Virtual GPU Software

Quick Start Guide

Table of Contents

About this Guide

Virtual GPU Software Quick Start Guide provides minimal instructions for installing and configuring NVIDIA® vGPU software on the Citrix Hypervisor or VMware vSphere hypervisor and for installing and configuring a Cloud License Service (CLS) instance or a standalone Delegated License Service (DLS) instance. The instructions for configuring a DLS instance assume that the VM that hosts the DLS instance has been assigned an IP address automatically.

If you need complete instructions for installing and configuring NVIDIA vGPU software or are using other platforms, refer to *Virtual GPU Software User Guide*.

If you are hosting a DLS instance on a VM that has not been assigned an IP address automatically, or require high availability for a DLS instance, refer to *NVIDIA License System User Guide*.

Chapter 1. Getting 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 NVIDIA vGPU software and technical support from NVIDIA.

To get NVIDIA vGPU software and technical support from NVIDIA, you must have an NVIDIA Enterprise Account, which provides login access to the following NVIDIA web properties:

- **NVIDIA Licensing Portal**, which provides access to your entitlements, software downloads, and options for managing your NVIDIA vGPU software license servers
- **NVIDIA Enterprise Support Portal**, which provides access to NVIDIA vGPU software support services

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 NVIDIA GRID Certified Servers.
- 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.
- You have a valid NVIDIA software subscription.

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. Your NVIDIA Entitlement Certificate contains your product activation keys and provides instructions for using the certificate.

If you are a data center administrator, follow the instructions in the NVIDIA Entitlement Certificate to use the certificate. Otherwise, forward your order confirmation message, including the attached NVIDIA Entitlement Certificate, to a data center administrator in your organization.

1.3. NVIDIA Enterprise Account Requirements

To get NVIDIA vGPU software, you must have a suitable NVIDIA Enterprise Account for getting NVIDIA vGPU software and technical support from NVIDIA.

> **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.

Whether or not you have a suitable NVIDIA Enterprise Account depends on whether you have previously purchased NVIDIA vGPU software.

- If you have previously purchased NVIDIA vGPU software, you already have a suitable NVIDIA Enterprise Account.

  To use this account to get NVIDIA vGPU software, follow the **Login** link in the instructions for using the certificate to log in to the NVIDIA Enterprise Application Hub, go to the NVIDIA Licensing Portal, and download your NVIDIA vGPU software. For details, refer to Downloading NVIDIA vGPU Software.
- If you have obtained an evaluation license but have not previously purchased NVIDIA vGPU software, you do **not** have a suitable NVIDIA Enterprise Account.

  To create a suitable NVIDIA Enterprise Account, 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 Creating your NVIDIA Enterprise Account, 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 Linking an Evaluation Account to an

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

- If you have not previously purchased NVIDIA vGPU software, you do **not** have a suitable NVIDIA Enterprise Account.

  To create a suitable NVIDIA Enterprise Account, follow the **Register** link in the instructions for using the certificate to create your account. For details, refer to Creating your NVIDIA Enterprise Account.

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 web properties for getting NVIDIA vGPU software and technical support from NVIDIA.

For details of these web properties, refer to Getting NVIDIA vGPU Software.

If you already have an account, skip this task and go to Downloading NVIDIA vGPU Software.

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 Linking an Evaluation Account to an NVIDIA Enterprise Account for Purchased Licenses, 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**.
   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**.

   > **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**.

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

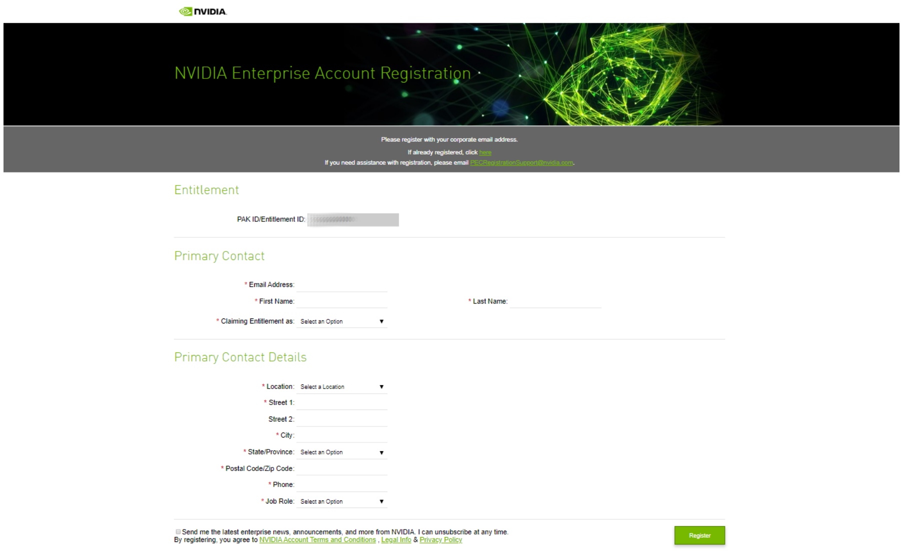
   You are then automatically directed to log in to the NVIDIA Licensing Portal with your new password.

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 Creating your NVIDIA Enterprise Account, 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 e-mail address with which you created your existing account, and click **Register**.



NVIDIA Enterprise Account Registration

Please register with your corporate email address.
If already registered, click here
If you need assistance with registration, please email PECRegistrationSupport@nvidia.com.

Entitlement

PAK ID/Entitlement ID: [redacted]

Primary Contact

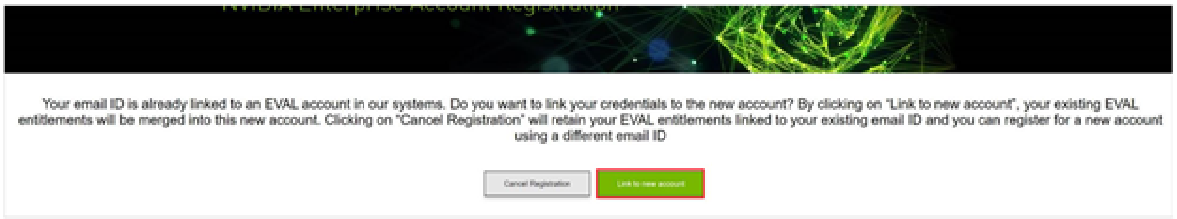
* Email Address:
* First Name:
* Last Name:
* Claiming Entitlement as: Select an Option ▾

Primary Contact Details

* Location: Select a Location ▾
* Street 1:
Street 2:
* City:
* State/Province: Select an Option ▾
* Postal Code/Zip Code:
* Phone:
* Job Role: Select an Option ▾

□ Send me the latest enterprise news, announcements, and more from NVIDIA. I can unsubscribe at any time.
By registering, you agree to NVIDIA Account Terms and Conditions , Legal Info & Privacy Policy

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



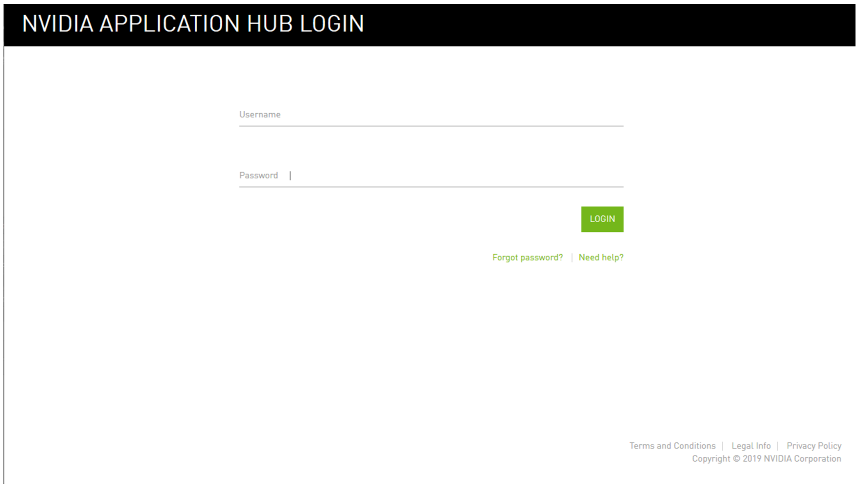
Your email ID is already linked to an EVAL account in our systems. Do you want to link your credentials to the new account? By clicking on "Link to new account", your existing EVAL entitlements will be merged into this new account. Clicking on "Cancel Registration" will retain your EVAL entitlements linked to your existing email ID and you can register for a new account using a different email ID

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

1.6. Downloading NVIDIA vGPU Software

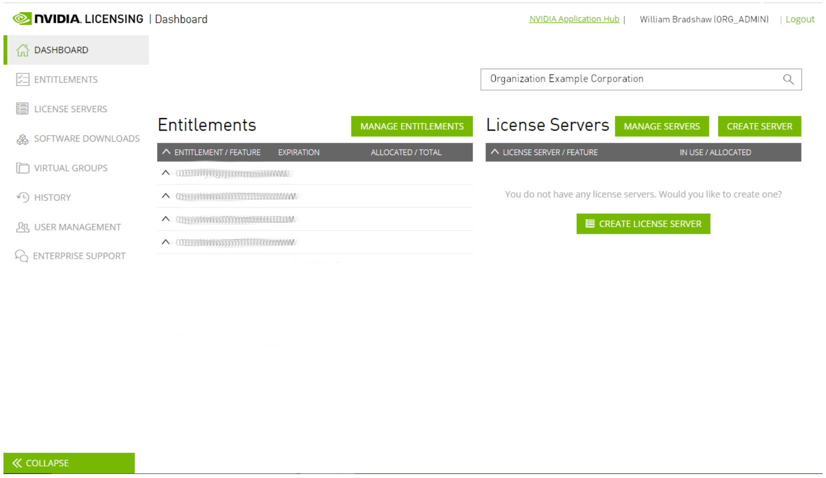
Before you begin, ensure that you have your order confirmation message and have created an NVIDIA Enterprise Account.

1. Visit the NVIDIA Enterprise Application Hub 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 APPLICATION HUB LOGIN
Username
Password
LOGIN
Forgot password? | Need help?
Terms and Conditions | Legal Info | Privacy Policy
Copyright © 2019 NVIDIA Corporation

3. On the **NVIDIA APPLICATION HUB** page that opens, click **NVIDIA LICENSING PORTAL**.

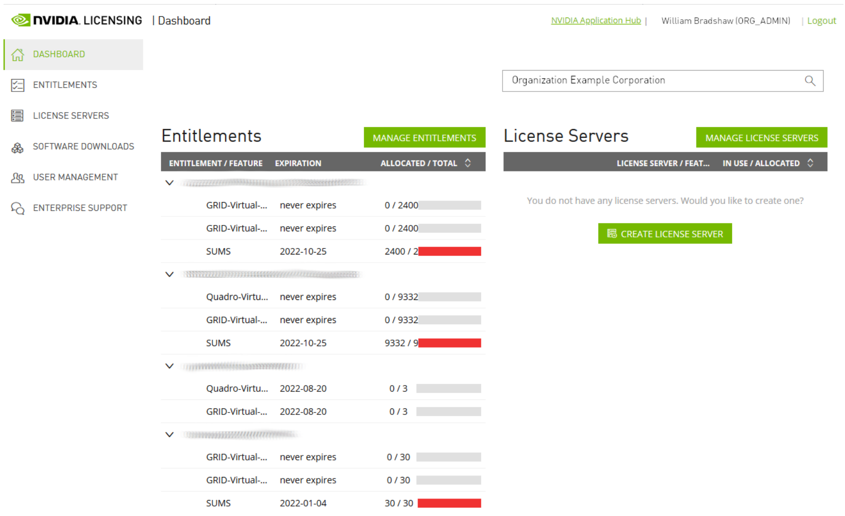
   The NVIDIA Licensing Portal dashboard page opens.



You do not have any license servers. Would you like to create one?

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.



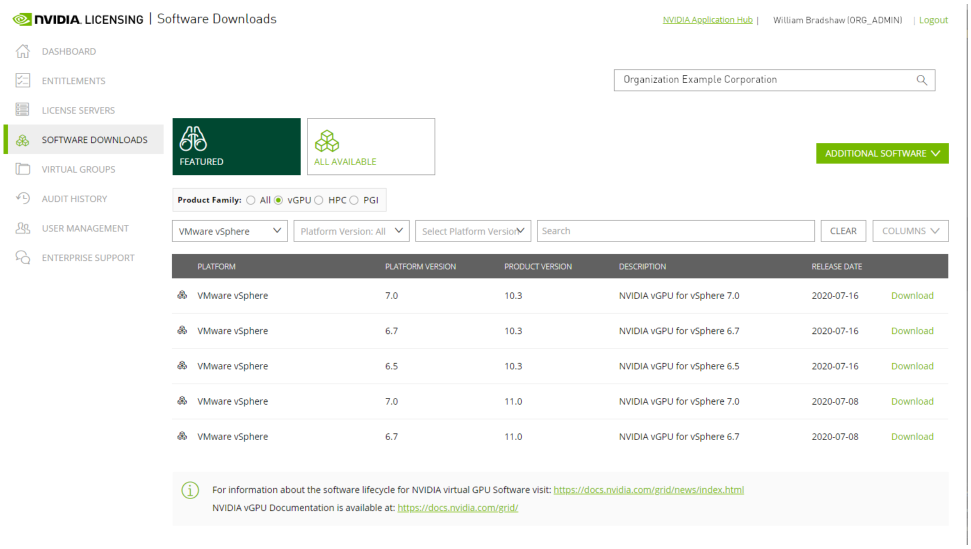
ENTITLEMENT / FEATURE	EXPIRATION	ALLOCATED / TOTAL
[redacted]		
GRID-Virtual-...	never expires	0 / 2400
GRID-Virtual-...	never expires	0 / 2400
SUMS	2022-10-25	2400 / 2
[redacted]		
Quadro-Virtu...	never expires	0 / 9332
GRID-Virtual-...	never expires	0 / 9332
SUMS	2022-10-25	9332 / 9
[redacted]		
Quadro-Virtu...	2022-08-20	0 / 3
GRID-Virtual-...	2022-08-20	0 / 3
[redacted]		
GRID-Virtual-...	never expires	0 / 30
GRID-Virtual-...	never expires	0 / 30
SUMS	2022-01-04	30 / 30

LICENSE SERVER / FEAT...	IN USE / ALLOCATED

You do not have any license servers. Would you like to create one?

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 7.0.2 for NVIDIA vGPU software release 15.4.

> **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.



Product Family: ○ All ● vGPU ○ HPC ○ PGI

VMware vSphere | Platform Version: All | Select Platform Version | Search

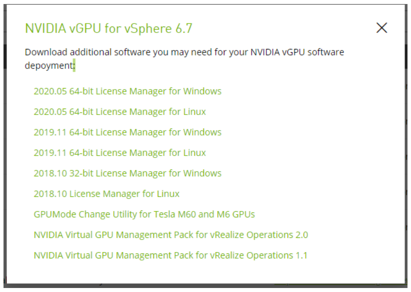
PLATFORM	PLATFORM VERSION	PRODUCT VERSION	DESCRIPTION	RELEASE DATE	
VMware vSphere	7.0	10.3	NVIDIA vGPU for vSphere 7.0	2020-07-16	Download
VMware vSphere	6.7	10.3	NVIDIA vGPU for vSphere 6.7	2020-07-16	Download
VMware vSphere	6.5	10.3	NVIDIA vGPU for vSphere 6.5	2020-07-16	Download
VMware vSphere	7.0	11.0	NVIDIA vGPU for vSphere 7.0	2020-07-08	Download
VMware vSphere	6.7	11.0	NVIDIA vGPU for vSphere 6.7	2020-07-08	Download

For information about the software lifecycle for NVIDIA virtual GPU Software visit: https://docs.nvidia.com/grid/news/index.html
NVIDIA vGPU Documentation is available at: https://docs.nvidia.com/grid/

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.



NVIDIA vGPU for vSphere 6.7

Download additional software you may need for your NVIDIA vGPU software depoyment:

- 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). If you are using Delegated License Service (DLS) instances to serve licenses, follow the link to DLS 1.0 for your chosen hypervisor, for example, **DLS 1.0 for VMware vSphere**.

   For information about installing and configuring DLS instances, refer to *NVIDIA License System User Guide*.

   b). If you are using NVIDIA GPU Operator, follow the **GPU Operator vGPU Driver Catalogs** link.

   c). 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**.

   d). 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 Switching the Mode of a Tesla M60 or M6 GPU.

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

The NVIDIA License System is used to serve a pool of floating licenses to licensed NVIDIA software products. The NVIDIA License System is configured with licenses obtained from the NVIDIA Licensing Portal.

> **Note:** These instructions cover only the configuration of a Cloud License Service (CLS) instance or a standalone Delegated License Service (DLS) instance. The instructions for configuring a DLS instance assume that the VM that hosts the DLS instance has been assigned an IP address automatically. If you need complete instructions, are hosting a DLS instance on a VM that has not been assigned an IP address automatically, or require high availability for a DLS instance, refer to *NVIDIA License System User Guide*.

2.1. Introduction to NVIDIA Software Licensing

To activate licensed functionalities, a licensed client must obtain a software license when it is booted.

A client with a network connection obtains a license by leasing it from a NVIDIA License System service instance. The service instance serves the license to the client over the network from a pool of floating licenses obtained from the NVIDIA Licensing Portal. The license is returned to the service instance when the licensed client no longer requires the license.

2.2. Configuring a CLS Instance

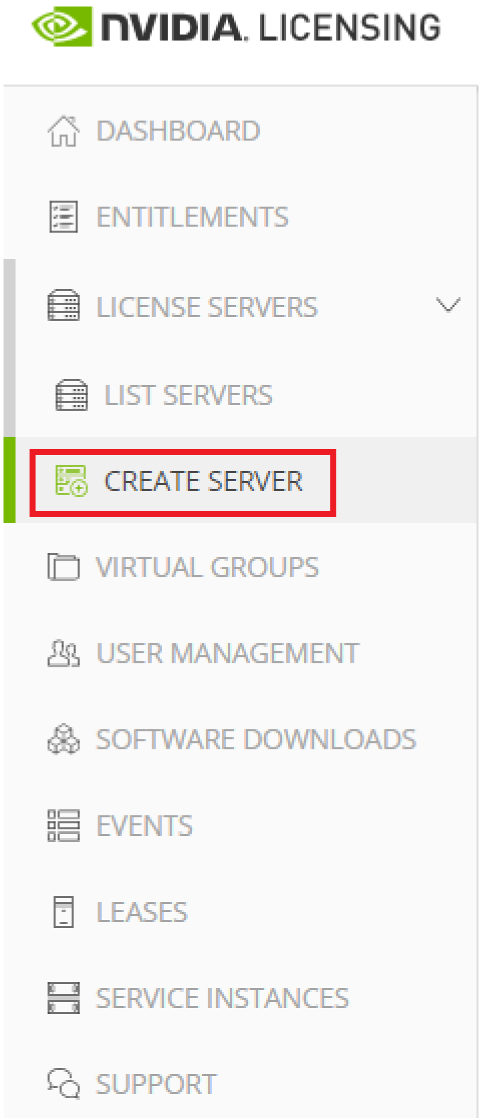
A Cloud License Service (CLS) instance is hosted on the NVIDIA Licensing Portal.

2.2.1. Creating a License Server on the NVIDIA Licensing Portal

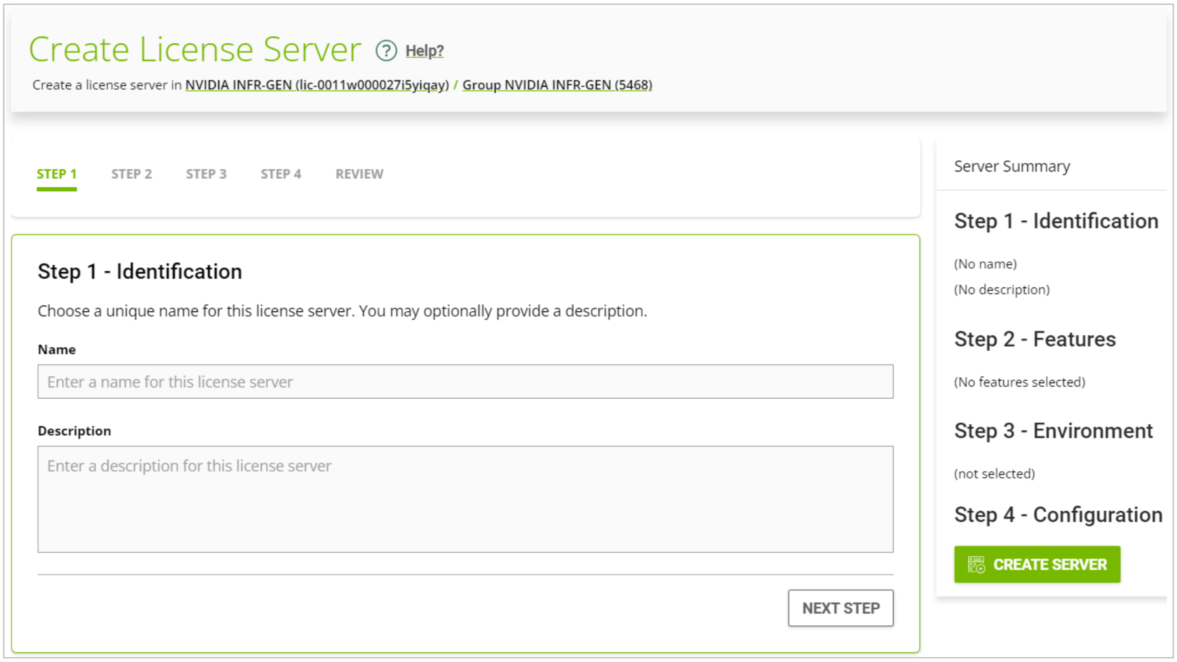
To be able to allot licenses to an NVIDIA License System instance, you must create at least one license server on the NVIDIA Licensing Portal. Creating a license server defines the set of licenses to be allotted.

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 NVIDIA Enterprise Application Hub 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, click **View settings** at the top right of the page and in the **My Info** window that opens, select the virtual group from the **Virtual Group** drop-down list, and close the **My Info** window.

    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.
2. In the left navigation pane of the NVIDIA Licensing Portal dashboard, expand **LICENSE SERVER** and click **CREATE SERVER**. The Create License Server wizard is started.



The **Create License Server** wizard opens.



Create License Server

Create a license server in NVIDIA INFR-GEN (lic-0011w000027i5yiqay) / Group NVIDIA INFR-GEN (5468)

Step 1 - Identification

Choose a unique name for this license server. You may optionally provide a description.

Name

Enter a name for this license server

Description

Enter a description for this license server

Server Summary

Step 1 - Identification

(No name)

(No description)

Step 2 - Features

(No features selected)

Step 3 - Environment

(not selected)

Step 4 - Configuration

3. On the Create License Server page of the wizard, step through the configuration requirements to provide the details of your license server.
   a). **Step 1 – Identification**: In the **Name** field, enter your choice of name for the license server and in the **Description** field, enter a text description of the license server.

      The description is required and will be displayed on the details page for the license server that you are creating.
   b). **Step 2 – Features**: Select one or more available features from your entitlements to allot to this license server.
   c). **Step 3 - Environment**: Select **Cloud (CLS)** or **On-Premises (DLS)** to install this license server.

      To make the selection after the license server has been created, select the **Deferred** option.
   d). **Step 4 – Configuration**: From the **Leasing mode** drop-down list, select one of the following leasing modes:

      Standard Networked Licensing
         Select this mode to simplify the management of licenses on a license server that supports networked licensing. In this mode, no additional configuration of the licenses on the server is required.

      Advanced Networked Licensing
         Select this mode if you require control over the management of licenses on a license server that supports networked licensing. This mode requires additional configuration to create license pools and fulfillment conditions on the server.

      Node-Locked Licensing
         Select this mode **only** if the license server will serve clients that cannot obtain a license from a remote license server over a network connection. In this mode,

the clients obtain a node-locked license from a file installed locally on the client system.

> ! **CAUTION:** This mode requires additional work to create the license file to be installed locally and to return licenses when the client is shut down. If this mode is set, the mode of the license server **cannot** be changed.

e). Click **REVIEW SUMMARY** to review the configuration summary before creating the license server.

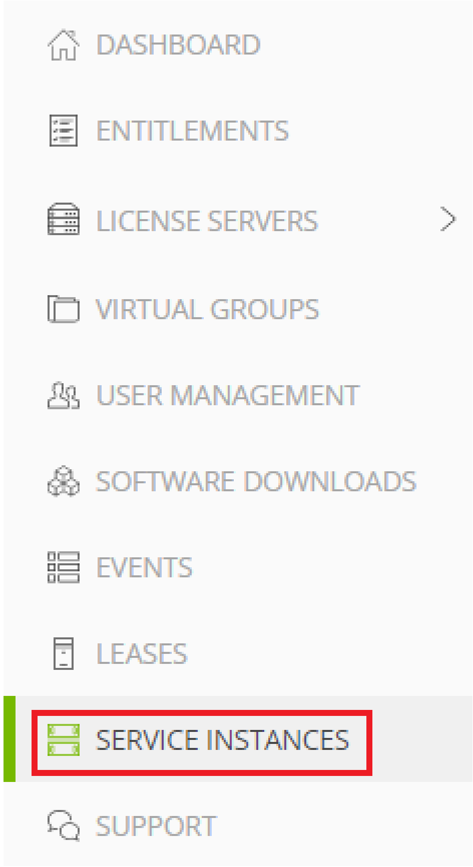
4. On the Create License Server page, from the **Step 4 – Configuration** menu, click the **CREATE SERVER** option to create this license server.

   Alternatively, you can click **CREATE SERVER** on the Server Summary page.

2.2.2. Creating a CLS Instance on the NVIDIA Licensing Portal

When you create a CLS instance, the instance is automatically registered with the NVIDIA Licensing Portal. This task is only necessary if you are not using the default CLS instance.

1. If you are not already logged in, log in to the NVIDIA Enterprise Application Hub and click **NVIDIA LICENSING PORTAL** to go to the NVIDIA Licensing Portal.
2. In the left navigation pane of the NVIDIA Licensing Portal dashboard, click **SERVICE INSTANCES**.



3. On the Service Instances page, from the **Actions** menu, choose **Create cloud (CLS) instance**.

   The **Create cloud (CLS) instance** pop-up window opens.
4. Provide the details of your cloud service instance.
   a). In the **Name** field, enter your choice of name for the service instance.
   b). In the **Description** field, enter a text description of the service instance.
   This description is required and will be displayed on the **Service Instances** page when the entry for service instance that you are creating is expanding.
5. Click **CREATE CLS INSTANCE**.

2.2.3. Binding a License Server to a Service Instance

Binding a license server to a service instance ensures that licenses on the server are available only from that service instance. As a result, the licenses are available only to the licensed clients that are served by the service instance to which the license server is bound.

1. In the NVIDIA Licensing Portal, navigate to the organization or virtual group to which the **license server** belongs.

   a). If you are not already logged in, log in to the NVIDIA Enterprise Application Hub 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, click **View settings** at the top right of the page and in the **My Info** window that opens, select the virtual group from the **Virtual Group** drop-down list, and close the **My Info** window.

2. In the left navigation pane of the NVIDIA Licensing Portal dashboard, expand **LICENSE SERVERS** and click **LIST SERVERS**.
3. In the list of license servers on the **License Servers** page that opens, from the **Actions** menu for the license server, choose **Bind**.
4. In the **Bind Service Instance** pop-up window that opens, select the service instance to which you want to bind the license server and click **BIND**.
   The **Bind Service Instance** pop-up window confirms that the license server has been bound to the service instance.

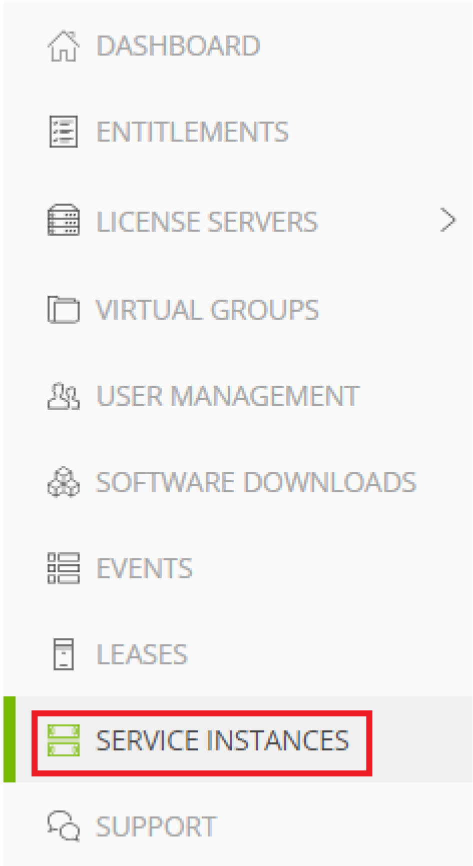
2.2.4. Installing a License Server on a CLS Instance

This task is necessary only if you are not using the default CLS instance.

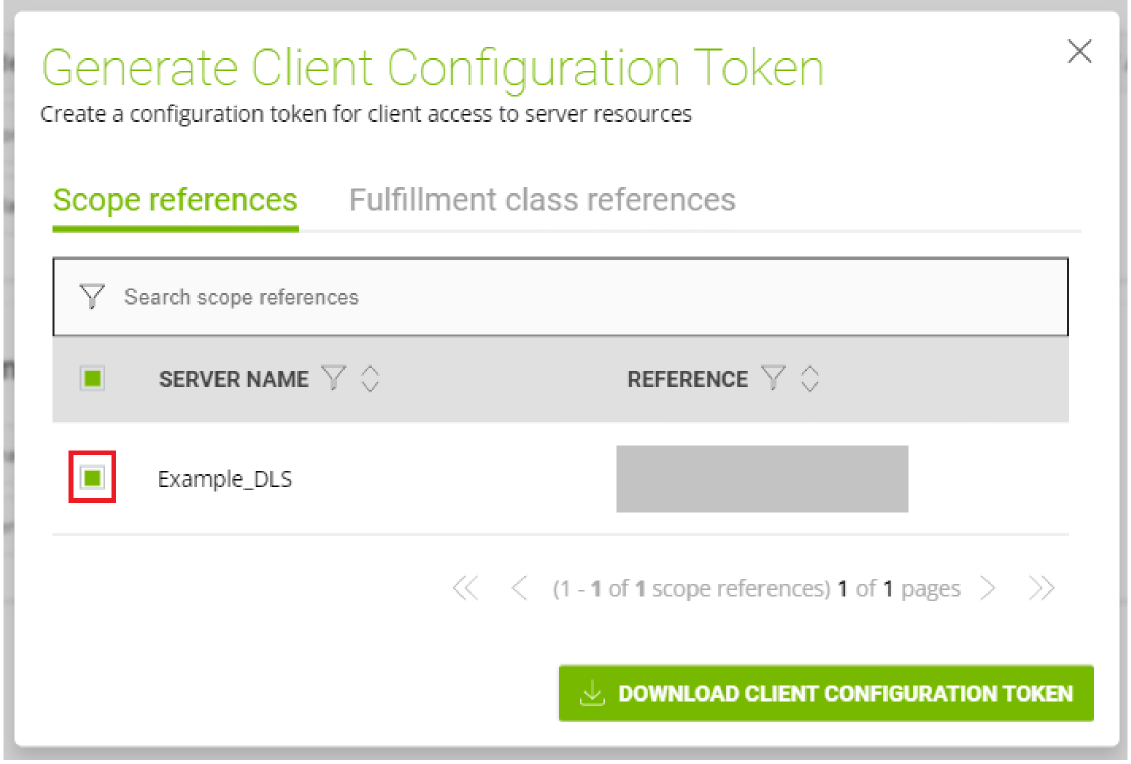
1. In the NVIDIA Licensing Portal, navigate to the organization or virtual group for which you want to install the license server.
   a). If you are not already logged in, log in to the NVIDIA Enterprise Application Hub and click **NVIDIA LICENSING PORTAL** to go to the NVIDIA Licensing Portal.
   b). **Optional:** If your assigned roles give you access to multiple virtual groups, click **View settings** at the top right of the page and in the My Info window that opens, select the virtual group from the **Virtual Group** drop-down list, and close the **My Info** window.
2. In the left navigation pane of the NVIDIA Licensing Portal dashboard, expand **LICENSE SERVER** and click **LIST SERVERS**.
3. In the list of license servers on the **License Servers** page that opens, click the name of the license server that you want to install.
4. In the **License Server Details** page that opens, from the **Actions** menu, choose **Install**.
5. In the **Install License Server** pop-up window that opens, click **INSTALL SERVER**.

2.2.5. Generating a Client Configuration Token for a CLS Instance

1. Log in to the NVIDIA Enterprise Application Hub and click **NVIDIA LICENSING PORTAL** to go to the NVIDIA Licensing Portal.
2. If your assigned roles give you access to multiple virtual groups, select the virtual group for which you are managing licenses from the list of virtual groups at the top right of the NVIDIA Licensing Portal dashboard.
3. In the left navigation pane, click **SERVICE INSTANCES**.



4. On the Service Instances page that opens, from the **Actions** menu for the CLS instance for which you want to generate a client configuration token, choose **Generate client configuration token**.
5. In the **Generate Client Configuration Token** pop-up window that opens, select the references that you want to include in the client configuration token.
   a). From the list of scope references, select the scope references that you want to include.


Generate Client Configuration Token
Scope references
Fulfillment class references
Search scope references
(1 - 1 of 1 scope references) 1 of 1 pages
DOWNLOAD CLIENT CONFIGURATION TOKEN

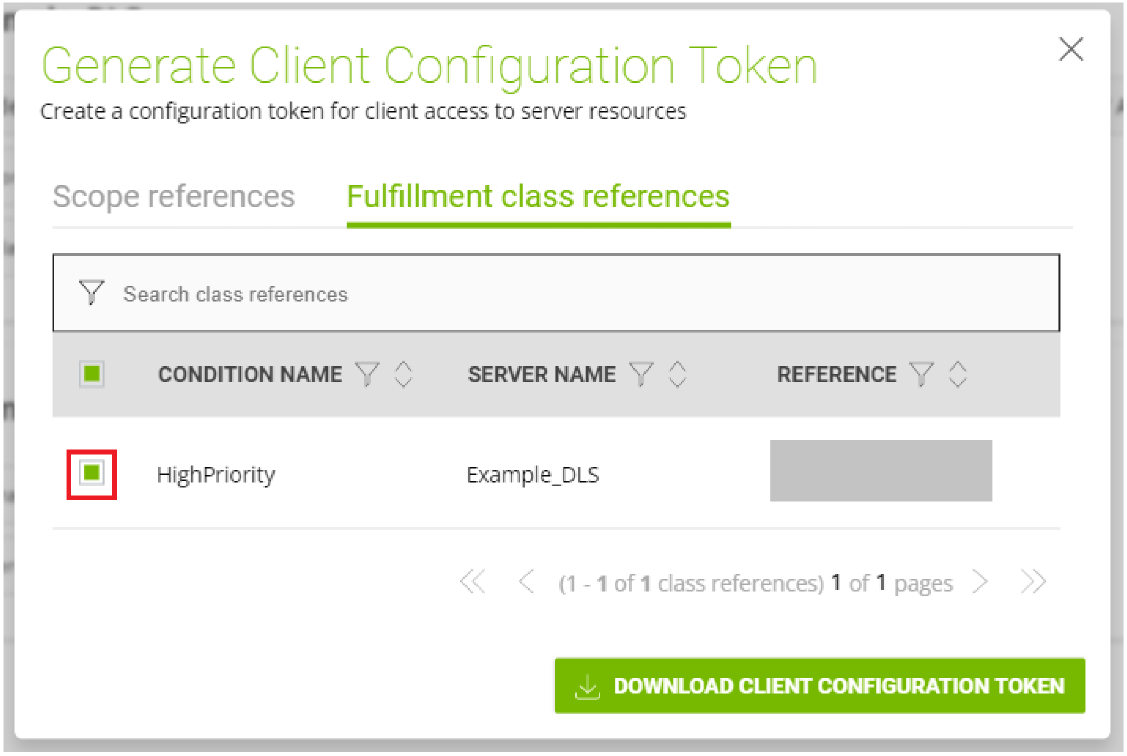
Create a configuration token for client access to server resources

☑	SERVER NAME	REFERENCE
☑	Example_DLS	[redacted]

You must select **at least one** scope reference.

Each scope reference specifies the license server that will fulfil a license request.

b). **Optional:** Click the **Fulfillment class references** tab, and from the list of fulfillment class references, select the fulfillment class references that you want to include.



Generate Client Configuration Token

Create a configuration token for client access to server resources

Scope references
Fulfillment class references
Search class references

■	CONDITION NAME	SERVER NAME	REFERENCE
■	HighPriority	Example_DLS	[redacted]

(1 - 1 of 1 class references) 1 of 1 pages
DOWNLOAD CLIENT CONFIGURATION TOKEN

Including fulfillment class references is optional.

c). **Optional:** In the **Expiration** section, select an expiration date for the client configuration token. If you do not select a date, the default token expiration time is 12 years.

d). Click **DOWNLOAD CLIENT CONFIGURATION TOKEN**.

A file named `client_configuration_token_`*`mm-dd-yyyy-hh-mm-ss`*`.tok` is saved to your default downloads folder.

After creating a client configuration token from a service instance, copy the client configuration token to each licensed client that you want to use the combination of license servers and fulfillment conditions specified in the token. For more information, see Configuring a Licensed Client.

2.3. Configuring a DLS Instance

A Delegated License Service (DLS) instance is hosted on-premises at a location that is accessible from your private network, such as inside your data center.

Before configuring a DLS instance, ensure that the DLS appliance is installed in a suitable VM or deployed in a suitable container on a bare-metal OS as explained in Installing and Configuring the DLS Virtual Appliance in *NVIDIA License System User Guide*.

2.3.1. Registering the DLS Administrator User

Each DLS virtual appliance is configured with a user account specifically for administering the DLS. This account provides access through a web-based management interface to the **NVIDIA Licensing** application on the appliance. Before administering a DLS virtual appliance, you must register this user to be able to access this management interface.

1. Open a web browser and connect to the URL `https://`*`dls-vm-ip-address`*.

   dls-vm-ip-address

   The IP address or, if defined, the fully qualified domain name or the CNAME of the VM on which the DLS virtual appliance is installed.

   You can get the IP address from the management console of your hypervisor.
2. On the **Set Up** page that opens, click **NEW INSTALLATION**.
3. On the **Register User** page that opens, provide the credentials for the DLS administrator user.

   > **Note:** If the DLS administrator user has already been registered, the login page opens instead of the **Register User** page.

   a). **Optional:** If you want to change the user name from the preset name **`dls_admin`**, replace the text in the **Username** field with your choice of user name.

   b). Provide a password for the DLS administrator user and confirm the password.

   The password must be at least eight characters long and is case sensitive.

   > **Note:** You can change the DLS administrator user name and password at any time after the DLS administrator user is registered.
4. Determine whether you want to enable an additional user that will be able to access the log files for the DLS virtual appliance.
   - If you **want** to enable this additional user, ensure that the **Create a diagnostic user** option remains selected.
   - Otherwise, deselect the **Create a diagnostic user** option.
5. Click **REGISTER**.
   The **Register User** page is refreshed to confirm that the user has been registered and displays a local reset secret to enable you to reset the user's password.
6. Copy the local reset secret and store it securely, for example, by clicking the clipboard icon and pasting the local reset secret into a plain text file that is readable only by you.

   You will need this key to reset the DLS administrator user's password.
7. Click **CONTINUE TO LOGIN**.
8. On the login page that opens, type the user name of the DLS administrator user, provide the password that you set for this user, and click **LOGIN**.

2.3.2. Configuring a Standalone DLS Instance

A standalone DLS instance must be registered before it can be used.

Ensure that the following prerequisites are met:

- The DLS virtual appliance that will host the instance has been installed and started.
- The DLS administrator user has been registered on the virtual appliance that will host the DLS instance.
- The DLS instance has **not** been configured as a member of a highly available (HA) cluster of DLS instances.

1. Log in to the DLS virtual appliance that will host the DLS instance.
2. In the left navigation pane, click **SERVICE INSTANCE**.
3. On the **Service Instance** page that opens, under **Node Configuration**, ensure that the **Enable High Availability** option is **not** set.
4. Click **CREATE STANDALONE** to start the configuration and wait for it to complete. The **Service Instance** page displays the progress of the standalone DLS instance configuration.

When the configuration is complete, the **Service Instance** page is updated to show the node health of the standalone DLS instance.

2.3.3. Changing the Name and Description of a DLS Instance

By default, a DLS instance is created with the name `DEFAULT_`*`timestamp`* and the description `ON_PREM_SERVICE_INSTANCE`. To distinguish a DLS instance on the NVIDIA Licensing Portal when multiple DLS instances are configured, change these defaults to a meaningful name and the description.
Perform this task from the DLS virtual appliance.

1. Log in to the DLS virtual appliance that is hosting the instance whose name and description you want to change.
2. In the left navigation pane of the **NVIDIA Licensing** dashboard, click **SERVICE INSTANCE**.
3. On the **Service Instance** page that opens, click **EDIT**.
4. In the **Edit Service Instance** dialog box that opens, type your choice of name and description for the instance and click **UPDATE**.

   Note: The instance name cannot contain special characters.

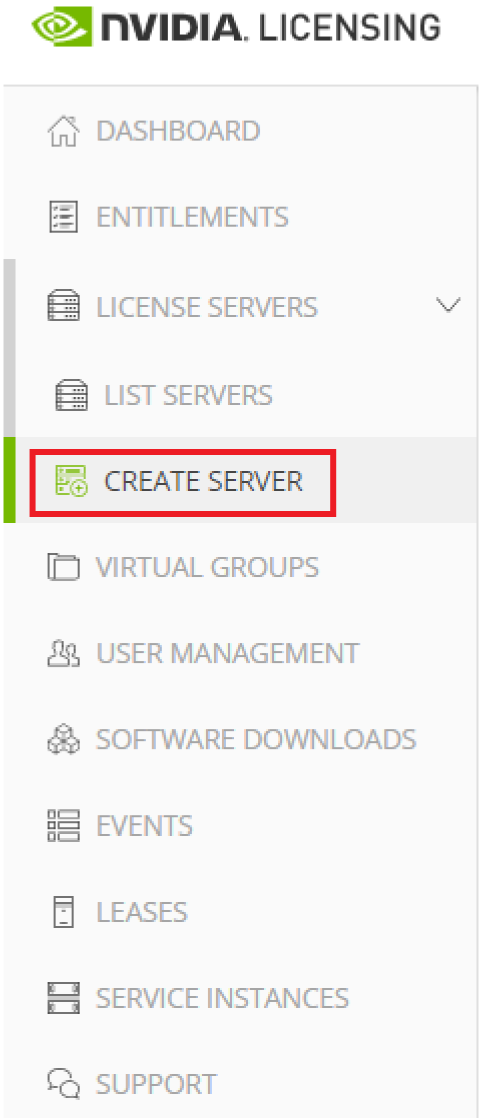
   The name and description of the instance are updated on the **Service Instance** page.

2.3.4. Creating a License Server on the NVIDIA Licensing Portal

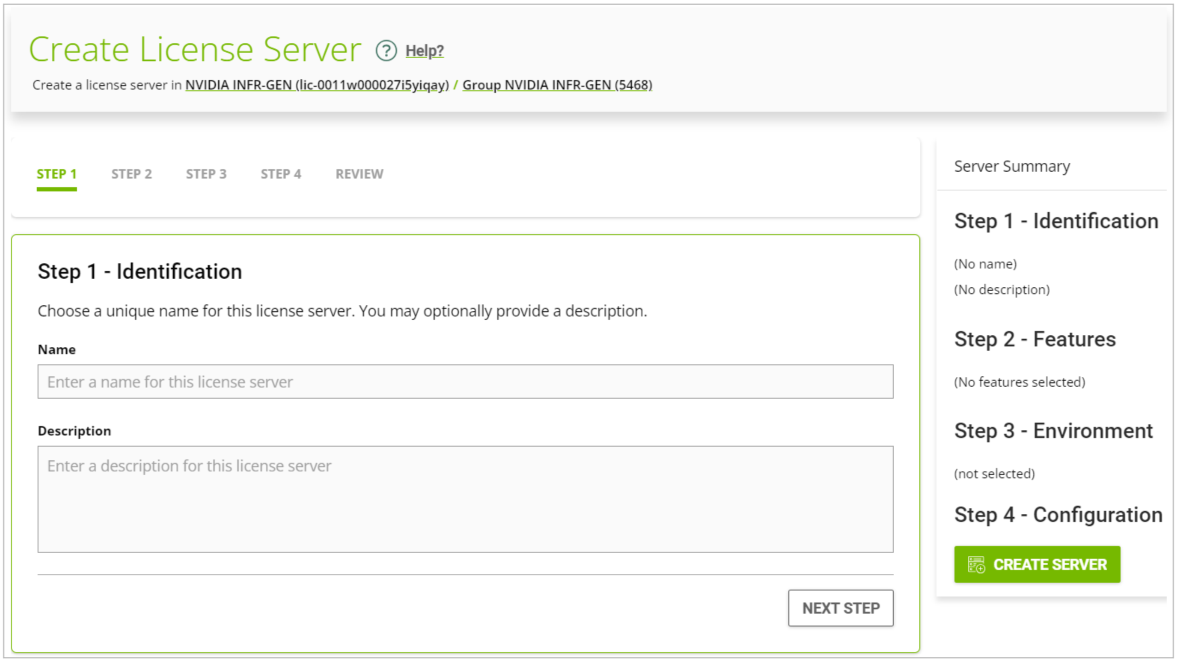
To be able to allot licenses to an NVIDIA License System instance, you must create at least one license server on the NVIDIA Licensing Portal. Creating a license server defines the set of licenses to be allotted.

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 NVIDIA Enterprise Application Hub 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, click **View settings** at the top right of the page and in the **My Info** window that opens, select the virtual group from the **Virtual Group** drop-down list, and close the **My Info** window.

   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.
2. In the left navigation pane of the NVIDIA Licensing Portal dashboard, expand **LICENSE SERVER** and click **CREATE SERVER**. The Create License Server wizard is started.



The **Create License Server** wizard opens.



Create License Server ? Help?

Create a license server in NVIDIA INFR-GEN (lic-0011w000027i5yiqay) / Group NVIDIA INFR-GEN (5468)

Step 1 - Identification

Choose a unique name for this license server. You may optionally provide a description.

Name

Enter a name for this license server

Description

Enter a description for this license server

Server Summary

Step 1 - Identification

(No name)

(No description)

Step 2 - Features

(No features selected)

Step 3 - Environment

(not selected)

Step 4 - Configuration

3. On the Create License Server page of the wizard, step through the configuration requirements to provide the details of your license server.

    a). **Step 1 – Identification**: In the **Name** field, enter your choice of name for the license server and in the **Description** field, enter a text description of the license server.

    The description is required and will be displayed on the details page for the license server that you are creating.

    b). **Step 2 – Features**: Select one or more available features from your entitlements to allot to this license server.

    c). **Step 3 - Environment**: Select **Cloud (CLS)** or **On-Premises (DLS)** to install this license server.

    To make the selection after the license server has been created, select the **Deferred** option.

    d). **Step 4 – Configuration**: From the **Leasing mode** drop-down list, select one of the following leasing modes:

    Standard Networked Licensing

    Select this mode to simplify the management of licenses on a license server that supports networked licensing. In this mode, no additional configuration of the licenses on the server is required.

    Advanced Networked Licensing

    Select this mode if you require control over the management of licenses on a license server that supports networked licensing. This mode requires additional configuration to create license pools and fulfillment conditions on the server.

    Node-Locked Licensing

    Select this mode **only** if the license server will serve clients that cannot obtain a license from a remote license server over a network connection. In this mode, the clients obtain a node-locked license from a file installed locally on the client system.

> ! **CAUTION:** This mode requires additional work to create the license file to be installed locally and to return licenses when the client is shut down. If this mode is set, the mode of the license server **cannot** be changed.

e). Click **REVIEW SUMMARY** to review the configuration summary before creating the license server.

4. On the Create License Server page, from the **Step 4 – Configuration** menu, click the **CREATE SERVER** option to create this license server.

   Alternatively, you can click **CREATE SERVER** on the Server Summary page.

2.3.5. Registering an on-Premises DLS Instance with the NVIDIA Licensing Portal

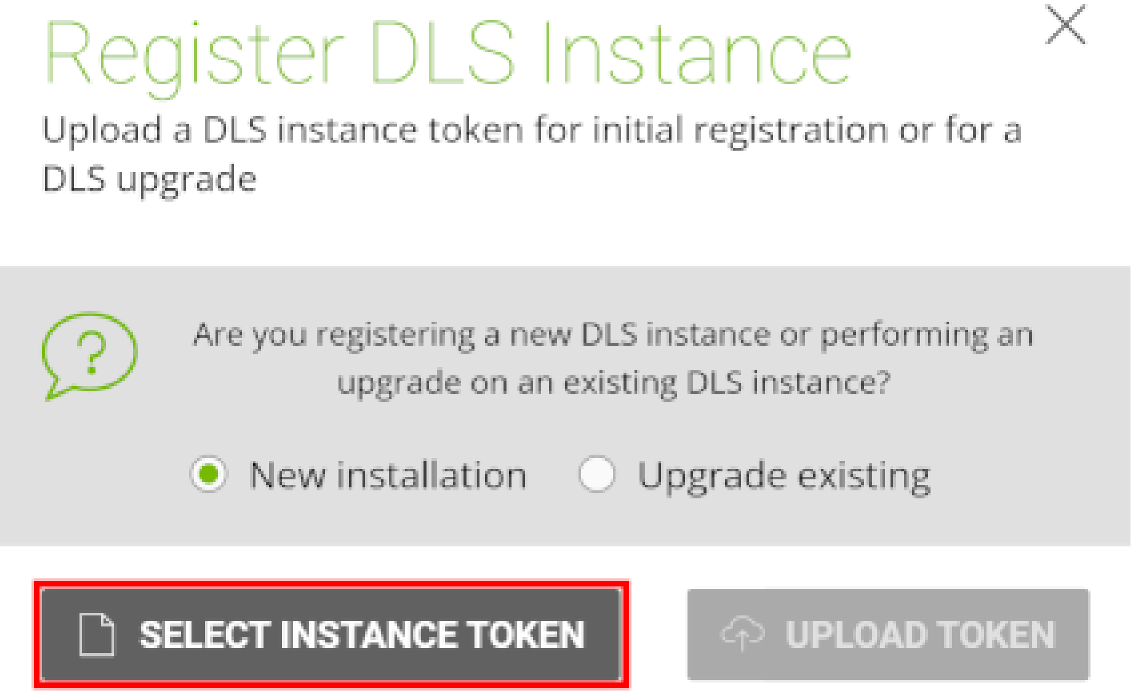
A DLS instance is created automatically when the virtual appliance on which the instance resides is installed. However, to enable the instance to be bound to a license server, you must register the instance with the NVIDIA Licensing Portal.

Registering an on-premises DLS instance with the NVIDIA Licensing Portal involves the exchange of a **DLS instance token** between the instance and the NVIDIA Licensing Portal.

A DLS instance token is created by a DLS instance. It identifies the DLS instance to the NVIDIA Licensing Portal and enables it to locate the NVIDIA Licensing Portal. After downloading the token from the DLS instance, you must upload it to the NVIDIA Licensing Portal to complete the registration of the service instance.

1. If you are not already logged in, log in to the **NVIDIA Licensing** application at the IP address of the VM on which the DLS virtual appliance is installed.
2. In the left navigation pane of the **NVIDIA Licensing** dashboard, click **SERVICE INSTANCE**.
3. On the **Service Instance Details** page that opens, from the **ACTIONS** menu, choose **Download DLS Instance Token**.
   A DLS instance token file that is named
   `dls_instance_token_`*`mm-dd-yyyy-hh-mm-ss`*`.tok` is downloaded.
4. In the NVIDIA Licensing Portal, navigate to the organization or virtual group for which you are registering the service instance.
   a). If you are not already logged in, log in to the NVIDIA Enterprise Application Hub 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, click **View settings** at the top right of the page and in the **My Info** window that opens, select the virtual group from the **Virtual Group** drop-down list, and close the **My Info** window.
5. On **Service Instances** page that opens, from the **Actions** menu, choose **Register DLS Instance**.

6. In the **Register DLS Instance** window that opens, select the **New installation** option and click **SELECT INSTANCE TOKEN**.



Register DLS Instance

Upload a DLS instance token for initial registration or for a DLS upgrade

Are you registering a new DLS instance or performing an upgrade on an existing DLS instance?

● New installation ○ Upgrade existing

7. In the file browser that opens, navigate to the folder that contains the DLS instance token file that is named `dls_instance_token_mm-dd-yyyy-hh-mm-ss.tok` that you downloaded and select the file.
8. Back in the **Register DLS Instance** window, click **UPLOAD TOKEN**.
   The service instance is added to the list of registered service instances.

2.3.6. Binding a License Server to a Service Instance

Binding a license server to a service instance ensures that licenses on the server are available only from that service instance. As a result, the licenses are available only to the licensed clients that are served by the service instance to which the license server is bound.

1. In the NVIDIA Licensing Portal, navigate to the organization or virtual group to which the **license server** belongs.
   a). If you are not already logged in, log in to the NVIDIA Enterprise Application Hub 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, click **View settings** at the top right of the page and in the **My Info** window that opens, select the virtual group from the **Virtual Group** drop-down list, and close the **My Info** window.

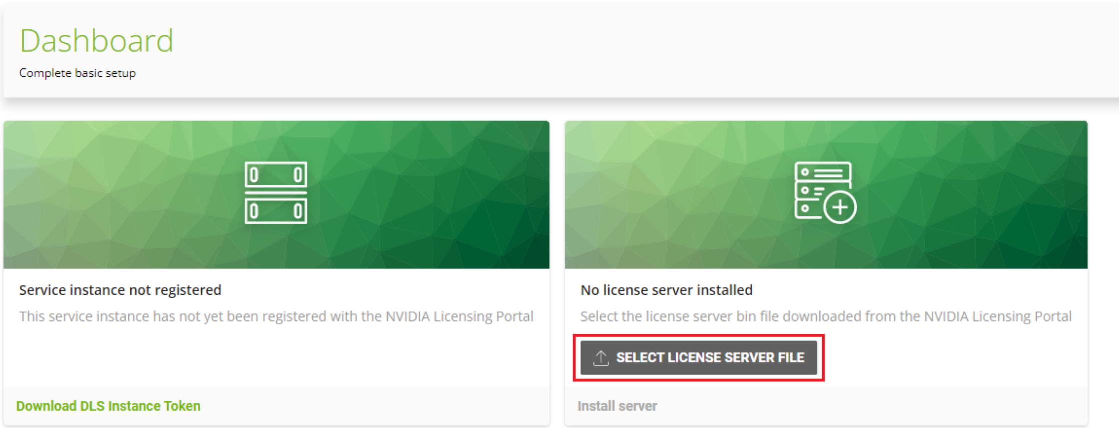
2. In the left navigation pane of the NVIDIA Licensing Portal dashboard, expand **LICENSE SERVERS** and click **LIST SERVERS**.
3. In the list of license servers on the **License Servers** page that opens, from the **Actions** menu for the license server, choose **Bind**.
4. In the **Bind Service Instance** pop-up window that opens, select the service instance to which you want to bind the license server and click **BIND**.
   The **Bind Service Instance** pop-up window confirms that the license server has been bound to the service instance.

2.3.7. Installing a License Server on a DLS Instance

1. In the NVIDIA Licensing Portal, navigate to the organization or virtual group for which the license server was created.
   a). If you are not already logged in, log in to the NVIDIA Enterprise Application Hub 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, click **View settings** at the top right of the page and in the My Info window that opens, select the virtual group from the **Virtual Group** drop-down list, and close the **My Info** window.
2. In the left navigation pane of the NVIDIA Licensing Portal dashboard, expand **LICENSE SERVER** and click **LIST SERVERS**.
3. In the list of license servers on the License Servers page that opens, click the name of the license server that you want to install.
4. In the **License Server Details** page that opens, from the **Actions** menu, choose **Download**.
5. In the **Download License File** window that opens, click **Download**.
   A license server file that is named `license_`*`mm-dd-yyyy-hh-mm-ss`*`.bin` is downloaded.
6. If you are not already logged in, log in to the **NVIDIA Licensing** application at the IP address of the VM on which the DLS virtual appliance is installed.
   After you log in, the information that is displayed on the **NVIDIA Licensing** dashboard depends on whether a license server has already been installed on the DLS virtual appliance.
   - If a license server has **not** been installed on the DLS virtual appliance, the **NVIDIA Licensing** dashboard displays a message asking if you want to install a license server.
   - Otherwise, the **NVIDIA Licensing** dashboard displays the **License Server Details** page for the installed license server.
7. Install or update the license server on the DLS virtual appliance.

   Whether you install or update the license server depends on whether a license server has already been installed on the DLS virtual appliance.

   - If a license server has **not** been installed on the DLS virtual appliance, on the **NVIDIA Licensing** dashboard, click **SELECT LICENSE SERVER FILE**.



Dashboard

Complete basic setup

Service instance not registered

This service instance has not yet been registered with the NVIDIA Licensing Portal

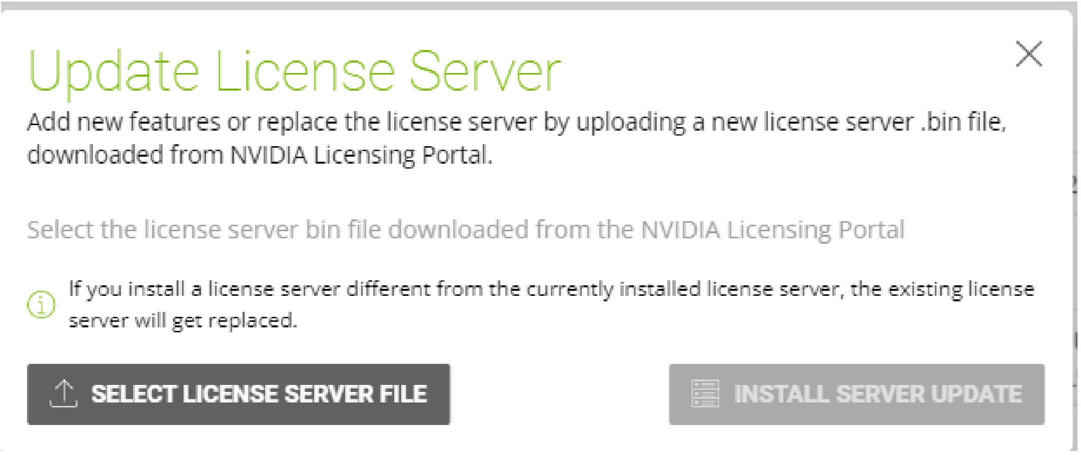
No license server installed

Select the license server bin file downloaded from the NVIDIA Licensing Portal

- If a license server has already been installed on the DLS virtual appliance, update the license server.

  a). From the **ACTIONS** menu for the license server, choose **Update server from NLP**.

  b). In the **Update License Server** pop-up window that opens, click **SELECT LICENSE SERVER FILE**.



Update License Server

Add new features or replace the license server by uploading a new license server .bin file, downloaded from NVIDIA Licensing Portal.

Select the license server bin file downloaded from the NVIDIA Licensing Portal

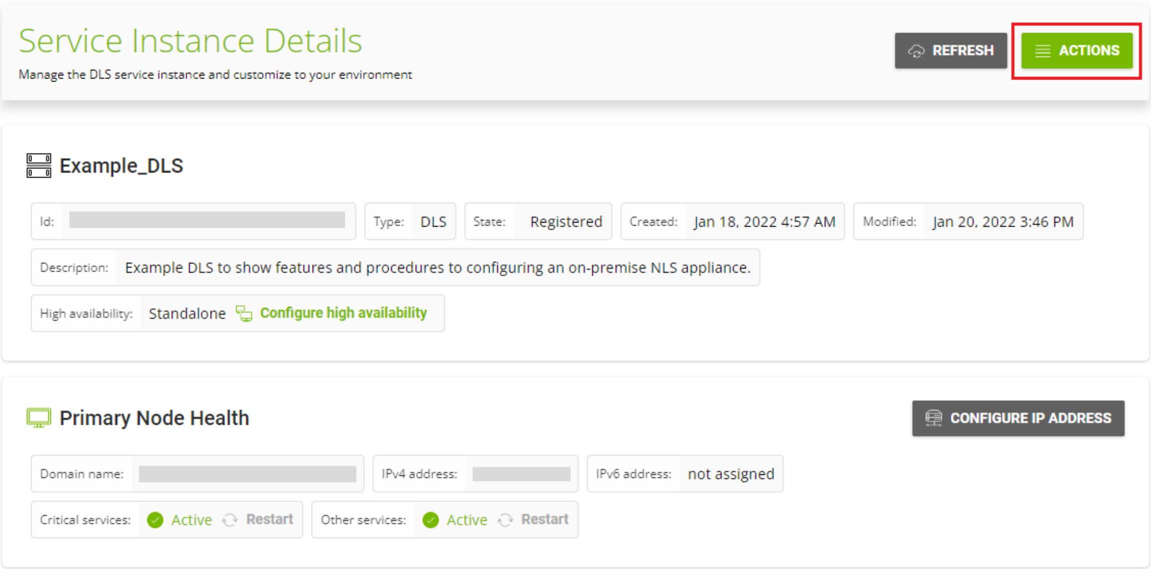
If you install a license server different from the currently installed license server, the existing license server will get replaced.

8. In the file browser that opens, navigate to the folder that contains the license server file named `license_`*`mm-dd-yyyy-hh-mm-ss`*`.bin` that you downloaded and select the file.
9. When asked if you want to install the selected file, click **INSTALL**.

NVIDIA Licensing dashboard is updated with the details of the license server that you installed.

2.3.8. Generating a Client Configuration Token for a DLS Instance

1. If you are not already logged in, log in to the **NVIDIA Licensing** application at the IP address of the VM on which the DLS instance resides.
2. In the left navigation pane, click **SERVICE INSTANCE**.
3. On the **Service Instance** page that opens, from the **Actions** menu for the DLS instance for which you want to generate a client configuration token, choose **Generate client configuration token**.



Service Instance Details

Manage the DLS service instance and customize to your environment

Example_DLS

Id: [redacted] Type: DLS State: Registered Created: Jan 18, 2022 4:57 AM Modified: Jan 20, 2022 3:46 PM

Description: Example DLS to show features and procedures to configuring an on-premise NLS appliance.

High availability: Standalone

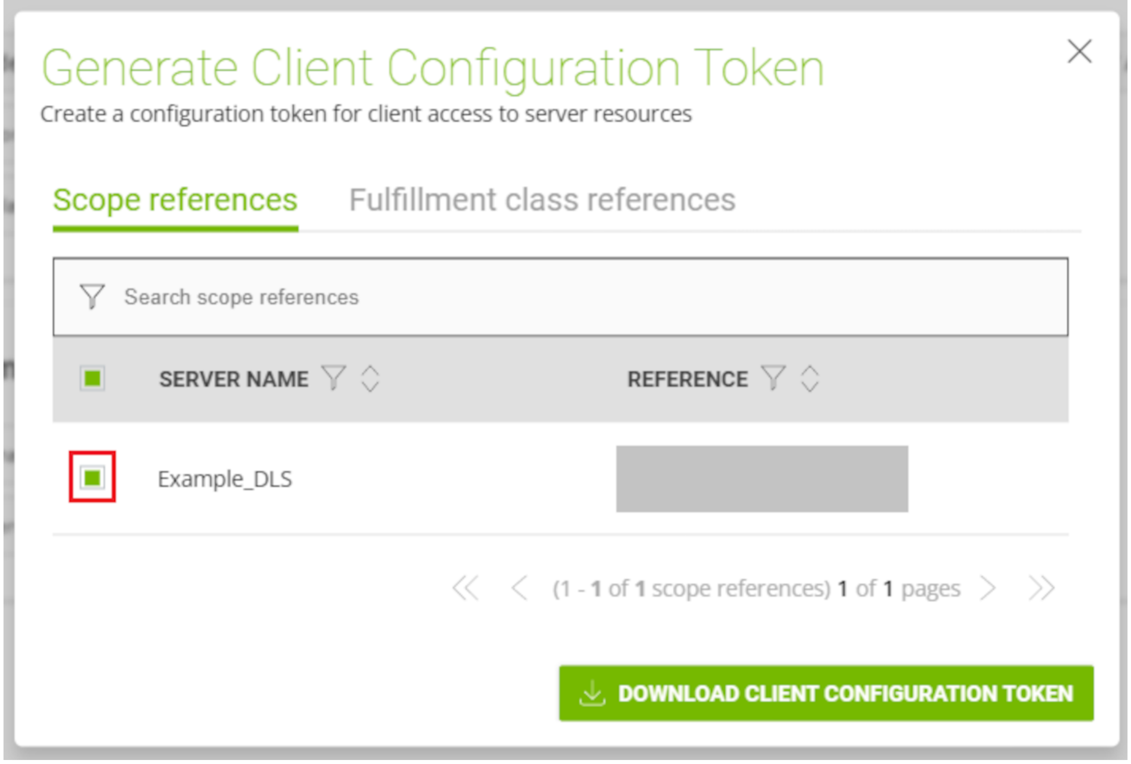
Primary Node Health

Domain name: [redacted] IPv4 address: [redacted] IPv6 address: not assigned

Critical services: Active Other services: Active

4. In the **Generate Client Configuration Token** pop-up window that opens, select the references that you want to include in the client configuration token.

   a). Click the **Scope references** tab, and from the list of scope references, select the scope references that you want to include.



Generate Client Configuration Token

Create a configuration token for client access to server resources

Scope references
Fulfillment class references
Search scope references

☑	SERVER NAME	REFERENCE
☑	Example_DLS	

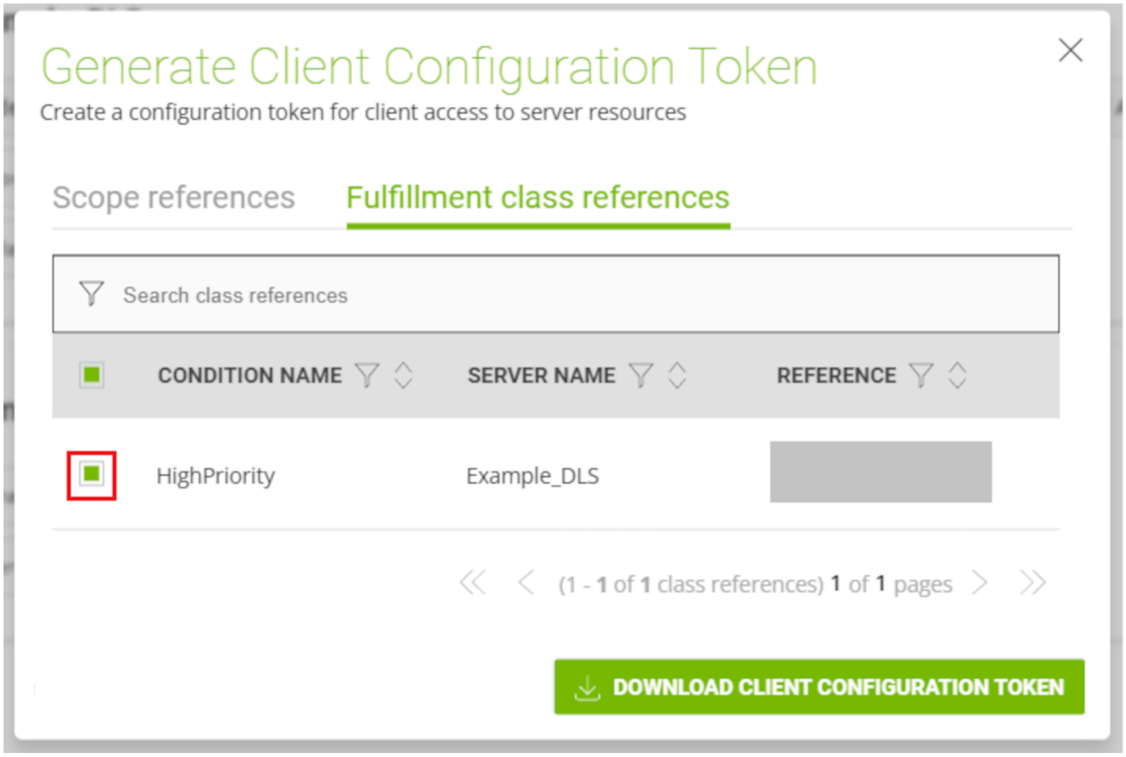
(1 - 1 of 1 scope references) 1 of 1 pages

DOWNLOAD CLIENT CONFIGURATION TOKEN

You must select **at least one** scope reference.

Each scope reference specifies the license server that will fulfil a license request.

b). **Optional:** Click the **Fulfillment class references** tab, and from the list of fulfillment class references, select the fulfillment class references that you want to include.



Generate Client Configuration Token

Create a configuration token for client access to server resources

Scope references
Fulfillment class references
Search class references

☑	CONDITION NAME	SERVER NAME	REFERENCE
☑	HighPriority	Example_DLS	[redacted]

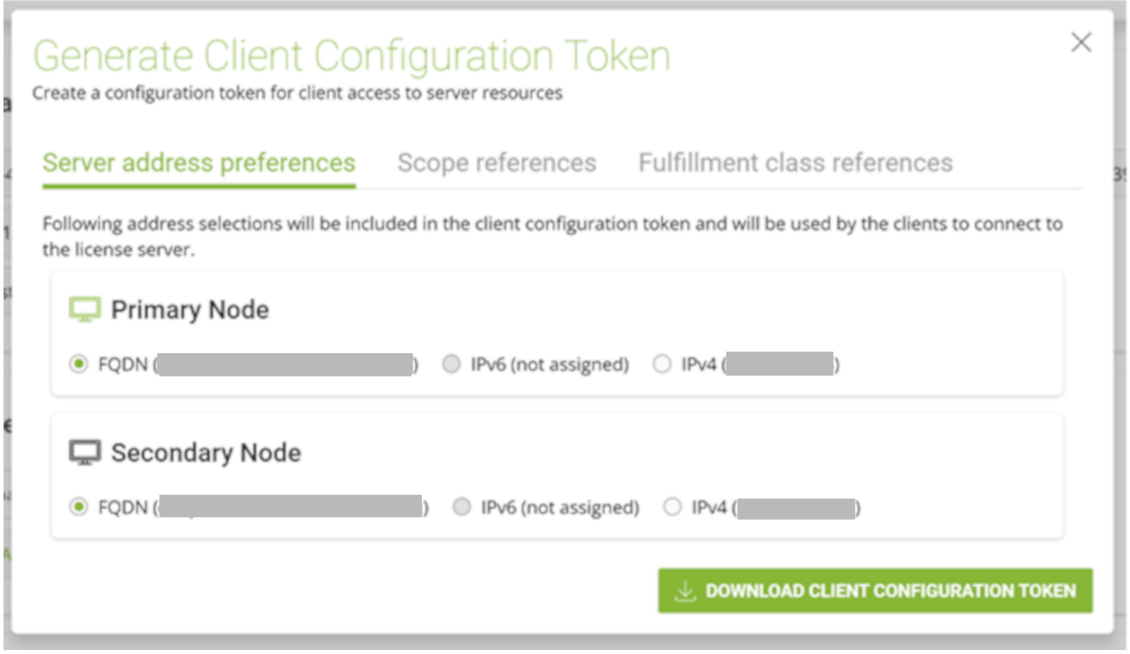
(1 - 1 of 1 class references) 1 of 1 pages

DOWNLOAD CLIENT CONFIGURATION TOKEN

Including fulfillment class references is optional.

c). **Optional:** If you want the service instance, or each node in an HA cluster of instances, to be identified through its IP address, click the **Server address preferences** tab and select the address for the IP version that you want: **IPv6** or **IP v4**.

By default, a service instance, or each node in an HA cluster of instances, is identified through its fully qualified domain name.



Generate Client Configuration Token

Create a configuration token for client access to server resources

Server address preferences | Scope references | Fulfillment class references

Following address selections will be included in the client configuration token and will be used by the clients to connect to the license server.

Primary Node

● FQDN ([redacted]) ○ IPv6 (not assigned) ○ IPv4 ([redacted])

Secondary Node

● FQDN ([redacted]) ○ IPv6 (not assigned) ○ IPv4 ([redacted])

DOWNLOAD CLIENT CONFIGURATION TOKEN

d). **Optional:** In the **Expiration** section, select an expiration date for the client configuration token. If you do not select a date, the default token expiration time is 12 years.

e). Click **DOWNLOAD CLIENT CONFIGURATION TOKEN**.

A file named `client_configuration_token_`*`mm-dd-yyyy-hh-mm-ss`*`.tok` is saved to your default downloads folder.

After creating a client configuration token from a service instance, copy the client configuration token to each licensed client that you want to use the combination of license servers and fulfillment conditions specified in the token. For more information, see Configuring a Licensed Client.

You can decouple the leasing port from the UI port for auth and lease operations. Once you have done so, you can block the UI port for the client VM.

Note: For backward compatibility, leasing operations will still be supported by the default HTTPS port (`443`) in the VM version.

- All UI and leasing operations will be supported on the default HTTPS port, `443`.
- Only leasing operations will be supported on the leasing port, `8082`.

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 display-off and display-enabled modes but must be used in NVIDIA vGPU software deployments in display-off 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 display-off mode, but other GPUs are supplied in a display-enabled mode.

GPU	Mode as Supplied from the Factory
NVIDIA A40	Display-off
NVIDIA L40	Display-off
NVIDIA RTX 6000 Ada	Display enabled
NVIDIA RTX A5000	Display enabled
NVIDIA RTX A5500	Display enabled
NVIDIA RTX A6000	Display enabled

A GPU that is supplied from the factory in display-off 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 NVIDIA Display Mode Selector Tool page on the NVIDIA Developer website.

> **Note:**
>
> Only the following GPUs support the `displaymodeselector` tool:
>
> - NVIDIA A40
> - NVIDIA L40
> - NVIDIA RTX A5000
> - NVIDIA RTX 6000 Ada
> - NVIDIA RTX A5500
> - 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 `gpumodeswitch` tool. Other GPUs that support NVIDIA vGPU do not support the `gpumodeswitch` tool and, except as stated in Switching the Mode of a GPU that Supports Multiple Display Modes, 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

The NVIDIA Virtual GPU Manager runs on the ESXi host. It is distributed as a number of software components in a ZIP archive.

The NVIDIA Virtual GPU Manager software components are as follows:

- A software component for the NVIDIA vGPU hypervisor host driver
- A software component for the NVIDIA GPU Management daemon

Before you begin, ensure that the following prerequisites are met:

- The ZIP archive that contains NVIDIA vGPU software has been downloaded from the NVIDIA Licensing Portal.
- The software components for the NVIDIA Virtual GPU Manager have been extracted from the downloaded ZIP archive.

1. Copy the NVIDIA Virtual GPU Manager component files to the ESXi host.
2. Put the ESXi host into maintenance mode.

   ```
   $ esxcli system maintenanceMode set --enable true
   ```

3. Install the NVIDIA vGPU hypervisor host driver and the NVIDIA GPU Management daemon from their software component files.

   a). Run the `esxcli` command to install the NVIDIA vGPU hypervisor host driver from its software component file.

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

   b). Run the `esxcli` command to install the NVIDIA GPU Management daemon from its software component file.

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

   datastore

   The name of the VMFS datastore to which you copied the software components.

   host-driver-component

   The name of the file that contains the NVIDIA vGPU hypervisor host driver in the form of a software component. Ensure that you specify the file that was extracted from the downloaded ZIP archive. For example, for VMware vSphere 7.0.2, *host-*

driver-component is **NVD-VMware-x86_64-525.147.01-1OEM.702.0.0.17630552-bundle-*build-number***.

gpu-management-daemon-component

The name of the file that contains the NVIDIA GPU Management daemon in the form of a software component. Ensure that you specify the file that was extracted from the downloaded ZIP archive. For example, for VMware vSphere 7.0.2, *gpu-management-daemon-component* is **VMW-esx-7.0.2-nvd-gpu-mgmt-daemon-1.0-0.0.0001**.

4. Exit maintenance mode.

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

5. Reboot the ESXi host.

```
$ reboot
```

6. Verify that the NVIDIA GPU Management daemon has started.

```
$ /etc/init.d/nvdGpuMgmtDaemon status
```

7. 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 double-bit 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 C-series 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. To determine whether ECC memory is enabled for a GPU, run **`nvidia-smi -q`** for the GPU.

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.

   ```
   # nvidia-smi -q

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

   Timestamp                           : Mon Nov 13 18:36:45 2023
   Driver Version                      : 525.147.01

   Attached GPUs                       : 1
   GPU 0000:02:00.0

   [...]

       Ecc Mode
           Current                     : Enabled
           Pending                     : 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==============

   Timestamp                           : Mon Nov 13 18:37:53 2023
   Driver Version                      : 525.147.01

   Attached GPUs                       : 1
   GPU 0000:02:00.0
   [...]

       Ecc Mode
           Current                     : Disabled
           Pending                     : Disabled

   [...]
   ```

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.

   ```
   # nvidia-smi -q

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

   Timestamp                           : Mon Nov 13 18:36:45 2023
   Driver Version                      : 525.147.01

   Attached GPUs                       : 1
   GPU 0000:02:00.0

   [...]

       Ecc Mode
           Current                     : Disabled
           Pending                     : 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                           : Mon Nov 13 18:37:53 2023
Driver Version                      : 525.147.01

Attached GPUs                       : 1
GPU 0000:02:00.0
[...]

    Ecc Mode
        Current                     : Enabled
        Pending                     : 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

The vGPU Manager VIB for VMware vSphere provides vSGA and vGPU functionality in a single VIB. After this VIB is installed, the default graphics type is Shared, which provides vSGA functionality. To enable vGPU support for VMs in VMware vSphere, you must change the default graphics type to Shared Direct.

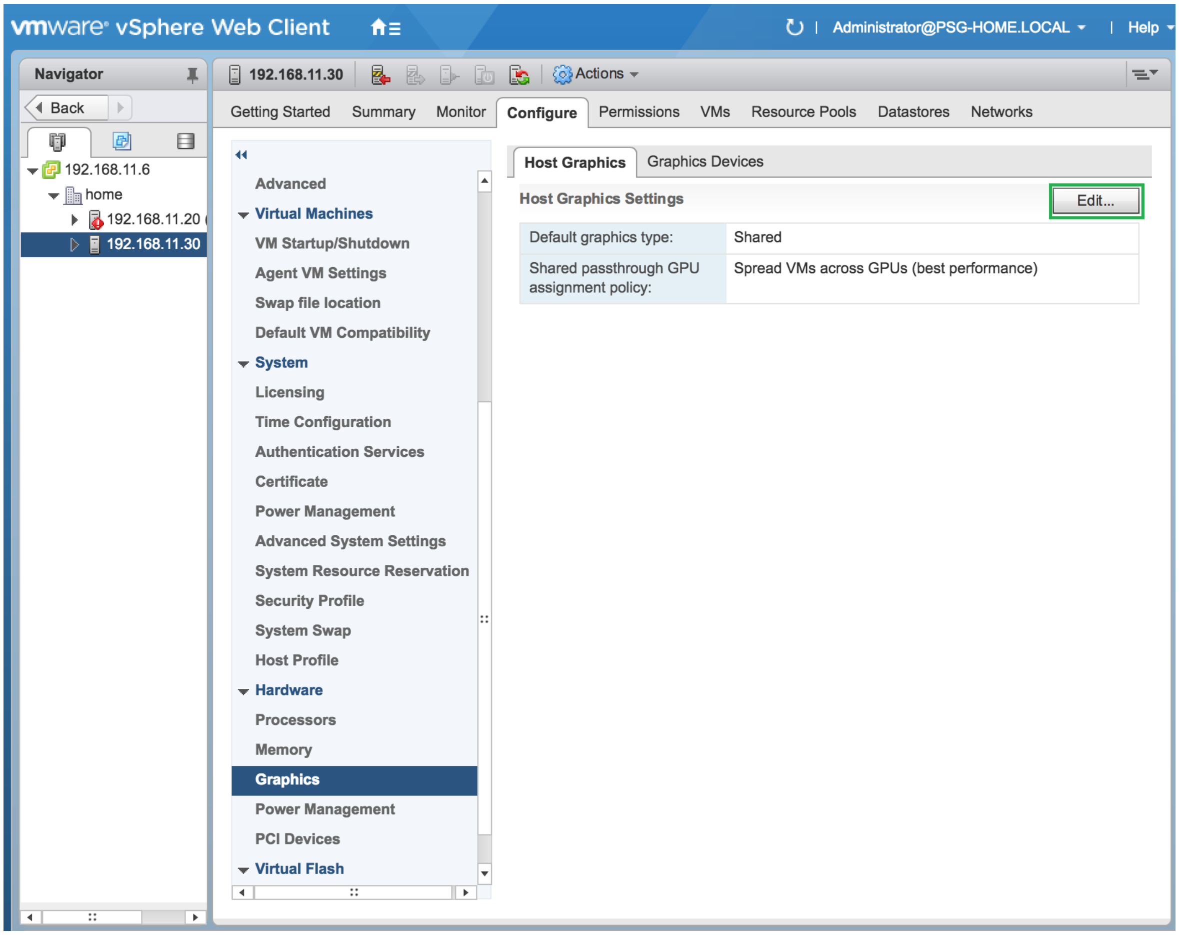
If you do not change the default graphics type, VMs to which a vGPU is assigned fail to start and the following error message is displayed:

```
The amount of graphics resource available in the parent resource pool is
 insufficient for the operation.
```

> **Note:** Change the default graphics type **before** configuring vGPU. Output from the VM console in the VMware vSphere Web Client is not available for VMs that are running vGPU.

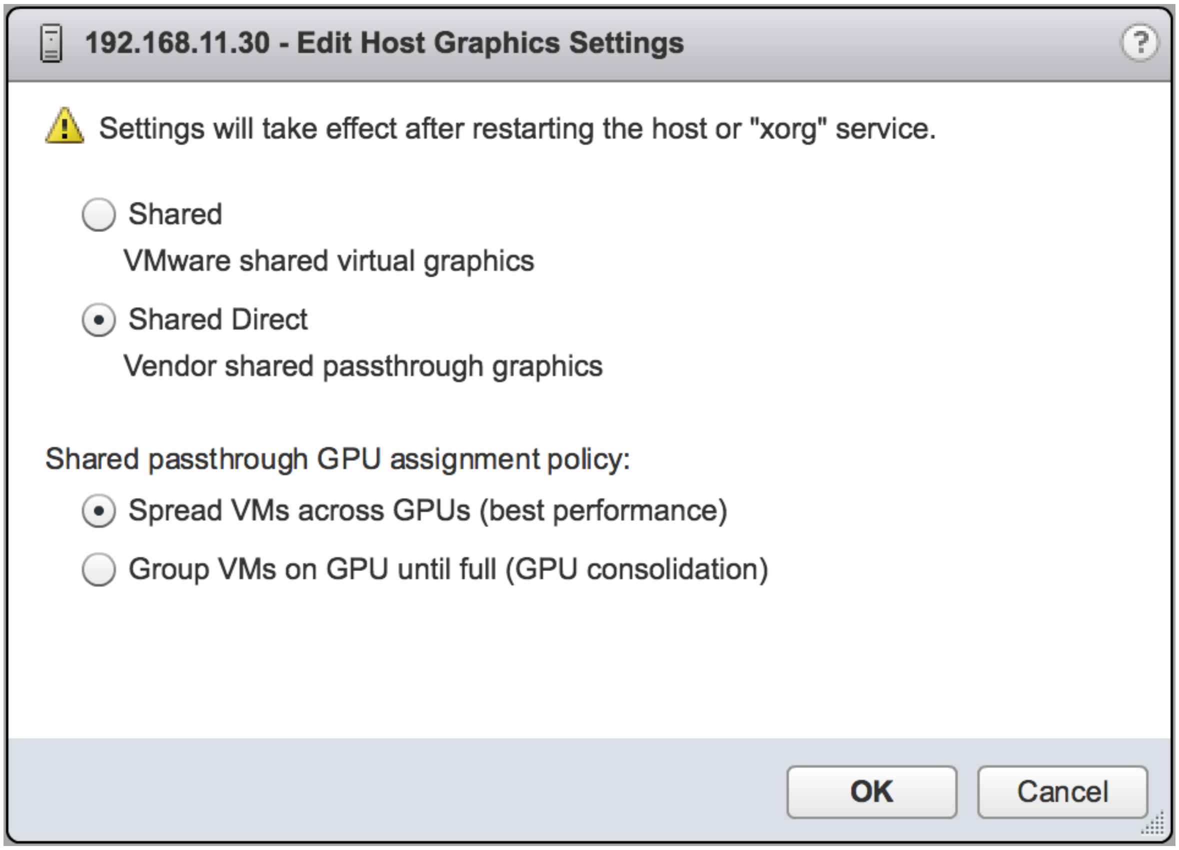
Before changing the default graphics type, ensure that the ESXi host is running and that all VMs on the host are powered off.

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**.



Default graphics type:	Shared
Shared passthrough GPU assignment policy:	Spread VMs across GPUs (best performance)

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

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)

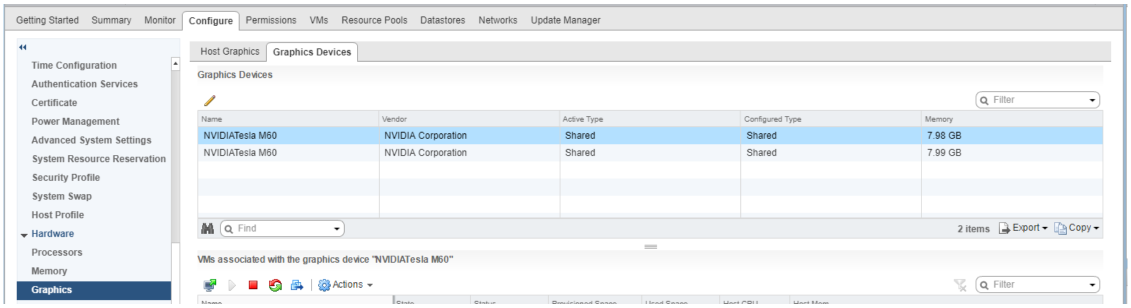
○ Group VMs on GPU until full (GPU consolidation)

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**.



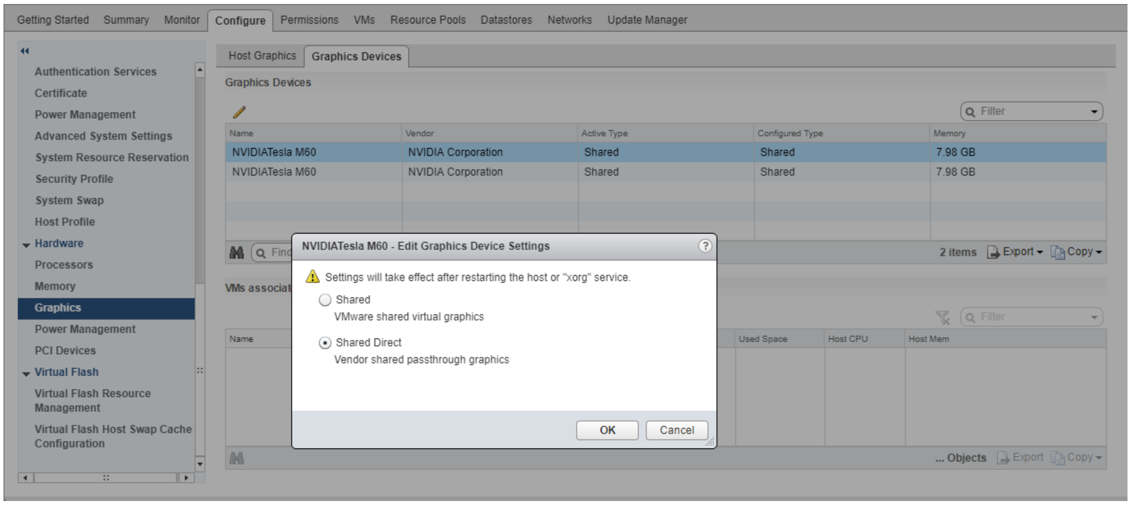
Graphics Devices

Name	Vendor	Active Type	Configured Type	Memory
NVIDIATesla M60	NVIDIA Corporation	Shared	Shared	7.98 GB
NVIDIATesla M60	NVIDIA Corporation	Shared	Shared	7.99 GB

2 items

VMs associated with the graphics device "NVIDIATesla M60"

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



Graphics Devices

Name	Vendor	Active Type	Configured Type	Memory
NVIDIATesla M60	NVIDIA Corporation	Shared	Shared	7.98 GB
NVIDIATesla M60	NVIDIA Corporation	Shared	Shared	7.98 GB

NVIDIATesla M60 - Edit Graphics Device Settings

Settings will take effect after restarting the host or "xorg" service.

○ Shared
VMware shared virtual graphics

● Shared Direct
Vendor shared passthrough graphics

OK Cancel

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.

   The Xorg service is not 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`.

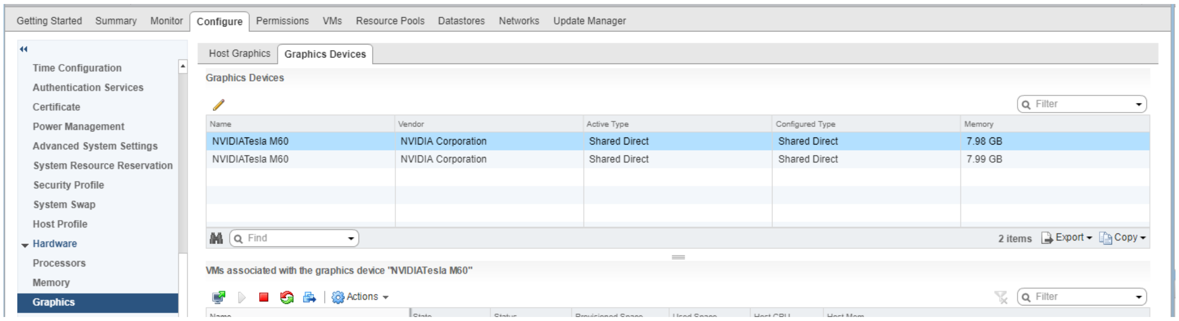
   ```
   [root@esxi:~] nv-hostengine -d
   ```

   e). **VMware vSphere releases before 7.0 Update 1 only:** Start the Xorg service.

   The Xorg service is not 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.



Graphics Devices

Name	Vendor	Active Type	Configured Type	Memory
NVIDIATesla M60	NVIDIA Corporation	Shared Direct	Shared Direct	7.98 GB
NVIDIATesla M60	NVIDIA Corporation	Shared Direct	Shared Direct	7.99 GB

VMs associated with the graphics 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.

How to configure a vSphere VM with a vGPU depends on your VMware vSphere version as explained in the following topics:

- Configuring a vSphere 8 VM with NVIDIA vGPU
- Configuring a vSphere 7 VM with NVIDIA vGPU

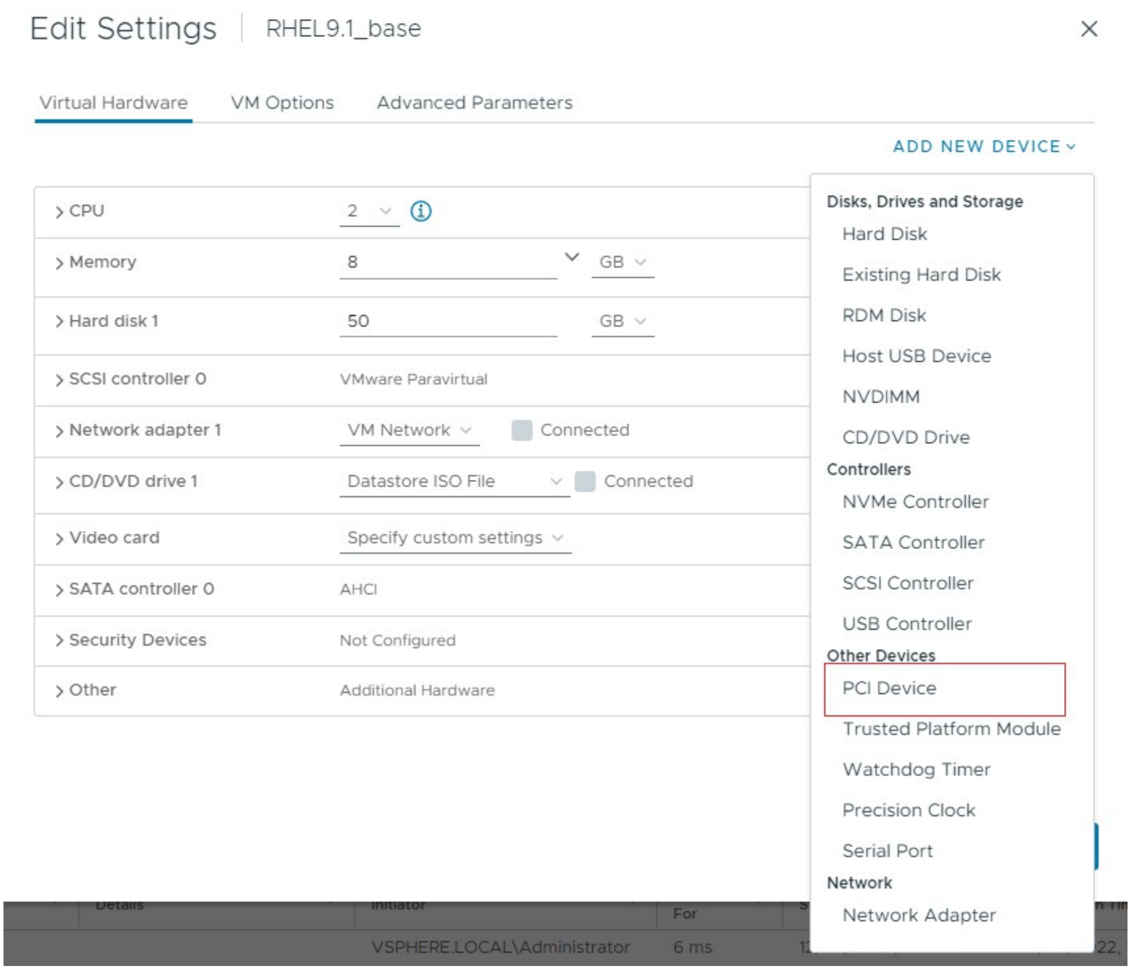
After you have configured a vSphere VM with a vGPU, start the VM. VM console in vSphere Web Client is not supported in this vGPU release. Therefore, use VMware Horizon or VNC to access the VM's desktop.

3.5.2.1. Configuring a vSphere 8 VM with NVIDIA vGPU

1. Open the vCenter Web UI.
2. In the vCenter Web UI, right-click the VM and choose **Edit Settings**.
3. In the **Edit Settings** window that opens, configure the vGPUs that you want to add to the VM.

   Add each vGPU that you want to add to the VM as follows:

   a). From the **ADD NEW DEVICE** menu, choose **PCI Device**.



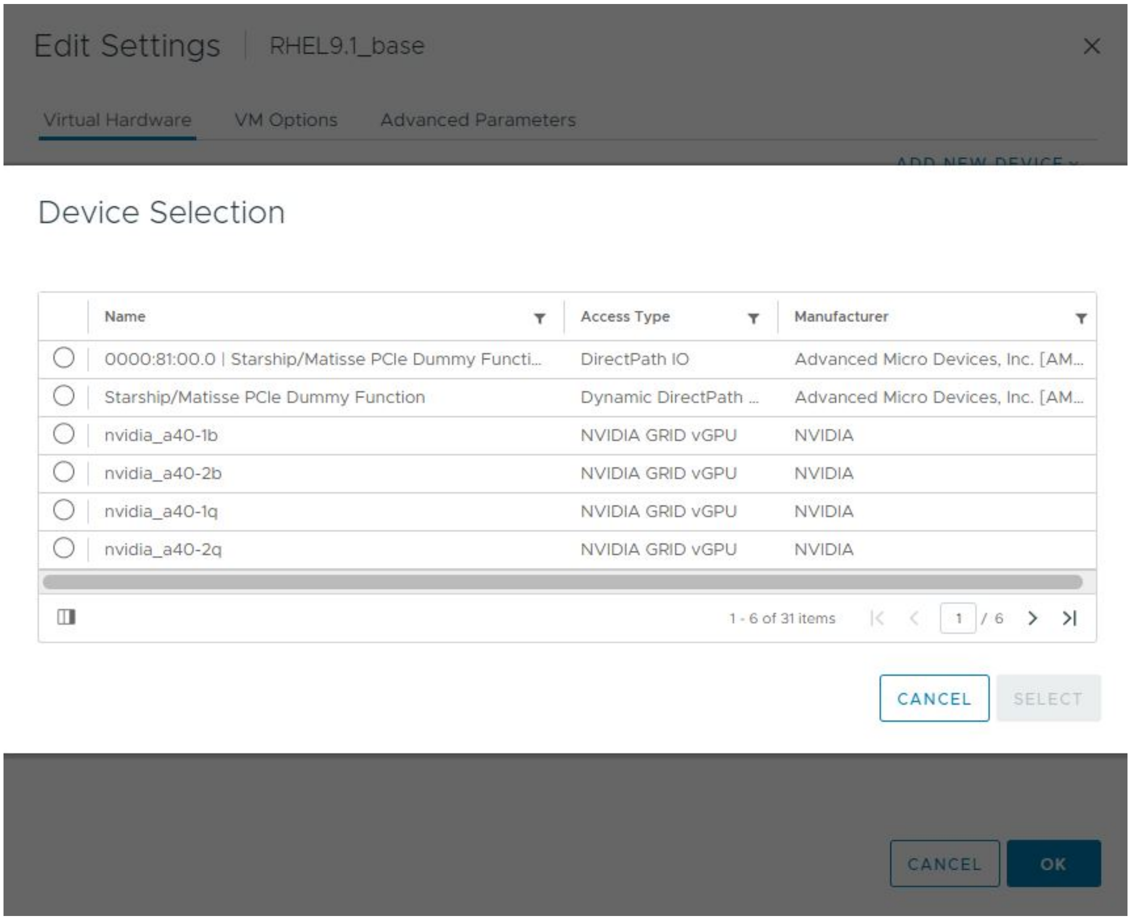
Edit Settings | RHEL9.1_base

Virtual Hardware | VM Options | Advanced Parameters

- CPU: 2
- Memory: 8 GB
- Hard disk 1: 50 GB
- SCSI controller 0: VMware Paravirtual
- Network adapter 1: VM Network □ Connected
- CD/DVD drive 1: Datastore ISO File □ Connected
- Video card: Specify custom settings
- SATA controller 0: AHCI
- Security Devices: Not Configured
- Other: Additional Hardware

b). In the **Device Selection** window that opens, select the type of vGPU you want to configure and click **SELECT**.

> **Note:** NVIDIA vGPU software does **not** support vCS on VMware vSphere. Therefore, C-series vGPU types are not available for selection in the **Device Selection** window.



	Name	Access Type	Manufacturer
○	0000:81:00.0 | Starship/Matisse PCIe Dummy Functi...	DirectPath IO	Advanced Micro Devices, Inc. [AM...
○	Starship/Matisse PCIe Dummy Function	Dynamic DirectPath ...	Advanced Micro Devices, Inc. [AM...
○	nvidia_a40-1b	NVIDIA GRID vGPU	NVIDIA
○	nvidia_a40-2b	NVIDIA GRID vGPU	NVIDIA
○	nvidia_a40-1q	NVIDIA GRID vGPU	NVIDIA
○	nvidia_a40-2q	NVIDIA GRID vGPU	NVIDIA

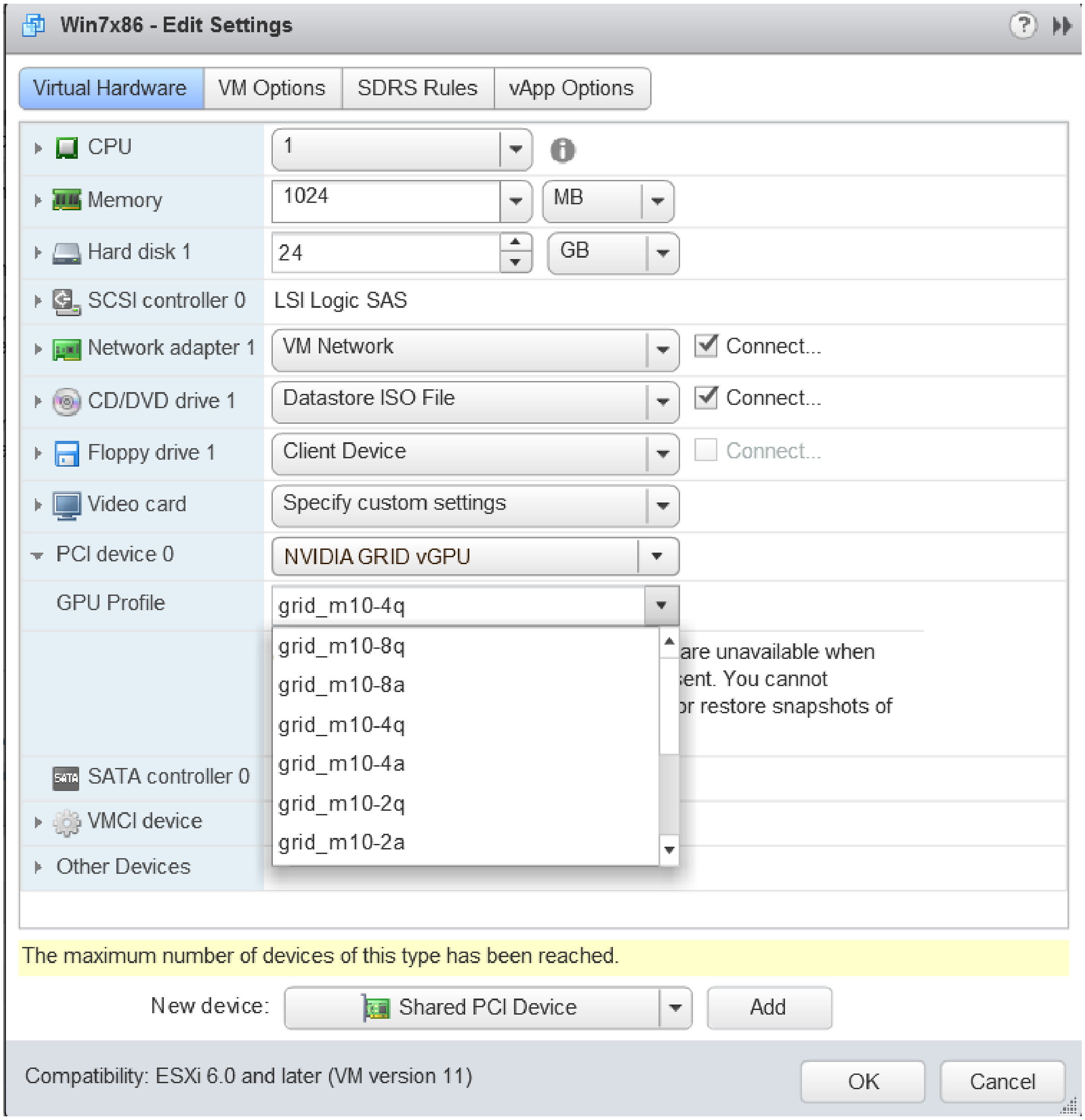
4. Back in the **Edit Settings** window, click **OK**.

3.5.2.2. Configuring a vSphere 7 VM with NVIDIA vGPU

If you are adding multiple vGPUs to a single VM, perform this task for each vGPU that you want to add to the VM.

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.
4. 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

CPU	1		
Memory	1024	MB	
Hard disk 1	24	GB	
SCSI controller 0	LSI Logic SAS		
Network adapter 1	VM Network	☑ Connect...	
CD/DVD drive 1	Datastore ISO File	☑ Connect...	
Floppy drive 1	Client Device	□ Connect...	
Video card	Specify custom settings		
PCI device 0	NVIDIA GRID vGPU		
GPU Profile	grid_m10-4q		
SATA controller 0			
VMCI device			
Other Devices			

are unavailable when
ent. You cannot
or restore snapshots of

The maximum number of devices of this type has been reached.

New device: Shared PCI Device

Compatibility: ESXi 6.0 and later (VM version 11)

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

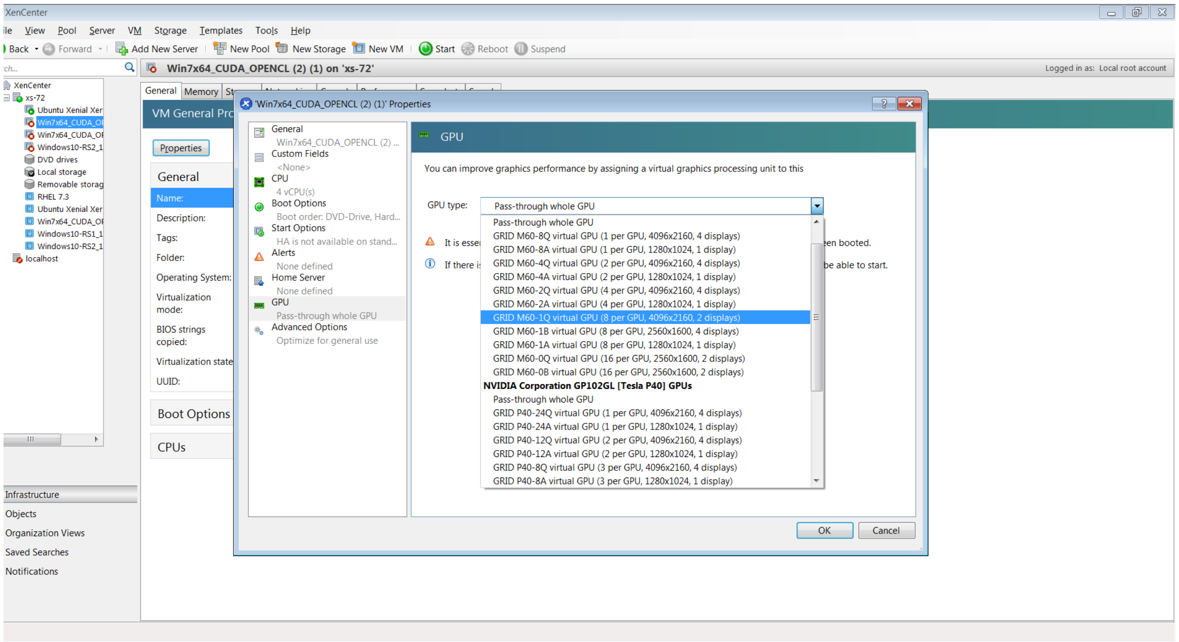
> **Note:** NVIDIA vGPU software does **not** support vCS on VMware vSphere. Therefore, C-series vGPU types are not available for selection from the **GPU Profile** drop-down menu.

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)**.

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.

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



'Win7x64_CUDA_OPENCL (2) (1)' Properties

GPU

You can improve graphics performance by assigning a virtual graphics processing unit to this

GPU type: Pass-through whole GPU

XenCenter
File View Pool Server VM Storage Templates Tools Help
Back Forward Add New Server New Pool New Storage New VM Start Reboot Suspend
Win7x64_CUDA_OPENCL (2) (1) on 'xs-72'
Logged in as: Local root account
General
Win7x64_CUDA_OPENCL (2) ...
Custom Fields
<None>
CPU
4 vCPU(s)
Boot Options
Boot order: DVD-Drive, Hard...
Start Options
HA is not available on stand...
Alerts
None defined
Home Server
None defined
GPU
Pass-through whole GPU
Advanced Options
Optimize for general use
Pass-through whole GPU
GRID M60-8Q virtual GPU (1 per GPU, 4096x2160, 4 displays)
GRID M60-8A virtual GPU (1 per GPU, 1280x1024, 1 display)
GRID M60-4Q virtual GPU (2 per GPU, 4096x2160, 4 displays)
GRID M60-4A virtual GPU (2 per GPU, 1280x1024, 1 display)
GRID M60-2Q virtual GPU (4 per GPU, 4096x2160, 4 displays)
GRID M60-2A virtual GPU (4 per GPU, 1280x1024, 1 display)
GRID M60-1Q virtual GPU (8 per GPU, 4096x2160, 2 displays)
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)
GRID M60-0B virtual GPU (16 per GPU, 2560x1600, 2 displays)
NVIDIA Corporation GP102GL [Tesla P40] GPUs
Pass-through whole GPU
GRID P40-24Q virtual GPU (1 per GPU, 4096x2160, 4 displays)
GRID P40-24A virtual GPU (1 per GPU, 1280x1024, 1 display)
GRID P40-12Q virtual GPU (2 per GPU, 4096x2160, 4 displays)
GRID P40-12A virtual GPU (2 per GPU, 1280x1024, 1 display)
GRID P40-8Q virtual GPU (3 per GPU, 4096x2160, 4 displays)
GRID P40-8A virtual GPU (3 per GPU, 1280x1024, 1 display)
OK
Cancel
Infrastructure
Objects
Organization Views
Saved Searches
Notifications

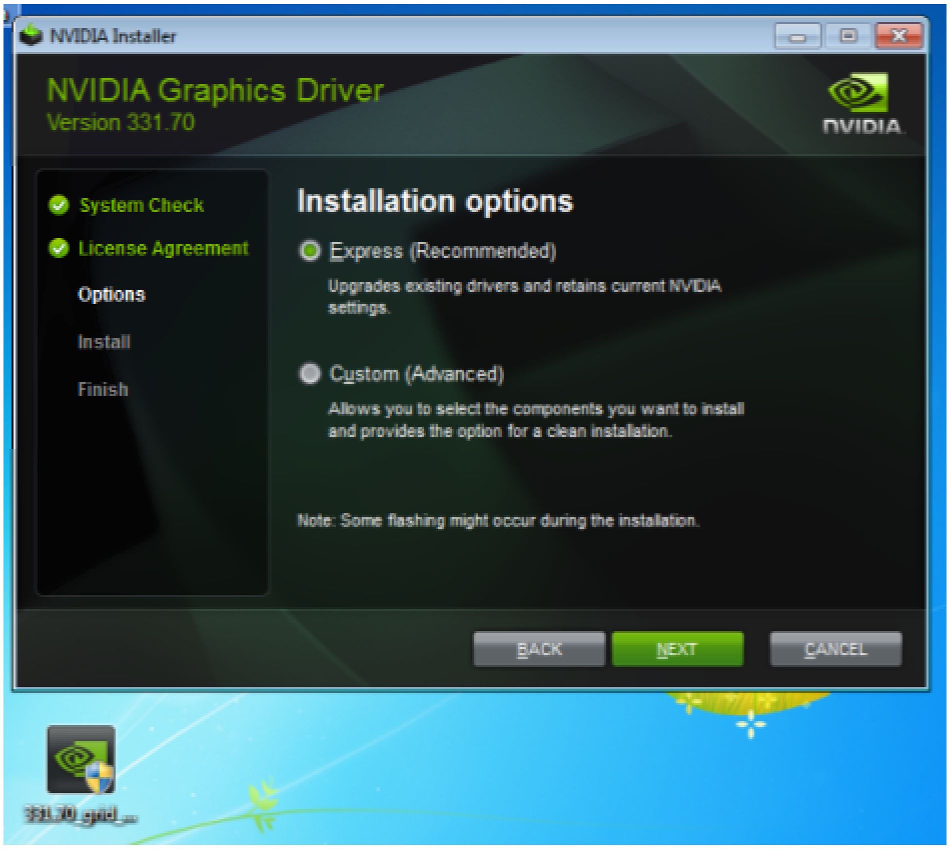
After you have configured a Citrix Hypervisor VM with a vGPU, start the VM, either from XenCenter or by using `xe vm-start` in a dom0 shell. You can view the VM's console in XenCenter.

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.



Installation options

● Express (Recommended)

Upgrades existing drivers and retains current NVIDIA settings.

○ Custom (Advanced)

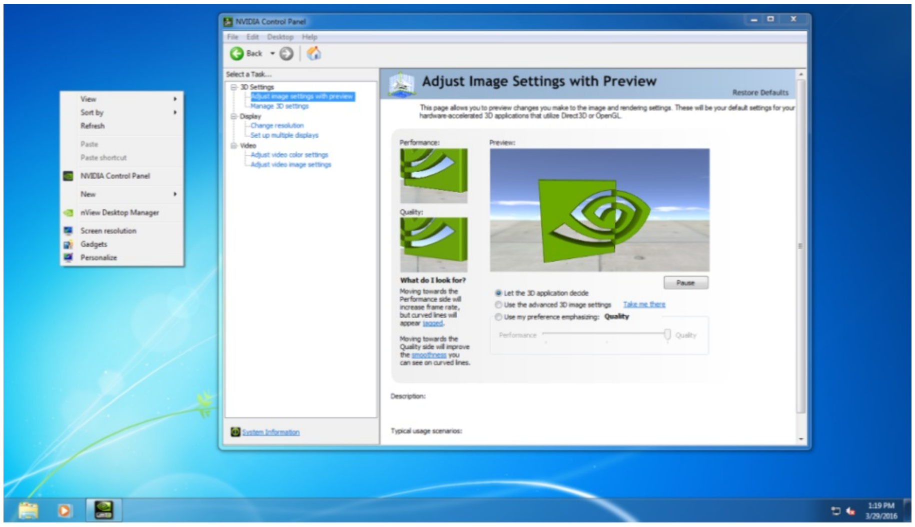
Allows you to select the components you want to install and provides the option for a clean installation.

Note: Some flashing might occur during the installation.

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.



This page allows you to preview changes you make to the image and rendering settings. These will be your default settings for your hardware-accelerated 3D applications that utilize Direct3D or OpenGL.

What do I look for?

Moving towards the Performance side will increase frame rate, but curved lines will appear jagged.

Moving towards the Quality side will improve the smoothness you can see on curved lines.

● Let the 3D application decide

○ Use the advanced 3D image settings

○ Use my preference emphasizing: **Quality**

3.7. Configuring a Licensed Client

A client with a network connection obtains a license by leasing it from a NVIDIA License System service instance. The service instance serves the license to the client over the network from a pool of floating licenses obtained from the NVIDIA Licensing Portal. The license is returned to the service instance when the licensed client no longer requires the license.

The graphics driver creates a default location in which to store the client configuration token on the client.

The process for configuring a licensed client is the same for CLS and DLS instances but depends on the OS that is running on the client.

3.7.1. Configuring a Licensed Client on Windows with Default Settings

Perform this task from the client.

1. Copy the client configuration token to the `%SystemDrive%:\Program Files\NVIDIA Corporation\vGPU Licensing\ClientConfigToken` folder.
2. Restart the `NvDisplayContainer` service.

The NVIDIA service on the client should now automatically obtain a license from the CLS or DLS instance.

3.7.2. Verifying the NVIDIA vGPU Software License Status of a Licensed Client

After configuring a client with an NVIDIA vGPU software license, verify the license status by displaying the licensed product name and status.

To verify the license status of a licensed client, run `nvidia-smi` with the `-q` or `--query` optionfrom the licensed client, **not** the hypervisor host. If the product is licensed, the expiration date is shown in the license status.

```
nvidia-smi -q
==============NVSMI LOG==============

Timestamp                                 : Wed Nov 23 10:52:59 2022
Driver Version                            : 525.60.06
CUDA Version                              : 12.0

Attached GPUs                             : 2
GPU 00000000:02:03.0
    Product Name                          : NVIDIA A2-8Q
    Product Brand                         : NVIDIA RTX Virtual Workstation
    Product Architecture                  : Ampere
    Display Mode                          : Enabled
    Display Active                        : Disabled
    Persistence Mode                      : Enabled
    MIG Mode
        Current                           : Disabled
        Pending                           : Disabled
    Accounting Mode                       : Disabled
    Accounting Mode Buffer Size           : 4000
    Driver Model
        Current                           : N/A
        Pending                           : N/A
    Serial Number                         : N/A
    GPU UUID                              : GPU-ba5b1e9b-1dd3-11b2-be4f-98ef552f4216
    Minor Number                          : 0
    VBIOS Version                         : 00.00.00.00.00
    MultiGPU Board                        : No
    Board ID                              : 0x203
    Board Part Number                     : N/A
    GPU Part Number                       : 25B6-890-A1
    Module ID                             : N/A
    Inforom Version
        Image Version                     : N/A
        OEM Object                        : N/A
        ECC Object                        : N/A
        Power Management Object           : N/A
    GPU Operation Mode
        Current                           : N/A
        Pending                           : N/A
    GSP Firmware Version                  : N/A
    GPU Virtualization Mode
        Virtualization Mode               : VGPU
        Host VGPU Mode                    : N/A
    vGPU Software Licensed Product
        Product Name                      : NVIDIA RTX Virtual Workstation
        License Status                    : Licensed (Expiry: 2022-11-23 10:41:16
 GMT)
    …
    …
```

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.

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

NVIDIA Corporation | 2788 San Tomas Expressway, Santa Clara, CA 95051
http://www.nvidia.com

NVIDIA.