>NVIDIA Enterprise Account for Purchased Licenses</u>, specifying the e-mail address with which you created your existing account.

If you already have a suitable NVIDIA Enterprise Account for accessing your licenses, follow the Login link in the instructions for using the certificate to log in to the <u>NVIDIA</u> <u>Enterprise Application Hub</u>, go to the NVIDIA Licensing Portal, and download your NVIDIA vGPU software. For details, refer to <u>Downloading your NVIDIA vGPU Software</u>.

1.4. Creating your NVIDIA Enterprise Account

If you do not have an NVIDIA Enterprise Account, you must create an account to be able to log in to the NVIDIA Licensing Portal.

If you already have an account, skip this task and go to <u>Downloading your NVIDIA vGPU</u> <u>Software</u>.

However, if you have an account that was created for an evaluation license and you want to access licenses that you purchased, you must repeat the registration process when you receive your purchased licenses. You can choose to create a separate account for your purchased licenses or link your existing account for an evaluation license to the account for your purchased licenses.

- To create a separate account for your purchased licenses, perform this task, specifying a different e-mail address than the address with which you created your existing account.
- To link your existing account for an evaluation license to the account for your purchased licenses, follow the instructions in <u>Linking an Evaluation Account to an NVIDIA Enterprise</u> <u>Account for Purchased Licenses</u>, specifying the e-mail address with which you created your existing account.

Before you begin, ensure that you have your order confirmation message.

- 1. In the instructions for using your NVIDIA Entitlement Certificate, follow the **Register** link.
- 2. Fill out the form on the NVIDIA Enterprise Account Registration page and click Register.

⊲anvidia NVIDIA Enterpris	ise Account Registration	
	Plenen myder with ynar copronte arwal address. I alwady ngalawrad, dak 'nr I yna need anddrada with galawrad (galawrad) a fallang arwadynan arwan arwan.	
Entitlement PAK IDEntitiement	wit (C)	
Primary Contact		
* First Name: * Claiming Entitlement as:		
Primary Contact Detai	ails	
* Street 1: Street 2:	t. 2	
* City: * State/Province: * Postal Code/Zip Code:	e: Select an Option 🔻	
* Phone: * Job Role:	er ler Select an Opsion •	
Send me the latest enterprise news, a By registering, you agree to INVIDIA Accord	a encouncements, and more from IV/DIA (can unsubscribe at any time.	

A message confirming that an account has been created appears, and an e-mail instructing you to set your NVIDIA password is sent to the e-mail address you provided.

3. Open the e-mail instructing you to set your password and click **SET PASSWORD**.

Dear
Thank you for your business and welcome to NVIDIA.
Login
To get started, please select the link below to set up your password:
SET PASSWORD
Once you have reset your password, please allow 24 business hours before attempting to <u>LOGIN</u> .
Questions? Please contact us at <u>NVIDIA Enterprise</u> <u>Support</u> .
Thank you for your interest in NVIDIA products.
Best Regards, NVIDIA Team

Note: After you have set your password during the initial registration process, you will be able to log in to your account within 15 minutes. However, it may take up to 24 business hours for your entitlement to appear in your account.

For your account security, the **SET PASSWORD** link in this e-mail is set to expire in 24 hours.

4. Enter and re-enter your new password, and click **SUBMIT**.

SET NEW PASSV	/ORD	
	New password: ••••••••	
	Re-type password:	
	Between 9 and 54 characters (inclusive)	SUBMIT
	At least one lowercase letter	
	At least one uppercase letter	
	✓ At least one number	
	At least one special character()	
	Password Match	
		Terms & Conditions Legal Info Privacy Policy Copyright © 2019 NVIDIA Corporation

A message confirming that your password has been set successfully appears.

NIDIA Password		
	SUCCESS Your password has been updated.	
	Terms & Conditions Legal Info Privacy Policy Copyright © 2019 NVIDIA Corporation	

You are then automatically directed to log in to the NVIDIA Licensing Portal with your new password.

1.5. Linking an Evaluation Account to an NVIDIA Enterprise Account for Purchased Licenses

If you have an account that was created for an evaluation license, you must repeat the registration process when you receive your purchased licenses. To link your existing account

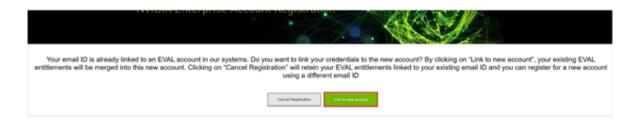
for an evaluation license to the account for your purchased licenses, register for an NVIDIA Enterprise Account with the e-mail address with which you created your existing account.

If you want to create a separate account for your purchased licenses, follow the instructions in <u>Creating your NVIDIA Enterprise Account</u>, specifying a different e-mail address than the address with which you created your existing account.

- 1. In the instructions for using the NVIDIA Entitlement Certificate **for your purchased licenses**, follow the **Register** link.
- 2. Fill out the form on the **NVIDIA Enterprise Account Registration** page, specifying the email address with which you created your existing account, and click **Register**.

≪∎ nvidia NVIDIA Enterpris	e Account	at Registration	
	If you need assist	Please register with your concrete erral address. If alwady negliaterad, click: en- stance with registration, please erral (
Entitlement			
PAK ID/Entitlement I	D: 11100000000001	_	
Primary Contact			
* Email Address: * First Name:		* Last Name:	
* Claiming Entitlement as:	Select an Option	Last name.	
Primary Contact Detail	S		
* Location:	Select a Location	•	
* Street 1:			
Street 2:			
* City:			
* State/Province:	Select an Option	•	
* Postal Code/Zip Code:			
	Select an Option	•	

3. When a message stating that your e-mail address is already linked to an evaluation account is displayed, click **LINK TO NEW ACCOUNT**.



Log in to the NVIDIA Licensing Portal with the credentials for your existing account.

1.6. Downloading your NVIDIA vGPU Software

Before you begin, ensure that you have your order confirmation message and have created an NVIDIA Enterprise Account.

- 1. Visit the <u>NVIDIA Enterprise Application Hub</u> by following the **Login** link in the instructions for using your NVIDIA Entitlement Certificate or when prompted after setting the password for your NVIDIA Enterprise Account.
- 2. When prompted, provide your e-mail address and password, and click **LOGIN**.

NVIDIA APPLICATIO	N HUB LOGIN		
	Username		
	Password	_	
		LOGIN	
		Forgot password? Need help	37
			Terms and Conditions Legal Info Privacy Policy Copyright © 2019 NVIDIA Corporation

On the NVIDIA APPLICATION HUB page that opens, click NVIDIA LICENSING PORTAL.
 The NVIDIA Licensing Portal dashboard page opens.

☆ DASHBOARD				
ENTITLEMENTS			Organization Example Corporation	Q
LICENSE SERVERS	E 100			
象 SOFTWARE DOWNLOADS	Entitlements	MANAGE ENTITLEMENTS	License Servers MANAGE SEF	RVERS CREATE SERVER
UIRTUAL GROUPS	ENTITLEMENT / FEATURE EXPIRATION	ALLOCATED / TOTAL	▲ LICENSE SERVER / FEATURE	IN USE / ALLOCATED
			You do not have any license servers. Wou	Ild you like to create one?
9, USER MANAGEMENT			E CREATE LICENSE SI	ERVER
C ENTERPRISE SUPPORT				

Note: Your entitlement might not appear on the NVIDIA Licensing Portal dashboard page until 24 business hours after you set your password during the initial registration process.

4. In the NVIDIA Licensing Portal dashboard page opens, click the down arrow next to each entitlement listed to view details of the NVIDIA vGPU software that you purchased.

NIDIA. LICENSING	Dashboard		NVIDIA Application Hub William Bradsh	haw (ORG_ADMIN)
G DASHBOARD				
ENTITLEMENTS			Organization Example Corporation	
LICENSE SERVERS				
SOFTWARE DOWNLOADS	Entitlements	MANAGE ENTITLEMENTS	License Servers	NAGE LICENSE SERVE
	ENTITLEMENT / FEATURE EXPIRATION	ALLOCATED / TOTAL 🗘	LICENSE SERVER / FEAT	IN USE / ALLOCATED
Ag USER MANAGEMENT	V	14400000		
C ENTERPRISE SUPPORT	GRID-Virtual never expires	0 / 2400	You do not have any license servers. Would yo	u like to create one?
	GRID-Virtual never expires	0 / 2400	圆 CREATE LICENSE SERVE	2
	SUMS 2022-10-25	2400 / 2		
	V			
	Quadro-Virtu never expires	0 / 9332		
	GRID-Virtual never expires	0 / 9332		
	SUMS 2022-10-25	9332 / 9		
	Quadro-Virtu 2022-08-20	0/3		
	GRID-Virtual 2022-08-20	0/3		
	V 000000000000000000000000000000000000			
	GRID-Virtual never expires	0/30		
	GRID-Virtual never expires	0/30		
	SUMS 2022-01-04	30/30		

- 5. In the left navigation pane of the NVIDIA Licensing Portal dashboard, click **SOFTWARE DOWNLOADS**.
- 6. On the **Product Download** page that opens, set the **Product Family** option to **vGPU** and follow the **Download** link for the brand and version of your chosen hypervisor for the release of NVIDIA vGPU software that you are using, for example, NVIDIA vGPU for vSphere 6.7 for NVIDIA vGPU software release 13.0.
 - Note: To be able to download any additional software that you need for your NVIDIA vGPU software deployment, for example, the license server software, you **must** set the **Product Family** option to vGPU. Otherwise, the **ADDITIONAL SOFTWARE** button does not appear on the **Product Download** page and the pop-up window for downloading additional software is not opened.

💽 NVIDIA. LICENSING	Software Downloads			NVIDIA Application Hub W	lliam Bradshaw (ORG_ADMI)	l) Logou
🖧 DASHBOARD						
ENTITLEMENTS				Organization Example Corporation		Q
LICENSE SERVERS						
SOFTWARE DOWNLOADS	68	8			ADDITIONAL SOF	
VIRTUAL GROUPS	FEATURED A	LL AVAILABLE			ADDITIONAL SOF	TWARE V
AUDIT HISTORY	Product Family: O All vGPU	HPC O PGI				
见 USER MANAGEMENT	VMware vSphere V Platf	orm Version: All 🗸 Select Platfor	rm Version Search		CLEAR CO	LUMNS 🗸
C ENTERPRISE SUPPORT	PLATFORM	PLATFORM VERSION	PRODUCT VERSION	DESCRIPTION	RELEASE DATE	
	& VMware vSphere	7.0	10.3	NVIDIA vGPU for vSphere 7.0	2020-07-16 D	ownload
	& VMware vSphere	6.7	10.3	NVIDIA vGPU for vSphere 6.7	2020-07-16 D	ownload
	& VMware vSphere	6.5	10.3	NVIDIA vGPU for vSphere 6.5	2020-07-16 D	ownload
	& VMware vSphere	7.0	11.0	NVIDIA vGPU for vSphere 7.0	2020-07-08 D	ownload
	W VMware vSphere	6.7	11.0	NVIDIA vGPU for vSphere 6.7	2020-07-08 D	ownload
	\odot	software lifecycle for NVIDIA virtual (cs.nvidia.com/grid/news/index.html		

If the brand and version of your chosen hypervisor for the release of NVIDIA vGPU software that you are using aren't displayed, click **ALL AVAILABLE** to display a list of all NVIDIA vGPU software available for download. Use the drop-down lists or the search box to filter the software listed.

- 7. When prompted to accept the license for the software that you are downloading, click **AGREE & DOWNLOAD**.
- 8. When the browser asks what it should do with the file, select the option to save the file.

After the download starts, a pop-up window opens for you to download any additional software that you might need for your NVIDIA vGPU software deployment.

ownload additional software you may need for your NVIDIA vGPU software	
epoyment	
2020.05 64-bit License Manager for Windows	
2020.05 64-bit License Manager for Linux	
2019.11 64-bit License Manager for Windows	
2019.11 64-bit License Manager for Linux	
2018.10 32-bit License Manager for Windows	
2018.10 License Manager for Linux	
GPUMode Change Utility for Tesla M60 and M6 GPUs	
NVIDIA Virtual GPU Management Pack for vRealize Operations 2.0	
NVIDIA Virtual GPU Management Pack for vRealize Operations 1.1	

- 9. In the pop-up window, follow the links to download any additional software that you need for your NVIDIA vGPU software deployment.
 - a). Follow the link to the NVIDIA vGPU software license server software for your license server host machine's operating system, for example, **License Manager for Windows**.
 - b). If you are using an NVIDIA Tesla[™] M60 or M6 GPU and think you might need to change its mode, follow the **Mode Change Utility** link.

For details about when you need to change the mode, see <u>Switching the Mode of a</u> <u>Tesla M60 or M6 GPU</u>.

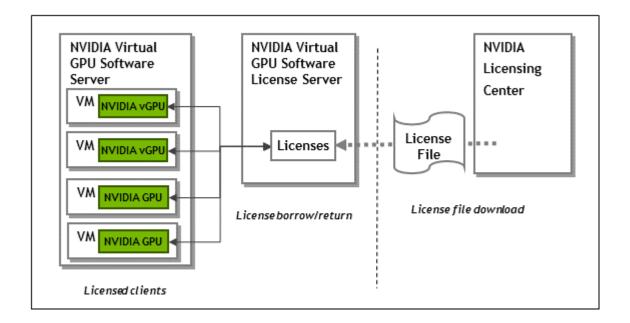
Chapter 2. Installing Your NVIDIA vGPU Software License Server and License Files

The NVIDIA vGPU software License Server is used to serve a pool of floating licenses to NVIDIA vGPU software licensed products. The license server is designed to be installed at a location that is accessible from a customer's network, and be configured with licenses obtained from the NVIDIA Licensing Portal.

Note: These instructions cover only the installation and configuration of a single-instance license server. If you require high availability for the license server, refer to <u>Virtual GPU License</u> <u>Server User Guide</u> for instructions for how to configure a primary and secondary license server. If you do not register a secondary server when the primary server is registered, you **cannot** register the secondary server later.

2.1. Overview of NVIDIA vGPU software Licensing

NVIDIA vGPU software is a licensed product. Licensed vGPU functionalities are activated during guest OS boot by the acquisition of a software license served over the network from an NVIDIA vGPU software license server. The license is returned to the license server when the guest OS shuts down.



The licenses served by the NVIDIA vGPU software license server are obtained from the NVIDIA Licensing Portal as downloadable license files and installed into the license server through its management interface.

2.2. Platform Configuration Requirements

- The platform must have a fixed (unchanging) IP address. The IP address may be assigned dynamically by DHCP or statically configured but must be constant.
- The platform must have at least one unchanging Ethernet MAC address to be used as a unique identifier when registering the server and generating licenses in the NVIDIA Licensing Portal.
- ▶ The platform's date and time must be set accurately. NTP is recommended.

2.3. Installing the Java Runtime Environment

If a suitable Java runtime environment (JRE) version is not already installed on your system, install a supported JRE before running the NVIDIA license server installer.

Install a JRE for the same the processor architecture as the license server that you will install: If you will install a 32-bit license server, install a 32-bit JRE; if you will install a 64-bit license server, install a 64-bit JRE. 1. Download a supported Oracle Java SE JRE or OpenJDK JRE.

CAUTION: To ensure that you download a 64-bit JRE, do **not** download the JRE from www.java.com. Either a 32-bit JRE is downloaded or the bitness of the downloaded version is determined from your browser, which might be 32-bit even on a 64-bit operating system.

- Download Oracle Java SE JRE from the <u>java.com</u>: <u>Java + You</u> page.
- Download OpenJDK JRE from the <u>Community builds using source code from OpenJDK</u> <u>project</u> on GitHub.
- 2. Install the JRE that you downloaded.
 - Oracle Java SE JRE installation:

🗲 🕞 📓 http://java.com	n/en/download/ie_manual.jsp P - C Download Java for Windows ×	☆ 않
الله المعاني (Java ^r	Search Download Help	∝ ^
Help Resources	Download Java for Windows	
 » What is Java? » Remove Older Versions » Sisable Java » Error Messages 	Recommended Version 8 Update 60 (filesize: 571 KB) Release date August 18, 2015	
» <u>Troubleshoot Java</u> » <u>Other Help</u>	Java Setup - Welcome	
Windows 64-bit Users		
Do you use both 32-bit and 64-bit browsers? » FAQ about 64-bit Java for	Welcome to Java	
Windows	Java provides safe and secure access to the world of amazing Java content. From business solutions to helpful utilities and entertainment, Java	ır
Offline Installation	makes your internet experience come to life.	
Trouble downloading? Try the <u>offline installer</u>	Note: No personal information is gathered as part of our install process. <u>Click here</u> for more information on what we do collect.	
	Click Install to accept the <u>license agreement</u> and install Java now.	-
	Change destination folder Cancel Install > 1, JVM, VM	<i>I</i> ,

• OpenJDK JRE installation:



3. Set the JAVA_HOME system variable to the full path to the jre folder of your JRE installation.

For example:

- ► For 64-bit Oracle Java SE JRE: C:\Program Files\Java\jre1.8.0_191
- ► For 32-bit Oracle Java SE JRE: C:\Program Files (x86)\Java\jre1.8.0_191
- For 64-bit OpenJDK JRE: C:\Program Files\ojdkbuild\java-1.8.0openjdk-1.8.0.201-1\jre
- For 32-bit OpenJDK JRE: C:\Program Files (x86)\ojdkbuild\java-1.8.0openjdk-1.8.0.201-1\jre

Ensure that the path does **not** include any trailing characters, such as a slash or a space.

The license server installer warns users if the path in the JAVA_HOME system environment variable appears to be incorrect.

If you are upgrading to a new version of the JRE, update the value of the JAVA_HOME system variable to the full path to the jre folder of your **new** JRE version.

 Ensure that the Path system variable contains the path to the java.exe executable file. The installer may have already added this path for you.

For example:

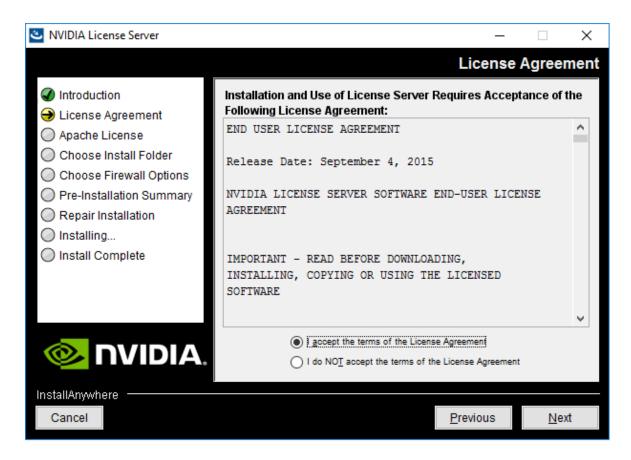
- ► For 64-bit Oracle Java SE JRE: C:\Program Files\Java\jre1.8.0_191\bin
- ► For 32-bit Oracle Java SE JRE: C:\Program Files (x86)\Java\jre1.8.0_191\bin
- For 64-bit OpenJDK JRE: C:\Program Files\ojdkbuild\java-1.8.0openjdk-1.8.0.201-1\bin
- For 32-bit OpenJDK JRE: C:\Program Files (x86)\ojdkbuild\java-1.8.0openjdk-1.8.0.201-1\bin

Note: If a 32-bit JRE and 64-bit JRE are installed, ensure that the path to the 64-bit JRE appears before the path to the 32-bit JRE in the Path system variable.

2.4. Installing the License Server Software

1. Unzip the license server installer and run setup.exe.

2. Accept the EULA for the license server software and the Apache Tomcat software used to support the license server's management interface.



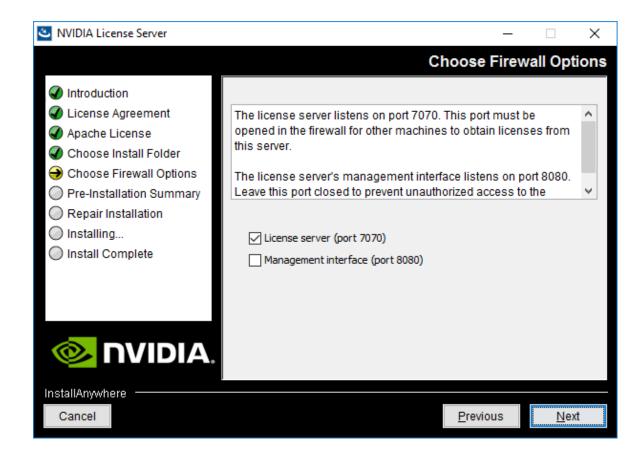
3. Choose the destination folder where you want the license server software to be installed.

NVIDIA License Server	- 🗆 X
	Choose Install Folde
 Introduction License Agreement Apache License Choose Install Folder Choose Firewall Options Pre-Installation Summary Repair Installation Installing 	Please choose a destination folder for this installation. Destination folder: C:\Program Files (x86)\WVIDIA\License Server
Install Complete	<u>R</u> estore Default Folder Ch <u>o</u> ose
InstallAnywhere Cancel	Previous Next

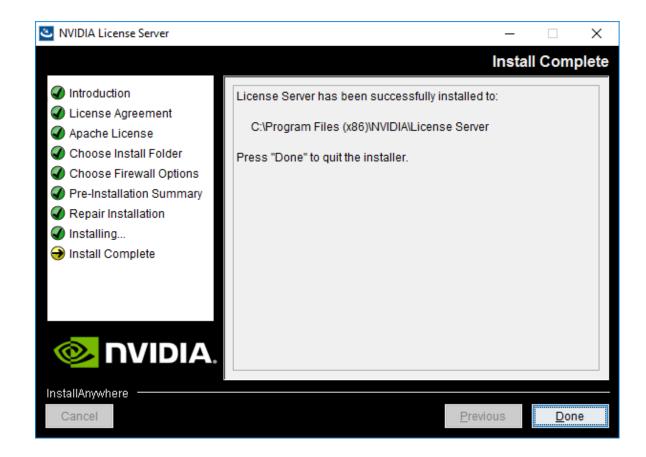
4. In the **Choose Firewall Options** dialog box, select the ports to be opened in the firewall.

To enable remote clients to access licenses from the server and prevent remote access to the management interface, use the default setting, which sets ports as follows:

- > Port 7070 is open to enable remote clients to access licenses from the server.
- Port 8080 is closed to ensure that the management interface is available only through a web browser running locally on the license server host.



5. After installation has completed successfully, click **Done** to exit the installer.



2.5. Obtaining the License Server's MAC Address

The license server's Ethernet MAC address uniquely identifies your server to the NVIDIA Licensing Portal. You will need this address to register your license server with the NVIDIA Licensing Portal to generate license files.

- 1. Open a web browser on the license server host and connect to the URL http://localhost:8080/licserver.
- 2. In the license server management interface, select **Configuration**.
- 3. On the **License Server Configuration** page that opens, in the **Server host ID** drop-down list, select the platform's ETHERNET address.

Attp://localhost8080/licsen	ver/serverProperties_view.acti	ion ρ → 🥺 NVIDIA License Client Mana ×	 ش
License Server ≯ Licensed Clients	License S Properties	erver Configuration	
 <u>Reservations</u> <u>Licensed Feature Usage</u> <u>License Management</u> <u>Configuration</u> 	Property Server host ID	Value O6407E06282C (ETHERNET)	Description Server's host ID used to uniquely identify the server to the NVIDIA licensing portal. If multiple ETHERNET IDs are available, select one and use consistently with licensing portal.
License Client Manager About Settings	General properties Server Version Server Status License Generation	2017.11 Alive	Server's executable version Indicates server state

2.6. Managing your License Servers and Getting your License Files

To be able to download NVIDIA vGPU software licenses, you must create at least one license server on the NVIDIA Licensing Portal and allocate licenses to the server. After creating a license server and allocating licenses to it, you can download your license file.

2.6.1. Creating a License Server on the NVIDIA Licensing Portal

To be able to download NVIDIA vGPU software licenses, you must create at least one license server on the NVIDIA Licensing Portal. Creating a license server on the NVIDIA Licensing Portal registers your license server host with the NVIDIA Licensing Portal through the MAC address of the host.

- 1. In the NVIDIA Licensing Portal, navigate to the organization or virtual group for which you want to create the license server.
 - a). If you are not already logged in, log in to the <u>NVIDIA Enterprise Application Hub</u> and click **NVIDIA LICENSING PORTAL** to go to the NVIDIA Licensing Portal.
 - b). **Optional:** If your assigned roles give you access to multiple virtual groups, select the virtual group for which you are creating the license server from the list of virtual groups at the top right of the page.

If no license servers have been created for your organization or virtual group, the NVIDIA Licensing Portal dashboard displays a message asking if you want to create a license server.

🔏 DASHBOARD				
ENTITLEMENTS			Organization Example Corporation	Q
LICENSE SERVERS	E 100			
SOFTWARE DOWNLOADS	Entitlements	MANAGE ENTITLEMENTS		CREATE SERVER
VIRTUAL GROUPS	ENTITLEMENT / FEATURE EXPIRATION	ALLOCATED / TOTAL	▲ LICENSE SERVER / FEATURE IN USE / ALL	OCATED
) HISTORY			You do not have any license servers. Would you like	to create one?
S USER MANAGEMENT			CREATE LICENSE SERVER	
☐ ENTERPRISE SUPPORT				

On the NVIDIA Licensing Portal dashboard, click CREATE LICENSE SERVER.
 The Create License Server pop-up window opens.

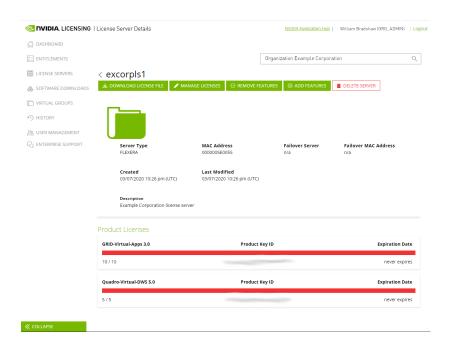
Server Name	Product	Licenses
Name this license server	Select a product	Y 1 ○ AI
Description	Added Products	
Provide a short description	Product	Count
	No pro	ducts have been added yet
MAC Address		
MAC Address (XX:XX:XX:XX:XX:XX:XX:XX:XX:XX:XX)		
Failover server configuration is optional. If configuring, you must provide a name AND MAC address		
Failover License Server		
Failover License Server		
Failover MAC Address		
Failover MAC Address		

- 3. Provide the details of your license server.
 - a). In the **Server Name** field, enter the host name of the license server.
 - b). In the **Description** field, enter a text description of the license server. This description is required and will be displayed on the details page for the license server that you are creating.
 - c). In the **MAC Address** field, enter the MAC address of your license server.
- 4. Add the licenses for the products that you want to allocate to this license server. For each product, add the licenses as follows:
 - a). From the **Product** drop-down list, select the product for which you want to add licenses.
 - b). In the **Licenses** field, enter the number of licenses for the product that you want to add.
 - c). Click **ADD**.
- 5. Leave the Failover License Server and Failover MAC Address fields unset.
- 6. Click **CREATE LICENSE SERVER**.

2.6.2. Downloading a License File

Each license server that you create has a license file associated with it. The license file contains all the licenses that you allocated to the license server. After downloading the license file, you can install it on the license server host associated with the license server on the NVIDIA Licensing Portal.

- 1. In the NVIDIA Licensing Portal, navigate to the organization or virtual group for which you want to download the license file.
 - a). If you are not already logged in, log in to the <u>NVIDIA Enterprise Application Hub</u> and click **NVIDIA LICENSING PORTAL** to go to the NVIDIA Licensing Portal.
 - b). **Optional:** If your assigned roles give you access to multiple virtual groups, select the virtual group for which you are downloading the license file from the list of virtual groups at the top right of the page.
- 2. In the list of license servers on the NVIDIA Licensing Portal dashboard, select the license server whose associated license file you want to download.
- 3. In the **License Server Details** page that opens, review the licenses allocated to the license server.



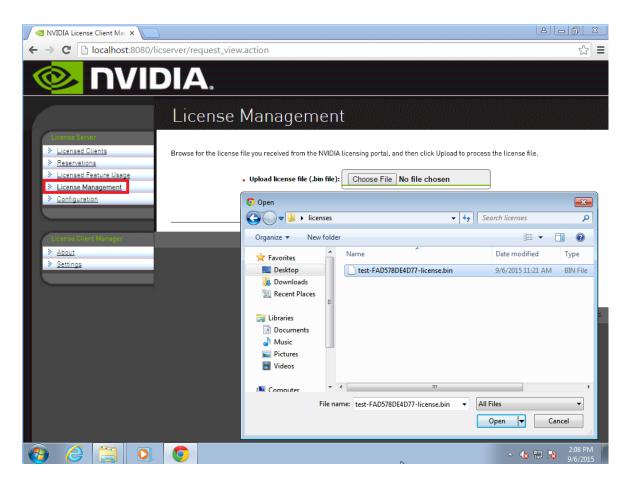
4. Click **DOWNLOAD LICENSE FILE** and save the .bin license file to your license server for installation.

2.7. Installing a License

NVIDIA vGPU software licenses are distributed as .bin files for download from the NVIDIA Licensing Portal.

Before installing a license, ensure that you have downloaded the license file from the NVIDIA Licensing Portal.

- 1. In the license server management interface, select **License Management**.
- 2. On the **License Management** page that opens, click **Choose File**.



- 3. In the file browser that opens, select the .bin file and click **Open**.
- 4. Back on the **License Management** page, click **Upload** to install the license file on the license server.

The license server should confirm the successful installation of the license file.

+ttp://localhost:8080/lics	erver/response_upload.action 🛛 🔎 マ 🖒 🔯 NVIDIA License Client Man 🗙 📑
	DIA.
	License Management
License Server <u>Licensed Clients</u>	VSuccessfully applied license file to license server.
<u>Reservations</u> <u>Licensed Feature Usage</u>	Browse for the license file you received from the NVIDIA licensing portal, and then click Upload to process the license file.
License Management Configuration	Upload license file (.bin file): Browse
License Client Manager	Cancel
> About	
Settings	Copyright [c] 2018 NVIDIA Corporation. All Rights Reserved. 2018.06.0.243045

Chapter 3. Installing and Configuring NVIDIA vGPU Manager and the Guest Driver

Before installing and configuring NVIDIA vGPU Manager and the Guest Driver, ensure that a VM running a supported Windows guest OS is configured in your chosen hypervisor.

The factory settings of some supported GPU boards are incompatible with NVIDIA vGPU software. Before configuring NVIDIA vGPU software on these GPU boards, you must configure the boards to change these settings.

3.1. Switching the Mode of a GPU that Supports Multiple Display Modes

Some GPUs support displayless and display-enabled modes but must be used in NVIDIA vGPU software deployments in displayless mode.

The GPUs listed in the following table support multiple display modes. As shown in the table, some GPUs are supplied from the factory in displayless mode, but other GPUs are supplied in a display-enabled mode.

GPU	Mode as Supplied from the Factory
NVIDIA A40	Displayless
NVIDIA RTX A5000	Display enabled
NVIDIA RTX A6000	Display enabled

A GPU that is supplied from the factory in displayless mode, such as the NVIDIA A40 GPU, might be in a display-enabled mode if its mode has previously been changed.

To change the mode of a GPU that supports multiple display modes, use the displaymodeselector tool, which you can request from the <u>NVIDIA Display Mode Selector</u> <u>Tool</u> page on the NVIDIA Developer website.



Note:

Only the following GPUs support the displaymodeselector tool:

- NVIDIA A40
- NVIDIA RTX A5000
- NVIDIA RTX A6000

Other GPUs that support NVIDIA vGPU software do not support the displaymodeselector tool and, unless otherwise stated, do not require display mode switching.

3.2. Switching the Mode of a Tesla M60 or M6 GPU

Tesla M60 and M6 GPUs support compute mode and graphics mode. NVIDIA vGPU requires GPUs that support both modes to operate in graphics mode.

Recent Tesla M60 GPUs and M6 GPUs are supplied in graphics mode. However, your GPU might be in compute mode if it is an older Tesla M60 GPU or M6 GPU or if its mode has previously been changed.

To configure the mode of Tesla M60 and M6 GPUs, use the gpumodeswitch tool provided with NVIDIA vGPU software releases. If you are unsure which mode your GPU is in, use the gpumodeswitch tool to find out the mode.

Note:

Only Tesla M60 and M6 GPUs support the <code>gpumodeswitch</code> tool. Other GPUs that support NVIDIA vGPU do not support the <code>gpumodeswitch</code> tool and, unless otherwise stated, do not require mode switching.

Even in compute mode, Tesla M60 and M6 GPUs do **not** support NVIDIA Virtual Compute Server vGPU types. Furthermore, vCS is not supported on any GPU on Citrix Hypervisor.

For more information, refer to *gpumodeswitch User Guide*.

3.3. Installing the NVIDIA Virtual GPU Manager

Before guests enabled for NVIDIA vGPU can be configured, the NVIDIA Virtual GPU Manager must be installed in your chosen hypervisor. The process for installing the NVIDIA Virtual GPU Manager depends on the hypervisor that you are using.

If you need more detailed instructions, refer to the appropriate NVIDIA vGPU installation guide.

3.3.1. Installing the NVIDIA Virtual GPU Manager on VMware vSphere

For all supported VMware vSphere releases, the NVIDIA Virtual GPU Manager package is distributed as a software component in a ZIP archive. For supported releases **before** VMware

vSphere 7.0, the NVIDIA Virtual GPU Manager package is also distributed as a vSphere Installation Bundle (VIB) file.

- 1. Copy the NVIDIA Virtual GPU Manager package file to the ESXi host.
- 2. Put the ESXi host into maintenance mode.
 - \$ esxcli system maintenanceMode set --enable true
- 3. Run the esxcli command to install the NVIDIA Virtual GPU Manager from the package file.
 - ► For a software component, run the following command:

```
$ esxcli software vib install -d /vmfs/volumes/datastore/software-component.zip
datastore
```

The name of the VMFS datastore to which you copied the software component. *software-component*

The name of the file that contains the software component.

► For a VIB file, run the following command:

```
$ esxcli software vib install -v directory/NVIDIA**.vib
directory
```

The absolute path to the directory to which you copied the VIB file. You must specify the absolute path even if the VIB file is in the current working directory.

4. Exit maintenance mode.

```
\$ esxcli system maintenanceMode set --enable false
```

5. Reboot the ESXi host.

\$ reboot

6. Verify that the NVIDIA kernel driver can successfully communicate with the physical GPUs in your system by running the nvidia-smi command without any options.

\$ nvidia-smi

If successful, the nvidia-smi command lists all the GPUs in your system.

3.3.2. Installing the NVIDIA Virtual GPU Manager on Citrix Hypervisor

The NVIDIA Virtual GPU Manager for Citrix Hypervisor is distributed as an RPM Package Manager (RPM) file. It runs in the Citrix Hypervisor Control Domain (dom0) shell.

- 1. Copy the NVIDIA Virtual GPU Manager RPM file to the Citrix Hypervisor dom0 shell.
- 2. Run the rpm command to install the package. [root@xenserver ~] # rpm -iv NVIDIA-**.rpm
- 3. Reboot the Citrix Hypervisor platform. [root@xenserver ~]# shutdown -r now
- 4. After the Citrix Hypervisor host has rebooted, verify the installation of the NVIDIA Virtual GPU Manager package for Citrix Hypervisor by checking for the NVIDIA kernel driver in the list of kernel-loaded modules.

```
[root@xenserver ~]# lsmod |grep nvidia
nvidia 8152994 0
i2c_core 20294 2 nvidia,i2c_
```

3.4. Disabling and Enabling ECC Memory

Some GPUs that support NVIDIA vGPU software support error correcting code (ECC) memory with NVIDIA vGPU. ECC memory improves data integrity by detecting and handling doublebit errors. However, not all GPUs, vGPU types, and hypervisor software versions support ECC memory with NVIDIA vGPU.

On GPUs that support ECC memory with NVIDIA vGPU, ECC memory is supported with Cseries and Q-series vGPUs, but not with A-series and B-series vGPUs. Although A-series and B-series vGPUs start on physical GPUs on which ECC memory is enabled, enabling ECC with vGPUs that do not support it might incur some costs.

On physical GPUs that do not have HBM2 memory, the amount of frame buffer that is usable by vGPUs is reduced. All types of vGPU are affected, not just vGPUs that support ECC memory.

The effects of enabling ECC memory on a physical GPU are as follows:

- ▶ ECC memory is exposed as a feature on all supported vGPUs on the physical GPU.
- In VMs that support ECC memory, ECC memory is enabled, with the option to disable ECC in the VM.
- ECC memory can be enabled or disabled for individual VMs. Enabling or disabling ECC memory in a VM does not affect the amount of frame buffer that is usable by vGPUs.

GPUs based on the Pascal GPU architecture and later GPU architectures support ECC memory with NVIDIA vGPU. These GPUs are supplied with ECC memory enabled.

Tesla M60 and M6 GPUs support ECC memory when used without GPU virtualization, but NVIDIA vGPU does not support ECC memory with these GPUs. In graphics mode, these GPUs are supplied with ECC memory disabled by default.

Some hypervisor software versions do not support ECC memory with NVIDIA vGPU.

If you are using a hypervisor software version or GPU that does not support ECC memory with NVIDIA vGPU and ECC memory is enabled, NVIDIA vGPU fails to start. In this situation, you must ensure that ECC memory is disabled on all GPUs if you are using NVIDIA vGPU.

3.4.1. Disabling ECC Memory

If ECC memory is unsuitable for your workloads but is enabled on your GPUs, disable it. You must also ensure that ECC memory is disabled on all GPUs if you are using NVIDIA vGPU with a hypervisor software version or a GPU that does not support ECC memory with NVIDIA vGPU. If your hypervisor software version or GPU does not support ECC memory and ECC memory is enabled, NVIDIA vGPU fails to start.

Where to perform this task depends on whether you are changing ECC memory settings for a physical GPU or a vGPU.

For a physical GPU, perform this task from the hypervisor host.

For a vGPU, perform this task from the VM to which the vGPU is assigned.

Note: ECC memory must be enabled on the physical GPU on which the vGPUs reside.

Before you begin, ensure that NVIDIA Virtual GPU Manager is installed on your hypervisor. If you are changing ECC memory settings for a vGPU, also ensure that the NVIDIA vGPU software graphics driver is installed in the VM to which the vGPU is assigned.

1. Use nvidia-smi to list the status of all physical GPUs or vGPUs, and check for ECC noted as enabled.

- [...]
- 2. Change the ECC status to off for each GPU for which ECC is enabled.
 - If you want to change the ECC status to off for all GPUs on your host machine or vGPUs assigned to the VM, run this command:

```
# nvidia-smi -e 0
```

If you want to change the ECC status to off for a specific GPU or vGPU, run this command:

```
# nvidia-smi -i id -e 0
```

id is the index of the GPU or vGPU as reported by nvidia-smi.

This example disables ECC for the GPU with index 0000:02:00.0.

nvidia-smi -i 0000:02:00.0 -e 0

- 3. Reboot the host or restart the VM.
- 4. Confirm that ECC is now disabled for the GPU or vGPU.

```
# nvidia-smi -q
==========NVSMI LOG===========
                                 : Mon Jul 26 18:37:53 2021
Timestamp
Driver Version
                                 : 470.63
Attached GPUs
                                 : 1
GPU 0000:02:00.0
[...]
   Ecc Mode
                                  : Disabled
       Current
                                  : Disabled
       Pending
[...]
```

3.4.2. Enabling ECC Memory

If ECC memory is suitable for your workloads and is supported by your hypervisor software and GPUs, but is disabled on your GPUs or vGPUs, enable it.

Where to perform this task depends on whether you are changing ECC memory settings for a physical GPU or a vGPU.

- For a physical GPU, perform this task from the hypervisor host.
- For a vGPU, perform this task from the VM to which the vGPU is assigned.



Note: ECC memory must be enabled on the physical GPU on which the vGPUs reside.

Before you begin, ensure that NVIDIA Virtual GPU Manager is installed on your hypervisor. If you are changing ECC memory settings for a vGPU, also ensure that the NVIDIA vGPU software graphics driver is installed in the VM to which the vGPU is assigned.

1. Use nvidia-smi to list the status of all physical GPUs or vGPUs, and check for ECC noted as disabled.

```
[...]
```

- 2. Change the ECC status to on for each GPU or vGPU for which ECC is enabled.
 - If you want to change the ECC status to on for all GPUs on your host machine or vGPUs assigned to the VM, run this command:

```
# nvidia-smi -e 1
```

If you want to change the ECC status to on for a specific GPU or vGPU, run this command:

```
# nvidia-smi -i id -e 1
```

id is the index of the GPU or vGPU as reported by nvidia-smi.

This example enables ECC for the GPU with index 0000:02:00.0.

nvidia-smi -i 0000:02:00.0 -e 1

- 3. Reboot the host or restart the VM.
- 4. Confirm that ECC is now enabled for the GPU or vGPU.

```
# nvidia—smi —q
```

=======NVSMI LOG==========

Timestamp Driver Version	: Mon Jul 26 18:37:53 2021 : 470.63
Attached GPUs GPU 0000:02:00.0 []	: 1
Ecc Mode Current Pending	: Enabled : Enabled
[]	

3.5. Attaching an NVIDIA vGPU Profile to a VM

To attach an NVIDIA vGPU profile to a virtual machine (VM), you must configure the VM hardware. The process for attaching an NVIDIA vGPU profile to a VM depends on the hypervisor that you are using.

3.5.1. Changing the Default Graphics Type in VMware vSphere 6.5 and Later

Before changing the default graphics type, ensure that the ESXi host is running and that all VMs on the host are powered off.

- 1. Log in to vCenter Server by using the vSphere Web Client.
- 2. In the navigation tree, select your ESXi host and click the **Configure** tab.
- 3. From the menu, choose **Graphics** and then click the **Host Graphics** tab.
- 4. On the **Host Graphics** tab, click **Edit**.

vmware [,] vSphere	Web Client ♠= Ů Administrator@PSG-HOME.LOCAL - Help -
Navigator I	🔋 192.168.11.30 🛛 🥵 🕞 🗊 🕵 🛛 🎯 Actions 🗸 📃
Back	Getting Started Summary Monitor Configure Permissions VMs Resource Pools Datastores Networks
 Image: Image: Im	Host Graphics Graphics Devices Advanced Host Graphics Settings VIItual Machines Default graphics type: Default graphics type: Shared
▶ 192.168.11.30	VM Startup/Shutdown Agent VM Settings Shared Shared Agent VM Settings Shared passthrough GPU assignment policy: Spread VMs across GPUs (best performance) Swap file location Image: Compatibility assignment policy: Spread VMs across GPUs (best performance) • System Licensing Image: Compatibility assignment policy: Spread VMs across GPUs (best performance) • Licensing Image: Compatibility assignment policy: Spread VMs across GPUs (best performance) • Time Configuration Authentication Services Spread VMs across GPUs (best performance) Certificate Power Management Advanced System Settings System Resource Reservation System Swap Image: Compatibility of the settings Image: Compatibility of the settings Image: Compatibility of the settings • Hardware Processors Image: Compatibility of the settings Image: Compatibility of the settings • Memory Graphics Image: Compatibility of the settings Image: Compatibility of the settings • Processors Image: Compatibility of the settings Image: Compatibility of the settings Image: Compatibility of the settings • Processors Image: Compatibility of the settings Image: Compatibility of the settings Image: Compatibility
<	PCI Devices ✓ Virtual Flash ✓ :: →

5. In the **Edit Host Graphics Settings** dialog box that opens, select **Shared Direct** and click **OK**.

192.168.11.30 - Edit Host Graphics Settings	?
A Settings will take effect after restarting the host or "xorg" service.	
 Shared VMware shared virtual graphics 	
 Shared Direct Vendor shared passthrough graphics 	
 Shared passthrough GPU assignment policy: Spread VMs across GPUs (best performance) Group VMs on GPU until full (GPU consolidation) 	
ОК	Cancel

After you click OK, the default graphics type changes to Shared Direct.

6. Click the **Graphics Devices** tab to verify the configured type of each physical GPU on which you want to configure vGPU.

The configured type of each physical GPU must be Shared Direct. For any physical GPU for which the configured type is Shared, change the configured type as follows:

a). On the **Graphics Devices** tab, select the physical GPU and click the **Edit icon**.

4	Host Graphics Graphics Dev	ices			
Time Configuration	Graphics Devices				
Authentication Services	Giaphics Devices				
Certificate	/				Q Filter
Power Management	Name	Vendor	Active Type	Configured Type	Memory
Advanced System Settings	NVIDIATesla M60	NVIDIA Corporation	Shared	Shared	7.98 GB
System Resource Reservation	NVIDIATesla M60	NVIDIA Corporation	Shared	Shared	7.99 GB
Security Profile					
System Swap					
Host Profile					
Hardware	A Find	•			2 items 📑 Export 👻 🎦 Copy
Processors	And a second state of south the second state	- device BURDIATESIS BROW	-		
Memory	VMs associated with the graphic	s device "NVIDIATESIa M60"			

b). In the **Edit Graphics Device Settings** dialog box that opens, select **Shared Direct** and click **OK**.

Getting Started Summary Monitor	Configure Pe	ermissions VMs	Resource Pools Datastores	Networks Update Manager			
"	Host Graphic	Graphics Devic	es				
Authentication Services Certificate	Graphics Dev	ices					
Power Management	/						Q Filter
Advanced System Settings	Name		Vendor	Active Type	Configu	red Type	Memory
System Resource Reservation	NVIDIATesla M60		NVIDIA Corporation	Shared	Share	d	7.98 GB
Security Profile	NVIDIATesla	M60	NVIDIA Corporation	Shared	Share	d	7.98 GB
System Swap							
Host Profile							
- Hardware		NVIDIATesla M60	- Edit Graphics Device Settin	gs	?		2 items 🕞 Export 🗝 🏥 Copy 🗸
Processors		A			_		Zitems Expert (Copy)
Memory	VMs associat		ake effect after restarting the ho	st or "xorg" service.			
Graphics		Shared VMware sha	ared virtual graphics				😨 🛛 Q Filter 🗸
Power Management	Name	 Shared Dire 	_ ·		Used Space	Host CPU	Host Mem
PCI Devices		0	shared passthrough graphics				
Virtual Flash Resource Management							
Virtual Flash Host Swap Cache Configuration				OK Cancel			
-	84						Objects 📑 Export 📑 Copy 🚽
4							

7. Restart the ESXi host **or** stop and restart the Xorg service if necessary and nv-hostengine on the ESXi host.

To stop and restart the Xorg service and nv-hostengine, perform these steps:

a). VMware vSphere releases before 7.0 Update 1 only: Stop the Xorg service.

As of VMware vSphere 7.0 Update 1, the Xorg service is no longer required for graphics devices in NVIDIA vGPU mode.

b). Stop nv-hostengine.

[root@esxi:~] nv-hostengine -t

- c). Wait for 1 second to allow nv-hostengine to stop.
- d). Start nv-hostengine. [root@esxi:~] **nv-hostengine -d**
- e). VMware vSphere releases before 7.0 Update 1 only: Start the Xorg service.

As of VMware vSphere 7.0 Update 1, the Xorg service is no longer required for graphics devices in NVIDIA vGPU mode.

[root@esxi:~] /etc/init.d/xorg start

8. In the **Graphics Devices** tab of the VMware vCenter Web UI, confirm that the active type and the configured type of each physical GPU are Shared Direct.

ie in the second se	Host Graphics Graphics Device	ces			
Time Configuration	Graphics Devices				
Authentication Services	Graphics Devices				
Certificate	/				Q Filter
Power Management	Name	Vendor	Active Type	Configured Type	Memory
Advanced System Settings	NVIDIATesia M60	NVIDIA Corporation	Shared Direct	Shared Direct	7.98 GB
System Resource Reservation	NVIDIATesia M60	NVIDIA Corporation	Shared Direct	Shared Direct	7.99 GB
Security Profile					
System Swap					
Host Profile					
Hardware	👭 (q. Find 🗸	•			2 items 🕒 Export 👻 🏠 Copy
Processors			—		
Memory	VMs associated with the graphics	s device "NVIDIATesla M60"			

3.5.2. Configuring a vSphere VM with NVIDIA vGPU

CAUTION: Output from the VM console in the VMware vSphere Web Client is not available for VMs that are running vGPU. Make sure that you have installed an alternate means of accessing the VM (such as VMware Horizon or a VNC server) before you configure vGPU.

VM console in vSphere Web Client will become active again once the vGPU parameters are removed from the VM's configuration.

- 1. Open the vCenter Web UI.
- 2. In the vCenter Web UI, right-click the VM and choose Edit Settings.
- 3. Click the Virtual Hardware tab.
- In the New device list, select Shared PCI Device and click Add.
 The PCI device field should be auto-populated with NVIDIA GRID vGPU.

🔂 Win7x86 - Edit Settings							(?) Þ	
Virtual Hardware VM C	Options	SDRS Rules	vAp	op Options)			
▶ 🔲 CPU	F 🔲 CPU 🚺		-	0				
► III Memory	1024	1024 • MB		MB	•			
▶ 🛄 Hard disk 1	24		•	GB	•			
▶ E SCSI controller 0	LSI Log	ic SAS						
▶ I Network adapter 1	VM Network				•	Connec	t	
▶	Datast	Datastore ISO File			•	Connec	t	
▶ Floppy drive 1	Client	Client Device			•	Connec	t	
▶ 🛄 Video card	Specif	Specify custom settings			•			
	NVIDI	DIA GRID vGPU			•			
GPU Profile	grid_m	ırid_m10-4q			•			
	grid_m10-8q grid_m10-8a grid_m10-4q grid_m10-4a grid_m10-2q				are unavailable when ent. You cannot			-
					ie Dr			
SATA controller 0					-			
Given VMCI device								
 Other Devices 	grid_m	10-2a	_		•			
The maximum number of	devices	of this type has t	been	reached.				
New device: 🛛 🕅 Shared PCI Device 🔹 Add								
Compatibility: ESXi 6.0 an	d later (\	/M version 11)					ОК	Cancel

5. From the **GPU Profile** drop-down menu, choose the type of vGPU you want to configure and click **OK**.

Note: VMware vSphere does **not** support vCS. Therefore, C-series vGPU types are not available for selection from the **GPU Profile** drop-down menu.

- 6. Ensure that VMs running vGPU have all their memory reserved:
 - a). Select Edit virtual machine settings from the vCenter Web UI.
 - b). Expand the Memory section and click Reserve all guest memory (All locked).

After you have configured a vSphere VM with a vGPU, start the VM. VM console in vSphere Web Client is not supported in this vGPU release. Therefore, use VMware Horizon or VNC to access the VM's desktop.

3.5.3. Configuring a Citrix Hypervisor VM with Virtual GPU

- 1. Ensure the VM is powered off.
- 2. Right-click the VM in XenCenter, select **Properties** to open the VM's properties, and select the **GPU** property.

XenCente - 6 % le View Bool Server VM Størage Templates Tools Help I Back - @ Forward - I 🖏 Add New Server I 🏪 New Pool 🕲 New Storage 📜 New VM I 🥹 Start ⊛ Reboot 🕕 Suspen Q Win7x64_CUDA_OPENCL (2) (1) on 'xs-72' Logged in as: Local root account General Memory St Win7x64_CUDA_OPENCL (2) (1)' Propertie ? × General Win7x64_CUDA_OPENCL (2) ... Custom Fields Vin7x64_CUDA_O Properties 9 <None> CPU 4 vCPU(s) Boot Options You can improve graphics performance by assigning a virtual graphics processing unit to this General HEL 7.3 GPU type: Pass-through whole GPL -Boot order: DVD-Drive, Hard... Start Options HA is not available on stand... Pass dmough whole GPU
 Pass dmough whole GPU
 Pass dmough whole GPU
 Pass dmough whole GPU
 GRD M60-8Q virtual GPU Lper GPU, 4056x2160, 4 display
 GRD M60-8Q virtual GPU Lper GPU, 1228x1024, 1 display
 GRD M60-4Q virtual GPU L2 per GPU, 4056x2160, 4 display
 GRD M60-4Q virtual GPU L2 per GPU, 4056x2160, 4 display
 GRD M60-4Q virtual GPU L2 per GPU, 1228x1024, 1 display
 GRD M60-4Q virtual GPU C2 per GPU, 1228x1024, 1 display
 GRD M60-4Q virtual GPU C2 per GPU, 1228x1024, 1 display
 GRD M60-2A virtual GPU C2 per GPU, 1228x1024, 1 display
 GRD M60-2A virtual GPU C2 per GPU, 1228x1024, 1 display
 GRD M60-2A virtual GPU C2 per GPU, 1228x1024, 1 display
 GRD M60-2A virtual GPU C2 per GPU, 1228x1024, 1 display
 GRD M60-2A virtual GPU C2 per GPU, 1228x1024, 1 display
 GRD M60-2A virtual GPU C2 per GPU, 1228x1024, 1 display
 GRD M60-4A virtual GPU C2 per GPU, 1228x1024, 1 display
 GRD M60-4A virtual GPU C2 per GPU, 1228x1024, 1 display
 GRD M60-4A virtual GPU C2 per GPU, 1228x1024, 1 display
 GRD M60-4A virtual GPU C2 per GPU, 1228x1024, 1 display
 GRD M60-4A virtual GPU C2 per GPU, 1228x1024, 1 display
 GRD M60-4A virtual GPU C2 per GPU, 1228x1024, 1 display
 GRD M60-4A virtual GPU C2 per GPU, 1228x1024, 1 display
 GRD M60-4A virtual GPU C2 per GPU, 1228x1024, 1 display
 GRD M60-4A virtual GPU C2 per GPU, 1228x1024, 1 display
 GRD M60-4A virtual GPU C2 per GPU, 1228x1024, 1 display
 GRD M60-4A virtual GPU C2 per GPU, 1228x1024, 1 display
 GRD M60-4A virtual GPU C2 per GPU, 1228x1024, 1 display
 GRD M60-4A virtual GPU C2 per GPU, 1228x1024, 1 display
 GRD M60-4A virtual GPU C2 per GPU, 1228x1024, 1 display
 GRD M60-4A virtual GPU C2 per GPU, 1228x1024, 1 display
 GRD M60-4A virtual GPU C2 per GPU, 1228x1024, 1 display
 GRD M60-4A virtual GPU C2 per GPU, 1228x1024, 1 display
 GRD M60-4A virtual GPU C2 per GPU, 1228x1024, 1 display
 GRD M60-4A virtual GPU C2 per GPU, 1228x1024, 1 display
 GRD M60-4A virtual G Description Tags: Alerts Folder None defined Home Server None defined GPU able to start Operating Syste Virtualization mode: Pass-through who Advanced Options BIOS strings GRID M60-1B virtual GPU (8 per GPU, 2560x1600, 4 displays) GRID M60-1A virtual GPU (8 per GPU, 1280x1024, 1 display) GRID M60-0Q virtual GPU (16 per GPU, 2560x1600, 2 displays) Virtualization st GRID M60-0B virtual GPU (16 per GPU, 2560x1600 VIDIA Corporation GP102GL [Tesla P40] GPUs 600, 2 displays UUID: Pass-through whole GPU GRID P40-24Q virtual GPU (1 per GPU, 4096x2160, 4 display Boot Options GRID P40-24Q virtual GPU (1 per GPU, 4096x2160, 4 displays GRID P40-24A virtual GPU (1 per GPU, 1280x1624, 1 displays GRID P40-12Q virtual GPU (2 per GPU, 4096x2160, 4 displays GRID P40-12A virtual GPU (2 per GPU, 1280x1024, 1 display) GRID P40-8A virtual GPU (3 per GPU, 1280x1024, 1 display) III + CPUs Infrastructure Objects OK Cance Organization Views Saved Searches Notifications

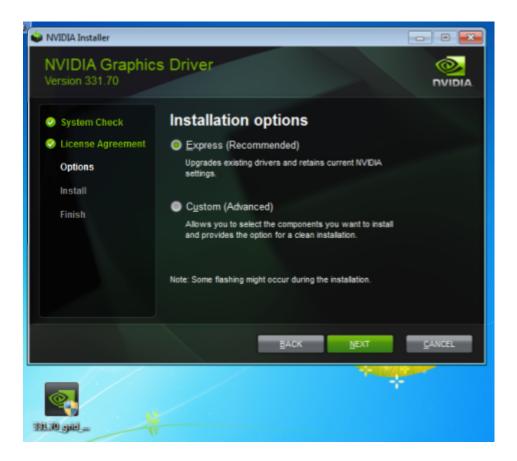
The available GPU types are listed in the GPU type drop-down list:

After you have configured a Citrix Hypervisor VM with a vGPU, start the VM, either from XenCenter or by using xe vm-start in a dom0 shell. You can view the VM's console in XenCenter.

3.6. Installing the NVIDIA vGPU Software Graphics Driver

After you create a Windows VM on the hypervisor and boot the VM, the VM should boot to a standard Windows desktop in VGA mode at 800×600 resolution. You can use the Windows screen resolution control panel to increase the resolution to other standard resolutions, but to fully enable GPU operation, the NVIDIA vGPU software graphics driver must be installed.

- 1. Copy the NVIDIA Windows driver package to the guest VM where you are installing the driver.
- 2. Execute the package to unpack and run the driver installer.

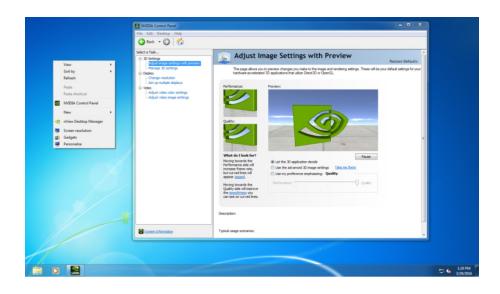


- 3. Click through the license agreement.
- 4. Select **Express Installation** and click **NEXT**. After the driver installation is complete, the installer may prompt you to restart the platform.
- 5. If prompted to restart the platform, do one of the following:
 - Select Restart Now to reboot the VM.
 - Exit the installer and reboot the VM when you are ready.

After the VM restarts, it boots to a Windows desktop.

- 6. Verify that the NVIDIA driver is running.
 - a). Right-click on the desktop.
 - b). From the menu that opens, choose NVIDIA Control Panel.
 - c). In the NVIDIA Control Panel, from the Help menu, choose System Information.

NVIDIA Control Panel reports the vGPU that is being used, its capabilities, and the NVIDIA driver version that is loaded.



3.7. Licensing an NVIDIA vGPU

Perform this task from the guest VM to which the vGPU is assigned.

The **NVIDIA Control Panel** tool that you use to perform this task detects that a vGPU is assigned to the VM and, therefore, provides no options for selecting the license type. After you license the vGPU, NVIDIA vGPU software automatically selects the correct type of license based on the vGPU type.

- Note: These instructions cover only licensing for a single-instance license server. If you configured a primary and secondary license server, refer to <u>Virtual GPU Client Licensing User</u>
 <u>Guide</u> for instructions for how to specify the address and port number of the primary and secondary license servers.
- 1. Open NVIDIA Control Panel:
 - ▶ Right-click on the Windows desktop and select **NVIDIA Control Panel** from the menu.
 - Open Windows Control Panel and double-click the NVIDIA Control Panel icon.
- 2. In **NVIDIA Control Panel**, select the **Manage License** task in the **Licensing** section of the navigation pane.

Note: If the **Licensing** section and **Manage License** task are not displayed in **NVIDIA Control Panel**, the system has been configured to hide licensing controls in **NVIDIA Control Panel**. For information about registry settings, refer to <u>Virtual GPU Client Licensing User Guide</u>.

The Manage License task pane shows that NVIDIA vGPU is currently unlicensed.

NVIDIA Control Panel		
<u>F</u> ile <u>E</u> dit Des <u>k</u> top <u>H</u> elp		
🚱 Back 👻 🚱 🐔		
Select a Task	Manage License	ń
Adjust image settings with preview Manage 3D settings Set PhysX Configuration	You can enable additional features by applying a license.	
Display Change resolution Set up multiple displays Chensing	License Edition: Your system does not have a valid GRID vGPU license. To access GRID vGPU features, enter license server details and apply.	
Anage License - Video Adjust video color settings	To access GRID vGPU teatures, enter license server details and apply. Primary License Server:	
∟Adjust video image settings	Port Number:	
		E
	Secondary License Server:	
	Port Number:	
	Description:	
	Typical usage scenarios:	
System Information	۲. III	

3. In the **Primary License Server** field, enter the address of yourNVIDIA vGPU software License Server.

The address can be a fully-qualified domain name such as gridlicensel.example.com, or an IP address such as 10.31.20.45.

- Leave the Port Number field under the Primary License Server field unset. The port defaults to 7070, which is the default port number used by NVIDIA vGPU software License Server.
- 5. Leave the Secondary License Server field and the Port Number field under it unset.
- Click Apply to assign the settings. The system requests the appropriate license for the current vGPU from the configured license server.

The vGPU within the VM should now operate at full capability without any performance degradation over time for as long as the vGPU is licensed.

You have now completed the process of installing NVIDIA vGPU software and setting up a single VM to use NVIDIA vGPU.

If the system fails to obtain a license, see <u>Virtual GPU Client Licensing User Guide</u> for guidance on troubleshooting.

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 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.

VESA DisplayPort

DisplayPort and DisplayPort Compliance Logo, DisplayPort Compliance Logo for Dual-mode Sources, and DisplayPort Compliance Logo for Active Cables are trademarks owned by the Video Electronics Standards Association in the United States and other countries.

HDMI

HDMI, the HDMI logo, and High-Definition Multimedia Interface are trademarks or registered trademarks of HDMI Licensing LLC.

OpenCL

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

Trademarks

NVIDIA, the NVIDIA logo, NVIDIA GRID, NVIDIA GRID vGPU, NVIDIA Maxwell, NVIDIA Pascal, NVIDIA Turing, NVIDIA Volta, GPUDirect, Quadro, and Tesla are trademarks 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

© 2013-2021 NVIDIA Corporation & affiliates. All rights reserved.

