

NVIDIA Virtual GPU Software Quick Start Guide

Release 21.0

NVIDIA Corporation

Nov 05, 2026

Contents

1 Getting NVIDIA vGPU Software **1**
1.1 Before You Begin . . . 1
1.2 Your Order Confirmation Message . . . 2
1.3 NVIDIA Enterprise Account Requirements . . . 2
1.4 Creating your NVIDIA Enterprise Account . . . 3
1.5 Linking an Evaluation Account to an NVIDIA Enterprise Account for Purchased Licenses 4
1.6 Downloading NVIDIA vGPU Software . . . 5

2 Installing Your NVIDIA vGPU Software License Server and License Files **7**
2.1 Introduction to NVIDIA Software Licensing . . . 7
2.2 Configuring a CLS Instance . . . 7
2.2.1 Creating a License Server on the NVIDIA Licensing Portal . . . 8
2.2.2 Creating a CLS Instance on the NVIDIA Licensing Portal . . . 9
2.2.3 Binding a License Server to a Service Instance . . . 9
2.2.4 Installing a License Server on a CLS Instance . . . 10
2.2.5 Generating a Client Configuration Token for a CLS Instance . . . 10
2.3 Configuring a DLS Instance . . . 11
2.3.1 Registering the DLS Administrator User . . . 11
2.3.2 Configuring a Standalone DLS Instance . . . 12
2.3.3 Changing the Name and Description of a DLS Instance . . . 12
2.3.4 Creating a License Server on the NVIDIA Licensing Portal . . . 13
2.3.5 Registering an on-Premises DLS Instance with the NVIDIA Licensing Portal . . . 14
2.3.6 Binding a License Server to a Service Instance . . . 15
2.3.7 Installing a License Server on a DLS Instance . . . 15
2.3.8 Generating a Client Configuration Token for a DLS Instance . . . 16

3 Installing and Configuring NVIDIA vGPU Manager and the Guest Driver **19**
3.1 Switching the Mode of a GPU that Supports Multiple Display Modes . . . 19
3.2 Installing the NVIDIA Virtual GPU Manager . . . 20
3.2.1 Installing the NVIDIA Virtual GPU Manager on VMware vSphere . . . 20
3.2.2 Installing the NVIDIA Virtual GPU Manager on XenServer . . . 22
3.3 Disabling and Enabling ECC Memory . . . 23
3.3.1 Disabling ECC Memory . . . 24
3.3.2 Enabling ECC Memory . . . 25
3.4 Attaching an NVIDIA vGPU Profile to a VM . . . 27
3.4.1 Changing the Graphics Type in VMware vSphere . . . 27
3.4.2 Configuring a vSphere VM with NVIDIA vGPU . . . 30
3.4.3 Configuring a XenServer VM with Virtual GPU . . . 32
3.5 Installing the NVIDIA vGPU Software Graphics Driver . . . 33
3.6 Configuring a Licensed Client . . . 35
3.6.1 Configuring a Licensed Client on Windows with Default Settings . . . 35
3.6.2 Verifying the NVIDIA vGPU Software License Status of a Licensed Client . . . 35

4 Notices 37

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

These web properties can be reached from the NVIDIA Application Hub.

Note

If you obtained NVIDIA vGPU software from Omnissa, contact Omnissa for all support requests for NVIDIA vGPU software. Do **not** use the NVIDIA Enterprise Support Portal for NVIDIA vGPU software support.

To activate the accounts for getting NVIDIA vGPU Software, create or link to an NVIDIA Enterprise Account:

- *Creating your NVIDIA Enterprise Account*
- *Linking an Evaluation Account to an NVIDIA Enterprise Account for Purchased Licenses*

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 XenServer 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.

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 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 a suitable NVIDIA Enterprise Account, you must create an account.

For details of the web properties that you can access through your NVIDIA Enterprise Account, 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. An e-mail instructing you to log in to your account on the NVIDIA Application Hub is sent to the e-mail address you provided.
3. Open the e-mail instructing you to log in to your account and click **Log In**.
4. On the NVIDIA Application Hub Login page that opens, in the text-entry field, type the e-mail address you provided and click **Sign In**.
5. On the Create Your Account page that opens, provide and confirm a password for the account and click **Create Account**.

   A message prompting you to verify your e-mail address appears. An e-mail instructing you to verify your e-mail address is sent to the e-mail address you provided.
6. Open the e-mail instructing you to verify your e-mail address and click **Verify Email Address**.

   A message confirming that your email address is confirmed appears.

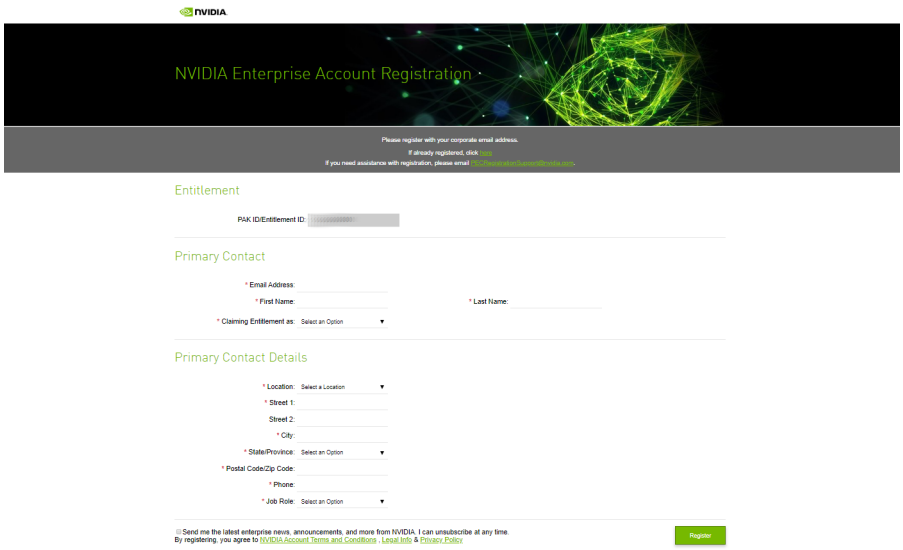
From the NVIDIA Application Hub page, you can now log in to the web properties that are listed in *Getting NVIDIA vGPU Software*.

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

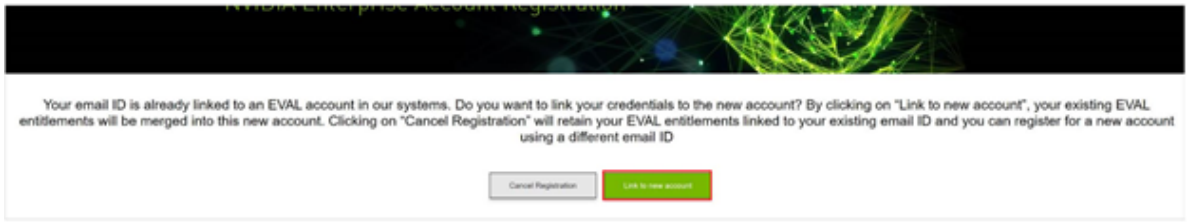
If you have an account that was created for an evaluation license, you must repeat the registration process when you receive your purchased licenses. To link your existing account for an evaluation license to the account for your purchased licenses, register for an NVIDIA Enterprise Account with the 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**.



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

   The NVIDIA Licensing Portal dashboard page opens.

   > **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 left navigation pane of the NVIDIA Licensing Portal dashboard page, click **ENTITLEMENTS** to view details of the NVIDIA vGPU software entitlements that you purchased.
5. In the left navigation pane of the NVIDIA Licensing Portal dashboard page, click **SOFTWARE DOWNLOADS**.
6. On the Software Downloads page that opens, download the software that you need for your NVIDIA vGPU software deployment.

   For an NVIDIA vGPU software deployment, you need at minimum the NVIDIA vGPU software release for your hypervisor that you want, for example, NVIDIA vGPU for vSphere 8.0.0 for NVIDIA vGPU software release 21.0.

   You might also need additional software for your deployment, for example:

   - If your system does not allow the installation of apps from the Microsoft Store, you need the standalone **NVIDIA Control Panel** installer.
   - If you are using Delegated License Service (DLS) instances to serve licenses, you need the **NLS License Server (DLS)** for the platform on which you want to deploy the DLS.

     For information about installing and configuring DLS instances, refer to NVIDIA License System User Guide.
   - If you want to use NVIDIA GPU Operator, you need the **vGPU Driver Catalog**.

   Find and download each software item that you need as follows:

   1. Filter and sort the items in the table to find the software item.
   2. In the row in the table for the software item, click **Download**.
   3. When prompted to accept the license for the software item that you are downloading, click **Agree and Download**.

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. NVIDIA License System provides role-based access control (RBAC) for DLS instances, enabling organizations to create multiple application users with different privilege levels. These application users are separate from OS-level users (dls_admin, rsu_admin) and are used solely for DLS Web UI and API access. Administrators can create local users or integrate with enterprise LDAP/Active Directory systems for centralized authentication. For more information, see User Accounts on a DLS Virtual Appliance.

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.
    1. If you are not already logged in, log in to the NVIDIA 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, 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.
3. On the Create License Server page of the wizard, step through the configuration requirements to provide the details of your license server.
    1. **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.
    2. **Step 2 – Features**: Select one or more available features from your entitlements to allot to this license server.
    3. **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.
    4. **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.

5. 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 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.
   1. In the **Name** field, enter your choice of name for the service instance.
   2. 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 the service instance that you are creating is expanded.
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.
   1. If you are not already logged in, log in to the NVIDIA 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, 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.
   1. If you are not already logged in, log in to the NVIDIA 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, 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 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 **LICENSE SERVERS** and click **LIST SERVERS**.
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.
   1. From the list of scope references, select the scope references that you want to include.

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

      Each scope reference specifies the license server that will fulfil a license request.
   2. Click the **Fulfillment class references** tab, and from the list of fulfillment class references, select the fulfillment class references that you want to include.

Including fulfillment class references is optional.

3. **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.
4. 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.

1. 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.
2. 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. 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.

5. 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.

6. Click **CONTINUE TO LOGIN**.

7. 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.
    1. If you are not already logged in, log in to the NVIDIA 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, 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.
3. On the Create License Server page of the wizard, step through the configuration requirements to provide the details of your license server.
    1. **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.
    2. **Step 2 – Features**: Select one or more available features from your entitlements to allot to this license server.
    3. **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.
    4. **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.

5. 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.

    1. If you are not already logged in, log in to the NVIDIA 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, 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 Infowindow.
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**.
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.
    1. If you are not already logged in, log in to the NVIDIA 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, 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.
    1. If you are not already logged in, log in to the NVIDIA 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, 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**.
   - If a license server has already been installed on the DLS virtual appliance, update the license server.
     1. From the **ACTIONS** menu for the license server, choose **Update server from NLP**.
     2. In the Update License Server pop-up window that opens, click **SELECT LICENSE SERVER FILE**.
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 **DASHBOARD**.
3. On the **License Server Details** page that opens, from the **Actions** menu , choose **Generate client configuration token**.
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.

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

   You must select **at least one** scope reference. Each scope reference specifies the license server that will fulfil a license request.

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

   Including fulfillment class references is optional.

3. 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.

4. **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.

5. 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 L40S	Display-off
NVIDIA L20	Display-off
NVIDIA L20 liquid cooled	Display-off
NVIDIA RTX 5000 Ada	Display enabled
NVIDIA RTX 6000 Ada	Display enabled
NVIDIA RTX A5000	Display enabled
NVIDIA RTX A5500	Display enabled
NVIDIA RTX A6000	Display enabled
NVIDIA RTX PRO 6000 Blackwell Server Edition	Display-off
NVIDIA RTX PRO 6000 Blackwell Server Edition liquid cooled	Display-off

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 GPUs listed in the table support the `displaymodeselector` tool. 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. 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.

3.2.1. 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.
   1. 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
```

   2. 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 8.0.0, `<host-driver-component>` is `NVD-VMware-x86_64-620.70.04-1OEM.800.1.0.20613240-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 8.0.0, `<gpu-management-daemon-component>` is `VMW-esx-8.0.0-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.

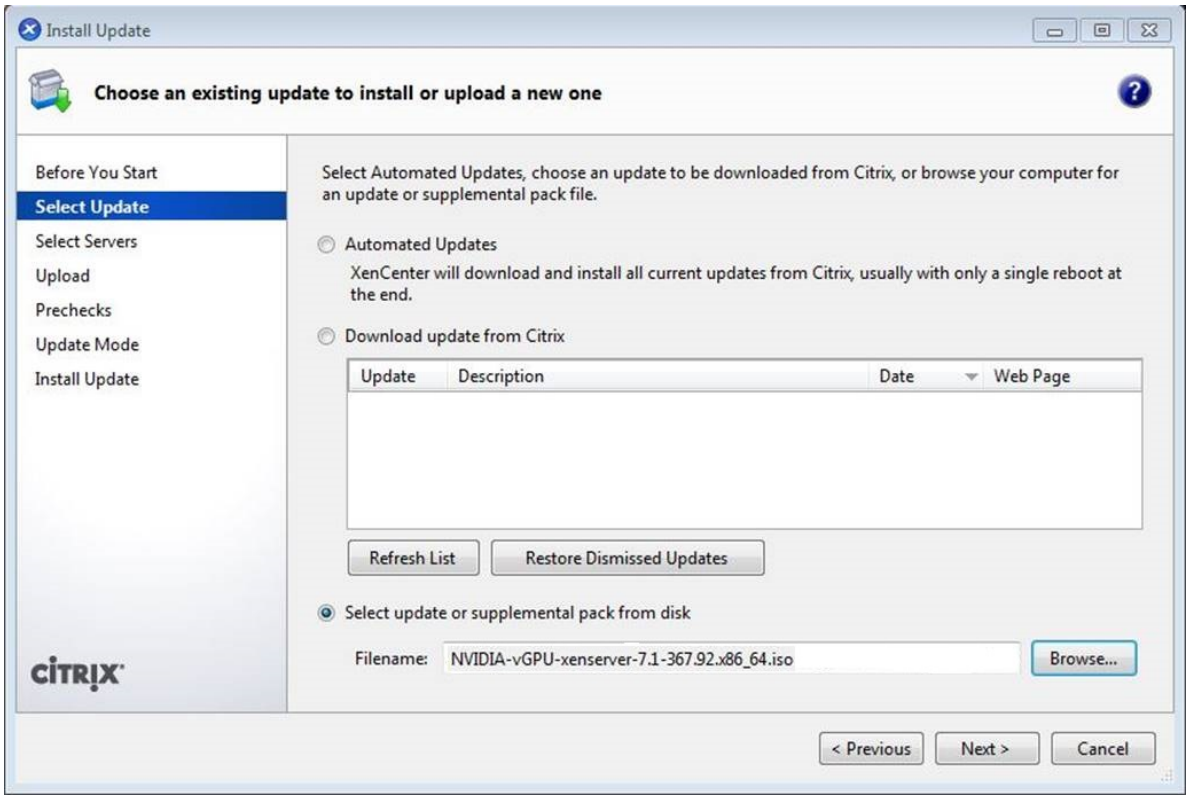
```
$ nvidia-smi
```

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

3.2.2. Installing the NVIDIA Virtual GPU Manager on XenServer

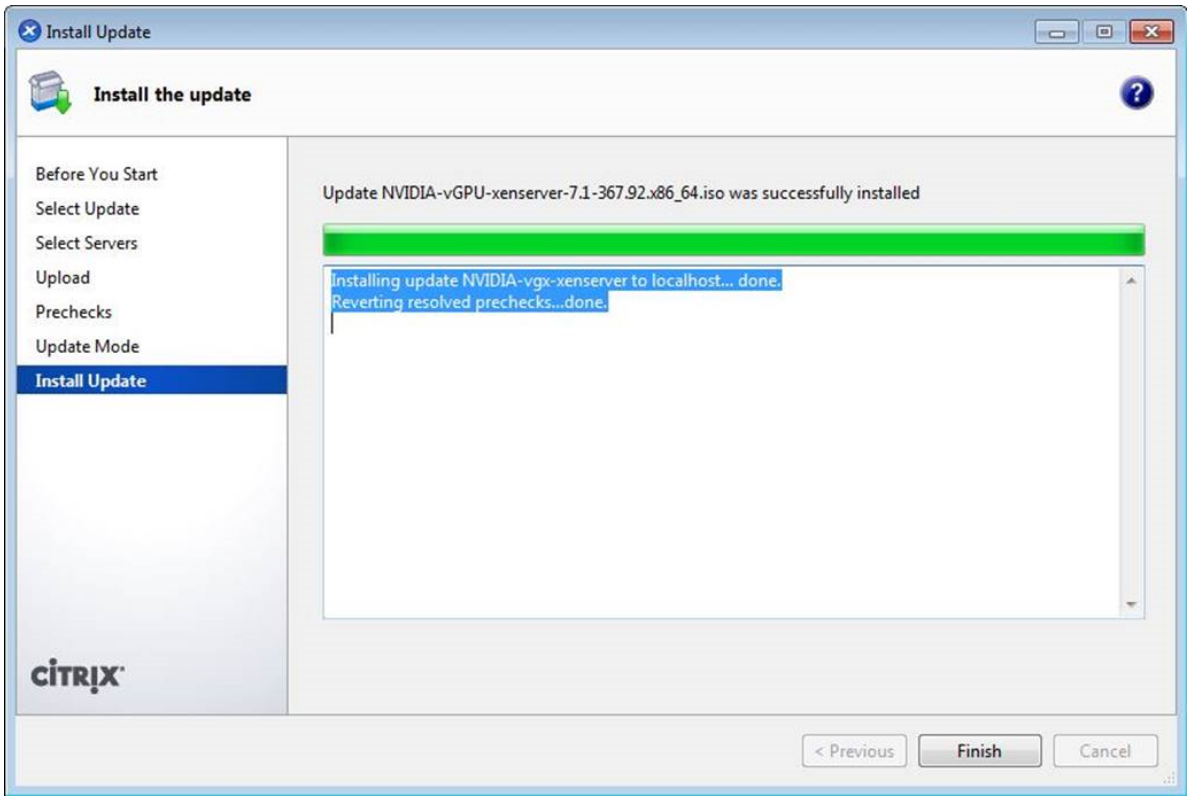
The NVIDIA Virtual GPU Manager for XenServer is distributed as a Supplemental Pack ISO.

1. In XenCenter, select **Install Update** from the **Tools** menu.
2. In the **Before You Start** section of the update wizard, click **Next**.
3. In the **Select Update** section, click **Select update or supplemental pack from disk** and open the NVIDIA Virtual GPU Manager Supplemental Pack ISO.



4. In the **Select Update** section, click **Next**.
5. In the **Select Servers** section, select all XenServer hosts on which the Supplemental Pack should be installed and click **Next**.
6. In the **Upload** section, after the Supplemental Pack has been uploaded to all selected hosts, click **Next**.
7. In the **Prechecks** section, click **Next**.

8. In the **Update Mode** section, click **Install Update**.
9. In the **Install Update** section, click **Finish**.



10. After the installation is complete, verify the installation of the NVIDIA Virtual GPU Manager package for XenServer 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_
```

3.3. 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.

3.3.1. 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 May 18 18:36:45 2026
Driver Version                      : 620.70.04

Attached GPUs                       : 1
GPU 0000:02:00.0
```

(continues on next page)

(continued from previous page)

```
[...]

    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 May 18 18:37:53 2026
Driver Version                      : 620.70.04

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

    Ecc Mode
        Current                     : Disabled
        Pending                     : Disabled

[...]
```

3.3.2. 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 May 18 18:36:45 2026
Driver Version                      : 620.70.04

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 disabled.
   - 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==============
```

(continues on next page)

(continued from previous page)

```
Timestamp                                 : Mon May 18 18:37:53 2026
Driver Version                            : 620.70.04

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

    Ecc Mode
        Current                           : Enabled
        Pending                           : Enabled

[...]
```

3.4. 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.

3.4.1. Changing the Graphics Type in VMware vSphere

After NVIDIA Virtual GPU Manager is installed, the graphics type is Shared. To enable vGPU support for VMs in VMware vSphere, you must change the graphics type to Shared Direct.

If you do not change the 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 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.

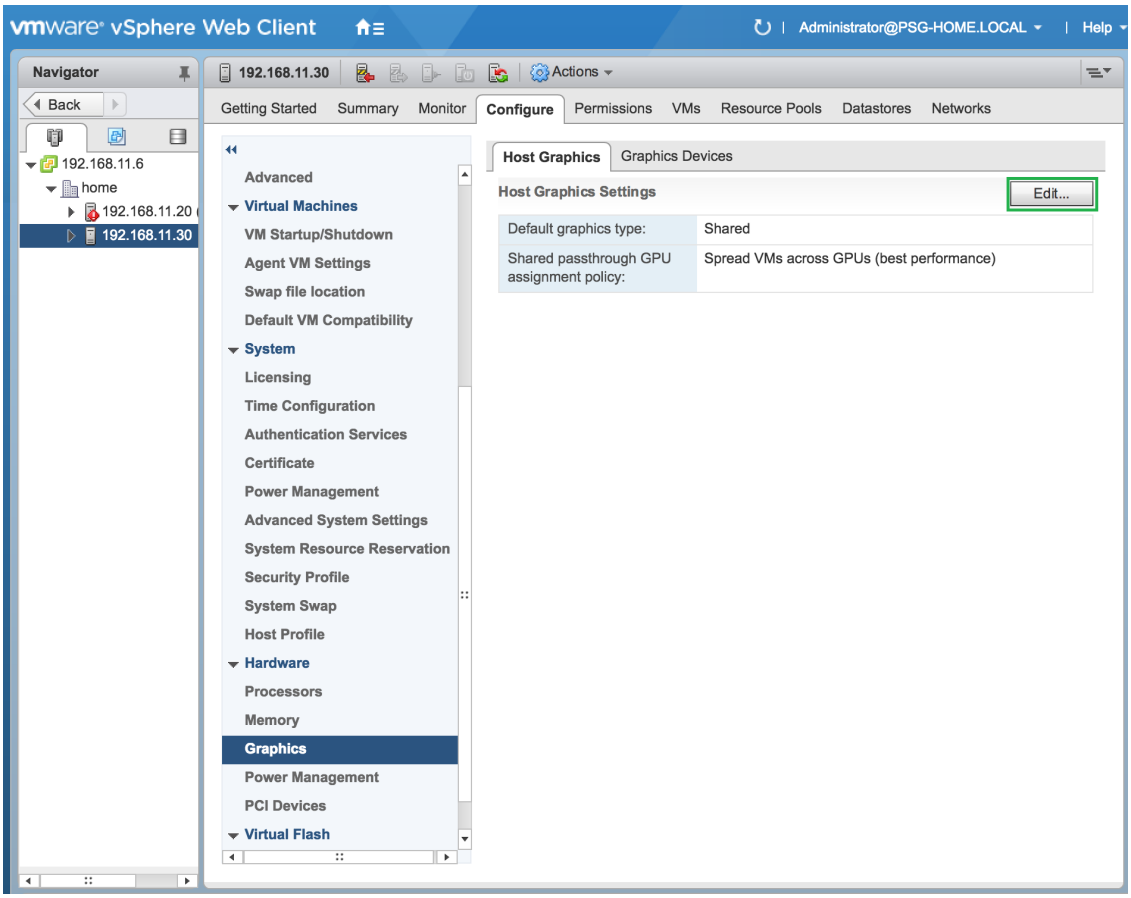
Before changing the graphics type, ensure that the ESXi host is running and that all VMs on the host are powered off.

Note

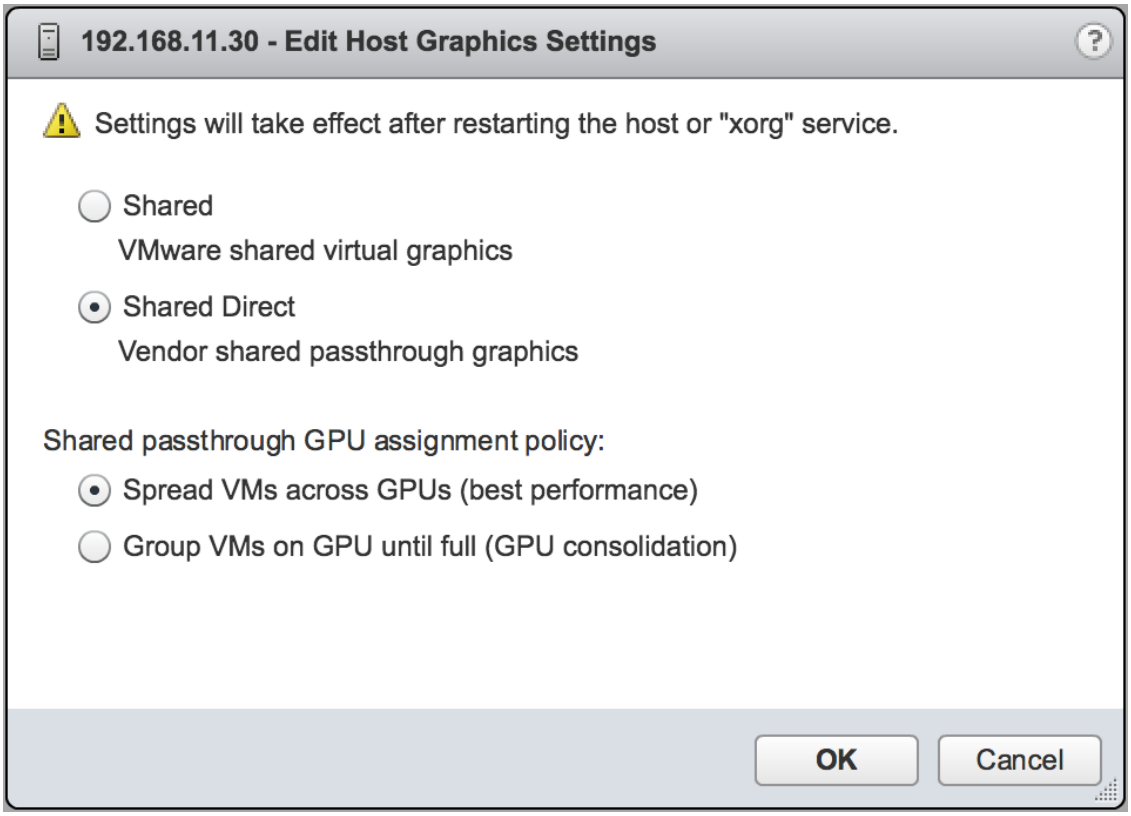
On VMware Cloud Foundation (VCF) 9 releases, you can change the graphics type only for individual physical GPUs. On earlier VMware vSphere releases, you can change the default graphics type for the host.

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**.

4. **Releases earlier than VCF 9 only:** Set the default graphics type for the host to Shared Direct.
   1. Click the **Host Graphics** tab.



   2. On the **Host Graphics** tab, click **Edit**.
   3. In the Edit Host Graphics Settings dialog box that opens, select **Shared Direct** and click **OK**.

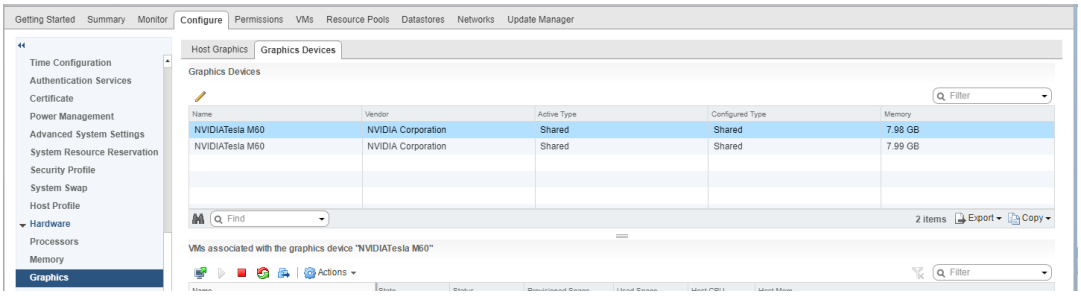


After you click OK, the default graphics type changes to Shared Direct.

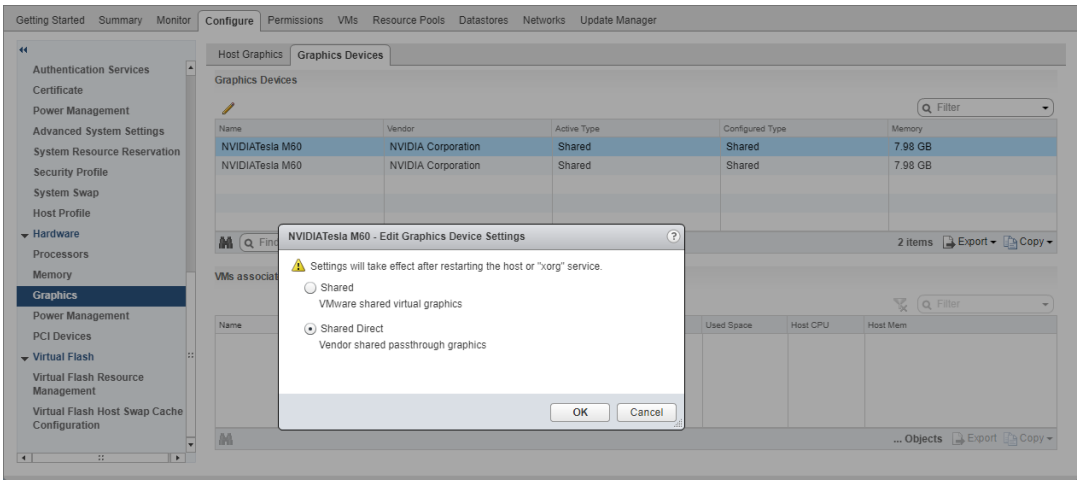
5. Verify the configured type of each physical GPU on which you want to configure vGPU.

   The configured type of each physical GPU must be Shared Direct.

   1. Open the **Graphics Devices** tab.
   2. For each physical GPU for which the configured type is Shared, change its configured type to Shared Direct.
      1. Select the physical GPU and click the **Edit icon**.



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



6. 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:

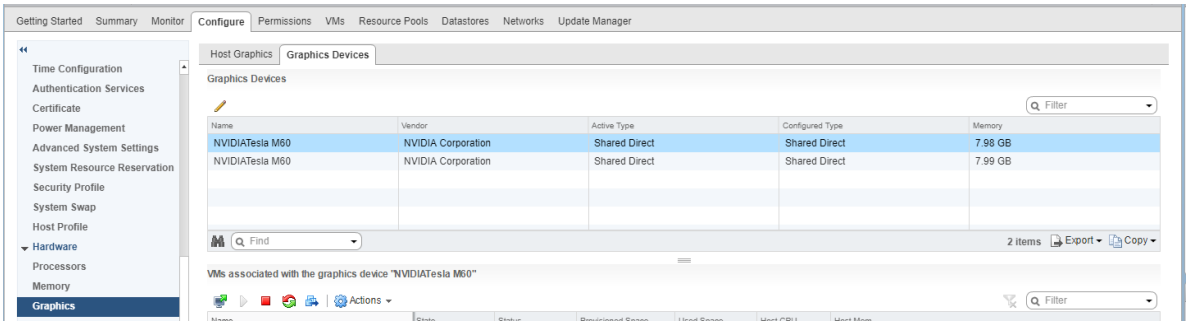
   1. Stop `nv-hostengine`.

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

   2. Wait for 1 second to allow `nv-hostengine` to stop.
   3. Start `nv-hostengine`.

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

7. 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.



3.4.2. Configuring a vSphere VM with NVIDIA vGPU

Caution

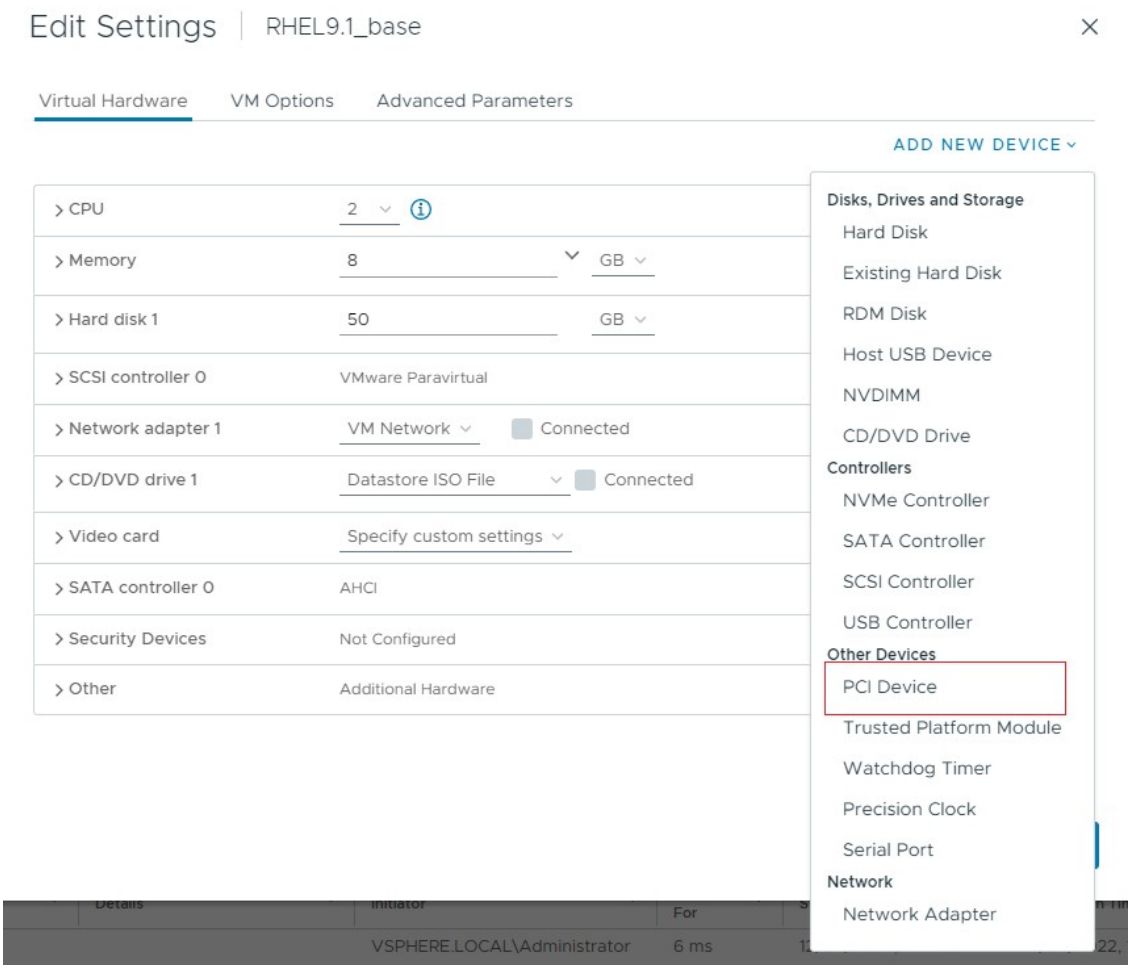
Output from the VM console in the VMware vSphere Web Client is not available for VMs that are running vGPU. Make sure that you have installed an alternate means of accessing the VM (such as Omnissa Horizon or a VNC server) before you configure vGPU.

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. In the Edit Settings window that opens, configure the vGPUs that you want to add to the VM.

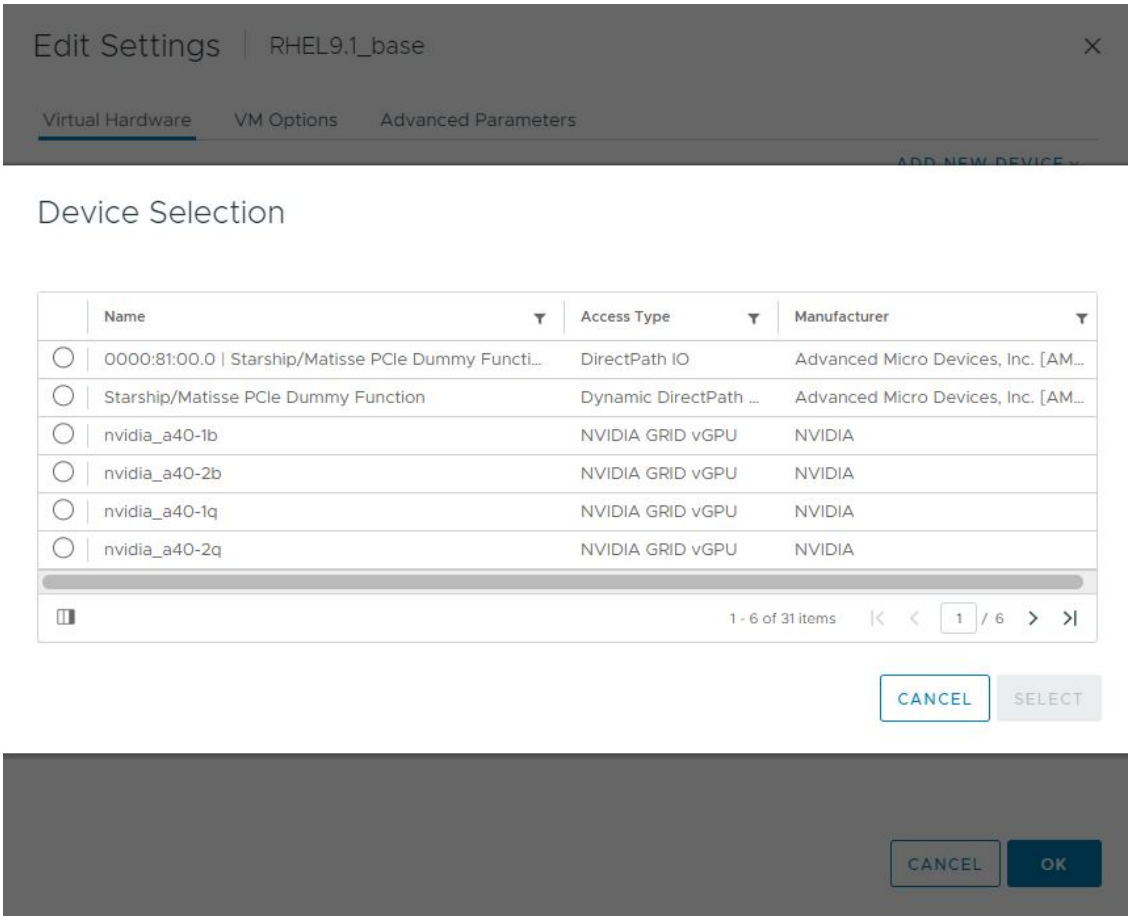
   Add each vGPU that you want to add to the VM as follows:

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



   2. 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 NVIDIA vGPU for Compute on VMware vSphere. Therefore, C-series vGPU types are not available for selection in the Device Selection window.



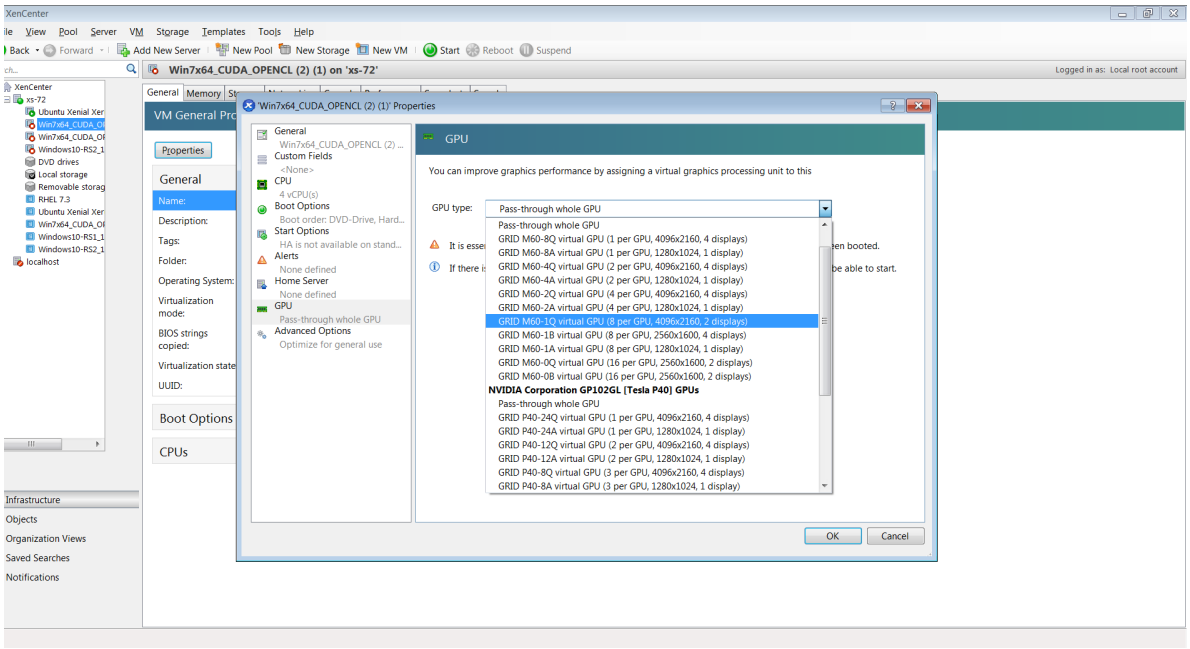
4. Back in the Edit Settings window, click **OK**.

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 Omnissa Horizon or VNC to access the VM's desktop.

3.4.3. Configuring a XenServer 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:



After you have configured a XenServer 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.

3.5. 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.

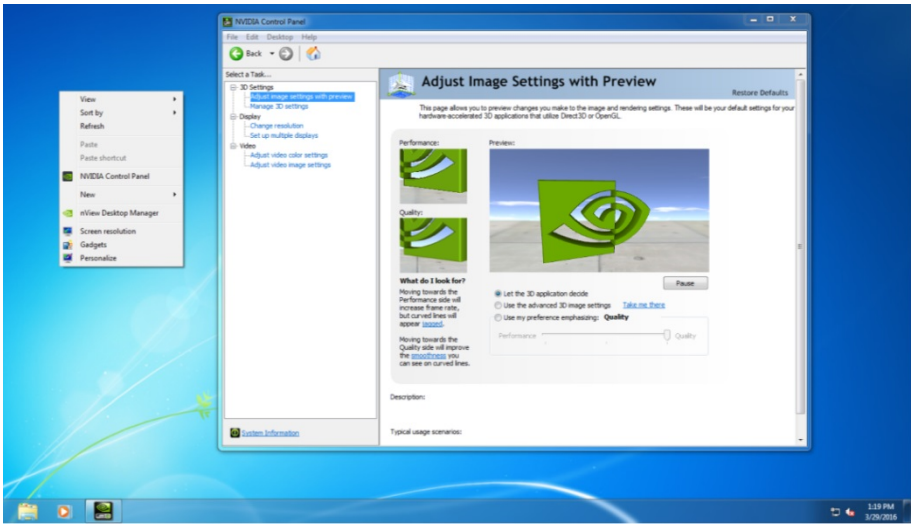


3. Click through the license agreement.
4. Select **Express Installation** and click **NEXT**.

   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.
   1. Right-click on the desktop.
   2. From the menu that opens, choose **NVIDIA Control Panel**.
   3. 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.



3.6. 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 NVIDIA vGPU software 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.

3.6.1. 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.

3.6.2. 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` option from 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)
    …
    …
```

Chapter 4. Notices

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

©2013-2026, NVIDIA Corporation