GPU Manager for VMware vCenter

User Guide

Table of Contents

Chapter 1. Introduction to the NVIDIA GPU Manager for VMware vCenter

NVIDIA GPU Manager for VMware vCenter enables you to manage NVIDIA GPUs from the vSphere client of VMware vCenter Server.

When NVIDIA GPU Manager for VMware vCenter is installed and configured, you can download and install NVIDIA GPU drivers by using a web-based tool integrated with the vSphere web client user interface instead of running a set of commands on the hypervisor. NVIDIA GPU Manager for VMware vCenter extends the capabilities of VMware vCenter Server by providing additional features and functions to manage NVIDIA GPUs for VMware vCenter clusters, VMware vSphere Hypervisor (ESXi) hosts, and guest VMs.

NVIDIA GPU Manager for VMware vCenter consists of a virtual appliance and a vSphere client plugin.

▶ **The NVIDIA GPU Manager for VMware vCenter virtual appliance** contains the following components:

  ▶ **NVIDIA GPU Manager** application, which is used for registering NVIDIA GPU Manager for VMware vCenter

  ▶ **NVIDIA GPU Manager drivers repository**, which is used for storing downloaded drivers

  ▶ Web-based management interface, which is used for accessing the **NVIDIA GPU Manager** application on the appliance

▶ **The vSphere client plugin** contains the user interface resources, which are copied from the virtual appliance when NVIDIA GPU Manager for VMware vCenter is registered. These resources are deployed on the vSphere client after registration.

Chapter 2. Installing and Configuring the NVIDIA GPU Manager for VMware vCenter Virtual Appliance

To simplify the installation and administration of the NVIDIA GPU Manager for VMware vCenter, it is distributed as a virtual appliance image to be installed on a supported release of VMware vSphere. This appliance is a secure, hardened environment in which access to the application software is strictly controlled.

Each NVIDIA GPU Manager for VMware vCenter appliance is configured with a single user account. You cannot modify this user account or add other user accounts to the appliance.

2.1. Platform Requirements for the NVIDIA GPU Manager for VMware vCenter Virtual Appliance

Before proceeding, ensure that you have a platform suitable for hosting the NVIDIA GPU Manager for VMware vCenter virtual appliance.

▶ The hosting platform must be a physical host running a supported release of VMware vSphere.

▶ The minimum resource requirements for the VM in which the NVIDIA GPU Manager for VMware vCenter virtual appliance will run are as follows:

  ▶ **Number of vCPUs:** 2

  ▶ **RAM:** 4 Gbytes

  ▶ **Disk Size:** 35 Gbytes

▶ The platform must have a fixed (unchanging) IP address. The IP address may be assigned dynamically by DHCP or statically configured, but must be constant.

▶ The platform's date and time must be set accurately. NTP is recommended. For information about how to configure NTP on an ESXi host, refer to VMware Knowledge Base Article: Configuring Network Time Protocol (NTP) on an ESXi host using the vSphere Client (57147).

> 💬 **Note:** Before proceeding with the installation, refer to *GPU Manager for VMware vCenter Release Notes* for details of supported VMware vSphere releases, web browsers, and known issues.

2.2. Host Name Resolution Requirements for an NVIDIA GPU Manager for VMware vCenter Virtual Appliance

The platform that hosts an NVIDIA GPU Manager for VMware vCenter virtual appliance must be identified by its IP address or its fully qualified domain name. If you want to identify the platform by its fully qualified domain name, ensure that the required DNS entries are set before installing the virtual appliance. If you want to identify the platform by its default host name, you must set a DNS entry that maps the default host name to the fully qualified domain name.

The process for setting these DNS entries is separate from the process for installing the virtual appliance. Use the standard interfaces of the name resolution service that you are using to set the required DNS entries.

For each mapping between a domain name and an IP address, ensure that you set both the forward pointer and reverse pointer DNS entries.

> 💬 **Note:** For the reverse pointer DNS entry on a Windows DNS server, use all lowercase for the fully qualified domain name.

To determine whether the forward pointer and reverse pointer DNS entries have been set correctly, type the following commands in a command or shell window on any host on the same network as the NVIDIA GPU Manager for VMware vCenter virtual appliance:

▶ For the forward pointer entry, type:

```
$ nslookup domain-name
```
domain-name
> The domain name for which you want to determine whether the forward pointer DNS entry is correct.

If the DNS entry has been set correctly, the command displays the IP address that is mapped to the domain name.

▶ For the reverse pointer entry, type:

```
$ nslookup ip-address
```

ip-address
> The IP address for which you want to determine whether the reverse pointer DNS entry is correct.

If the DNS entry has been set correctly, the command displays the domain name that is mapped to the IP address.

2.3.   Firewall Rules for NVIDIA GPU Manager for VMware vCenter

To enable communication through a firewall, firewall rules that allow traffic on specific URLs must be defined. To set these rules, use the standard interfaces of the firewall that you are using, for example Zscaler Firewall.

To enable communication between NVIDIA GPU Manager for VMware vCenter and NVIDIA Licensing Portal through a firewall, firewall rules that allow traffic on the URLs in the following table must be defined.

URL	Traffic
`api.licensing.nvidia.com`	NVIDIA vGPU software graphics driver metadata from the NVIDIA Licensing Portal
`griddownloads.nvidia.com`	NVIDIA vGPU software graphics driver binary files from NVIDIA Licensing Portal

2.4.   Installing the NVIDIA GPU Manager for VMware vCenter Virtual Appliance

The NVIDIA GPU Manager for VMware vCenter image is distributed as a ZIP archive that contains an Open Virtual Appliance (OVA) file. You use standard interfaces of VMware vSphere to install the virtual appliance.

The virtual appliance image specifies the minimum configuration for the VM as listed in Platform Requirements for the NVIDIA GPU Manager for VMware vCenter Virtual Appliance. You are not required to specify the VM configuration when you install the virtual appliance. After installing the NVIDIA GPU Manager for VMware vCenter virtual appliance, you can use standard interfaces of the hypervisor to change the configuration of the VM if necessary.

Use the **VMware vSphere Client** to perform this task on the ESXi server on which you want to run the NVIDIA GPU Manager for VMware vCenter virtual appliance.

For additional information, refer to the following topics on the VMware Docs site:

▶ Log in to vCenter Server by Using the vSphere Web Client

▶ <u>Deploy an OVF or OVA Template</u>

1. Download the ZIP archive that contains the OVA file that contains the NVIDIA GPU Manager for VMware vCenter image.
2. Extract the contents of the ZIP archive that you downloaded.
3. Log in to vCenter Server by using the **VMware vSphere Client**.
4. From the **VMware vSphere Client Actions** menu, choose **Deploy OVF Template**.
5. Select the **Local file** option, browse for and select the downloaded OVA file, and click **Next**.
6. Enter the your choice of virtual machine name, select a location for the virtual machine, and click **Next**.
7. Select a compute resource where the virtual machine will be created and click **Next**.
8. Review the details of the template that you are deploying and click **Next**.
9. Accept the NVIDIA Software License Agreement and click **Next**.
10. Select the storage for the virtual appliance configuration and disk files and click **Next**.
11. Leave the network properties blank to obtain network settings from a DHCP server or provide the network settings for a static IP address, and click **Next**.

> 🗨 **Note:** You can change these settings after the virtual appliance is installed as explained in <u>Updating the Network Settings for the NVIDIA GPU Manager for VMware vCenter Virtual Appliance</u>.

Provide the network settings for a static IP address as follows.

a). In the **IPv4 Address** text-entry field, type the IP address that you want to assign to the virtual appliance.

This address must be an unused IP address in the subnet range of the VM's virtual network.

b). In the **Netmask Prefix** text-entry field, type the subnet mask of the virtual appliance's network in classless inter-domain routing (CIDR) format **without the leading slash character (/)**.

To get a subnet mask in CIDR format from its decimal equivalent, refer to the table on page 2 of <u>IETF RFC 1878: Variable Length Subnet Table For IPv4</u>.

For example, the subnet mask in CIDR format of the decimal equivalent 255.255.255.0 is 24.

c). In the **Gateway** text-entry field, type the IP address of the virtual appliance's default gateway.

d). In the **DNS Server 1** text-entry field, type the IP address of the first DNS server to be used for name resolution.

e). **Optional:** In the **DNS Server 2** text-entry field, type the IP address of the second DNS server to be used for name resolution.

> 🗨 **Note:** If any field **except DNS Server 2** is blank or if any field contains invalid data, a static IP address is **not** assigned to the VM. Instead, the VM obtains network settings from a DHCP server.

12.Review all the details of the virtual machine that you are creating and click **Finish**.

13.Start the VM that you created.

2.5. Registering the NVIDIA GPU Manager for VMware vCenter Administrator User

Each NVIDIA GPU Manager for VMware vCenter virtual appliance is configured with a user account specifically for administering the NVIDIA GPU Manager for VMware vCenter. This account provides access through a web-based management interface to the **NVIDIA GPU Manager** application on the appliance. Before administering the NVIDIA GPU Manager for VMware vCenter 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://vcenter-gpu-mgr-vm-ip-address`.

    vcenter-gpu-mgr-vm-ip-address
    The IP address or, if defined, the fully qualified domain name of the VM on which the NVIDIA GPU Manager for VMware vCenter virtual appliance is installed.

    You can get the IP address from the management console of your hypervisor.

2. On the **Register User Account** page that opens, provide and confirm the password for the NVIDIA GPU Manager for VMware vCenter administrator user `vcp_admin`.

    > **Note:** If the NVIDIA GPU Manager for VMware vCenter administrator user has already been registered, the login page opens instead of the **Register User Account** page.

    The password must be at least eight characters long, and must contain at least one uppercase letter, one lowercase letter, one number, and one of the special characters `~`, `#`, `?`, `!`, `@`, `$`, `%`, `^`, `&`, `*`, `-`,`+`, `=`, and `_` .

    > **Note:** You can change the NVIDIA GPU Manager for VMware vCenter administrator password at any time after this user is registered. For instructions, refer to Changing the NVIDIA GPU Manager for VMware vCenter Administrator Password.

3. Click **Register**.
   The **Register User Account** 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.

4. 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.

    > **Note:** You will need this key to reset a forgotten NVIDIA GPU Manager for VMware vCenter administrator user's password. If you lose this secret, you will **not** be able to reset a forgotten NVIDIA GPU Manager for VMware vCenter administrator user's password.

5. Click **Continue to Login**.

6. On the login page that opens, provide the password that you set for the `vcp_admin` user, and click **Login**.

2.6. Changing the NVIDIA GPU Manager for VMware vCenter Administrator Password

For enhanced security, change the password for the NVIDIA GPU Manager for VMware vCenter administrator user periodically after this user is registered.

1. Open a web browser and connect to the URL `https://vcenter-gpu-mgr-vm-ip-address`.

   vcenter-gpu-mgr-vm-ip-address
   > The IP address or, if defined, the fully qualified domain name of the VM on which the NVIDIA GPU Manager for VMware vCenter virtual appliance is installed.
   >
   > You can get the IP address from the management console of your hypervisor.

2. On the login page that opens, provide the user credentials for the NVIDIA GPU Manager for VMware vCenter administrator user on the NVIDIA GPU Manager for VMware vCenter virtual appliance and click **Login**.

3. At the top right of the page in the **NVIDIA GPU Manager** application on the NVIDIA GPU Manager for VMware vCenter appliance that opens, click **Change Password**.

4. On the **Change Password** page that opens, enter the current password, provide a new password , and confirm the new password for the NVIDIA GPU Manager for VMware vCenter administrator user.

   a). In the **Current Password** text-entry field, type the current password for the NVIDIA GPU Manager for VMware vCenter administrator user.

   b). In the **New Password** text-entry field, type the new password that you want for the NVIDIA GPU Manager for VMware vCenter administrator user.

   The password must be at least eight characters long, and must contain at least one uppercase letter, one lowercase letter, one number, and one of the special characters **~**, **#**, **?**, **!**, **@**, **$**, **%**, **^**, **&**, *****, **−**,**+**, **=**, and **_** .

   c). In the **Confirm New Password** text-entry field, type the password that you typed in the **New Password** text-entry field.

5. Click **Change Password**.

2.7. Resetting a Forgotten NVIDIA GPU Manager for VMware vCenter Administrator Password

If you have forgotten the NVIDIA GPU Manager for VMware vCenter administrator password, you can use its local reset secret to reset it.

1. Open a web browser and connect to the URL `https://vcenter-gpu-mgr-vm-ip-address`.

   vcenter-gpu-mgr-vm-ip-address
   > The IP address or, if defined, the fully qualified domain name of the VM on which the NVIDIA GPU Manager for VMware vCenter virtual appliance is installed.
   >
   > You can get the IP address from the management console of your hypervisor.

2. On the login page that opens, follow the **Forgot Password?** link.

3. In the **Reset Password** window that opens, enter the local reset secret, provide a new password , and confirm the new password for the NVIDIA GPU Manager for VMware vCenter administrator user.

   a). In the **Local Reset Secret** text-entry field, type the local reset secret that was created when the NVIDIA GPU Manager for VMware vCenter administrator user was registered.

   b). In the **Password** text-entry field, type the new password that you want for the NVIDIA GPU Manager for VMware vCenter administrator user.

   The password must be at least eight characters long, and must contain at least one uppercase letter, one lowercase letter, one number, and one of the special characters **~**, **#**, **?**, **!**, **@**, **$**, **%**, **^**, **&**, *****, **−**,**+**, **=**, and **_** .

   c). In the **Confirm Password** text-entry field, type the password that you typed in the **Password** text-entry field.

4. Click **Reset Password**.

2.8. Logging in to the NVIDIA GPU Manager for VMware vCenter Virtual Appliance

Each NVIDIA GPU Manager for VMware vCenter virtual appliance is configured with a user account specifically for administering NVIDIA GPU Manager for VMware vCenter. This

account provides access through a web-based management interface to the **NVIDIA GPU Manager** application on the appliance.

Ensure that the NVIDIA GPU Manager for VMware vCenter administrator user has been registered for the appliance as explained in Registering the NVIDIA GPU Manager for VMware vCenter Administrator User.

1. Open a web browser and connect to the URL `https://vcenter-gpu-mgr-vm-ip-address`.

   vcenter-gpu-mgr-vm-ip-address
   > The IP address or, if defined, the fully qualified domain name of the VM on which the NVIDIA GPU Manager for VMware vCenter virtual appliance is installed.

   > You can get the IP address from the management console of your hypervisor.

2. On the login page that opens, provide the user credentials for the NVIDIA GPU Manager for VMware vCenter administrator user on the NVIDIA GPU Manager for VMware vCenter virtual appliance and click **Login**.

2.9. Configuring an NVIDIA GPU Manager for VMware vCenter Virtual Appliance with a Third-Party Signed SSL Certificate

By default, an NVIDIA GPU Manager for VMware vCenter virtual appliance is configured with a self-signed SSL certificate that is generated when the virtual appliance is started. If necessary, you can replace the self-signed certificate with an SSL certificate that is signed by a third party, such as a certificate authority (CA).

When **NVIDIA GPU Manager** is registered with **VMware vCenter Server**, the SSL certificate with which the virtual appliance is configured is imported into **VMware vCenter Server**. Therefore, try to configure the virtual appliance with a new SSL certificate before registering **NVIDIA GPU Manager** with **VMware vCenter Server**.

Otherwise, you must confirm that you can use vSphere Lifecycle Management (vLCM) to synchronize the new SSL certificate on all **VMware vCenter Server** instances with which **NVIDIA GPU Manager** is registered. This confirmation is required because **VMware vCenter Server** must be able to trust the new certificate before it allows any communication that is authenticated by the certificate.

To configure an NVIDIA GPU Manager for VMware vCenter virtual appliance with a third-party signed SSL certificate, follow this sequence of instructions:

1. Obtaining a Third-Party Signed SSL Certificate for an NVIDIA GPU Manager for VMware vCenter Virtual Appliance
2. Installing a Third-Party Signed SSL Certificate on an NVIDIA GPU Manager for VMware vCenter Virtual Appliance

2.9.1. Obtaining a Third-Party Signed SSL Certificate for an NVIDIA GPU Manager for VMware vCenter Virtual Appliance

Obtain a third-party signed SSL certificate by submitting a certificate signing request (CSR) to a suitable third party, such as a certificate authority (CA).

For each certificate that you require, submit a certificate signing request (CSR) to a CA.

Ensure that each certificate that you request meets these requirements:

▶ The certificate must be a PEM text file (**not** in Java keystore format) and secured with a private key.

▶ The certificate and the private key must be in separate files.

▶ To ensure that web browsers trust the domain, the domain name must be part of the Subject Alternate Name (SAN) attribute, **not** the Common Name (CN) attribute of the CSR.

▶ The SAN attribute of the CSR must specify the fully qualified domain name of any NVIDIA GPU Manager for VMware vCenter virtual appliance that will be configured with the certificate.

   If a fully qualified domain name is not assigned to the NVIDIA GPU Manager for VMware vCenter, specify its IP address as the DNS in the CSR.

▶ The certificate must use RSA, DSA, and DH keys that are at least 2048 bits long.

▶ If the certificate chain of trust includes intermediate certificates, the following conditions apply:

   ▶ The certificate must be bundled with the intermediate certificates in the following order:

      1. Domain name certificate
      2. Intermediate certificates
      3. Root certificate

   ▶ The `keyUsage` extension must have `Certificate Sign` usage for all the certificates in the chain.

If necessary, contact the CA that will provide your certificate for information about how to request a certificate that meets these requirements or convert an existing certificate to meet these requirements.

2.9.2. Installing a Third-Party Signed SSL Certificate on an NVIDIA GPU Manager for VMware vCenter Virtual Appliance

Ensure that you have obtained the SSL certificate that you are installing and its private key file.

1. Log in to the NVIDIA GPU Manager for VMware vCenter virtual appliance on which you are installing the SSL certificate.
2. In the left navigation pane, click **SETTINGS**.
3. On the **SETTINGS** page that opens, in the **SSL Configuration** section, click **Update**.
4. In the **Update SSL Configuration** window that opens, specify the SSL certificate that you are installing and its private key file.
    a). Click **Choose File** adjacent to **Certificate** and in the file browser that opens, navigate to the folder that contains the SSL certificate and select the file.
    b). Click **Choose File** adjacent to **Private Key** and in the file browser that opens, navigate to the folder that contains the SSL certificate's private key and select the file.
5. Click **Update**.

The **NVIDIA GPU Manager** attempts to import the new certificate into any **VMware vCenter Server** instances which **NVIDIA GPU Manager** is registered.
Confirm that the NVIDIA GPU Manager for VMware vCenter virtual appliance has been configured with the SSL certificate.

1. From the **vSphere Client** menu, choose **Lifecycle Manager**, click the **Settings** tab, choose **Administration** > **Patch Setup** , and confirm that the download source is set correctly.
2. Confirm that the **Sync updates** operation in vSphere Lifecycle Management (vLCM) succeeds.

2.10. Updating the Network Settings for the NVIDIA GPU Manager for VMware vCenter Virtual Appliance

If necessary, you can change the network settings for an NVIDIA GPU Manager for VMware vCenter virtual appliance any time after the virtual appliance has been installed.

Try to update the network settings of the virtual appliance before registering **NVIDIA GPU Manager** with **VMware vCenter Server**.

Otherwise, you must confirm that all **VMware vCenter Server** instances with which **NVIDIA GPU Manager** is registered reflect the updated network settings. This confirmation is required because the **NVIDIA GPU Manager** plugin and driver repository on **VMware vCenter Server** are linked to the IP address or fully qualified domain name of the NVIDIA GPU Manager for VMware vCenter virtual appliance.

1. Log in to the NVIDIA GPU Manager for VMware vCenter virtual appliance for which you are updating network settings.
2. In the left navigation pane, click **SETTINGS**.
3. On the **SETTINGS** page that opens, in the **Network Settings** section, click **Update**.
4. In the **Update Network Settings** window that opens, select **DHCP** to obtain network settings from a DHCP server or provide the network settings for a static IP address.

   Provide the network settings for a static IP address as follows.

   a). Ensure that the **Static** option is selected.

   b). In the **IP Address** text-entry field, type the IP address that you want to assign to the virtual appliance.

      This address must be an unused IP address in the subnet range of the VM's virtual network.

   c). In the **Netmask Prefix** text-entry field, type the subnet mask of the virtual appliance's network in classless inter-domain routing (CIDR) format **without the leading slash character (/)**.

      To get a subnet mask in CIDR format from its decimal equivalent, refer to the table on page 2 of [IETF RFC 1878: Variable Length Subnet Table For IPv4](#).

      For example, the subnet mask in CIDR format of the decimal equivalent 255.255.255.0 is 24.

   d). In the **Gateway** text-entry field, type the IP address of the virtual appliance's default gateway.

   e). **Optional:** In the **DNS Server 1** text-entry field, type the IP address of the first DNS server to be used for name resolution.

   f). **Optional:** In the **DNS Server 2** text-entry field, type the IP address of the second DNS server to be used for name resolution.

5. Click **Update**.
   Your browser will be disconnected from the instance after the update is complete.
6. Log in to the NVIDIA GPU Manager for VMware vCenter virtual appliance again by connecting to the URL `https://vcenter-gpu-mgr-vm-static-ip-address`.

   vcenter-gpu-mgr-vm-static-ip-address
      The static IP address that you set for the NVIDIA GPU Manager for VMware vCenter virtual appliance.

Confirm that the network settings have been updated as required.

1. Confirm that the vSphere client shows the correct IP address or fully qualified domain name of the NVIDIA GPU Manager for VMware vCenter.
2. Confirm that the **Sync updates** operation in vSphere Lifecycle Management (vLCM) succeeds.

2.11. Accessing the NVIDIA GPU Manager for VMware vCenter Virtual Appliance from the vCenter Web Console

If the NVIDIA GPU Manager for VMware vCenter virtual appliance network is incorrectly configured, you might not be able to access the web-based management interface to the **NVIDIA GPU Manager** application on the appliance. In this situation, you can reconfigure network settings and download log files for troubleshooting through the web console option of the vCenter web client. The virtual appliance is configured with a user account specifically for this purpose. This user account has access only to the directories and scripts necessary for reconfiguring network settings and downloading log files.

2.11.1. Downloading Log Files from an NVIDIA GPU Manager for VMware vCenter Virtual Appliance from the Hypervisor

Ensure that the secure shell (SSH) service is running and port 22 is open on the remote host to which you want to download the log files.

1. Use the web console option of the vCenter web client to log in as the user **`vcp_cli_user`** to the VM that hosts the NVIDIA GPU Manager for VMware vCenter virtual appliance.

   The first time you log in as **`vcp_cli_user`**, you are prompted to set a password for this user.

2. Run the `generate-log-bundle.sh` script with `sudo` user privileges.

   ```
   $ sudo ./generate-log-bundle.sh
   ```

   A ZIP archive named `log_bundle.zip` that contains the log files is created on the virtual appliance.

3. When prompted, enter the details of the remote host to which you want to transfer the `log_bundle.zip` file that was created in the previous step.

   a). Enter the user name of the user that will log in to the remote host, for example, **`vcp_root`**.

   b). Enter the IP address of the remote host.

   > **Note:** The IP address must be an IPv4 address. IPv6 addresses are **not** supported.

   c). Enter the full path to the directory on the remote to which you want the `log_bundle.zip` file to be copied, for example, **`/home/vcp_root`**.

   d). Enter the password of the user that you specified previously will log in to the remote host.

2.11.2. Assigning the IP Address of an NVIDIA GPU Manager for VMware vCenter Virtual Appliance by Using DHCP from the Hypervisor

1. Use the web console option of the vCenter web client to log in as the user **vcp_cli_user** to the VM that hosts the NVIDIA GPU Manager for VMware vCenter virtual appliance.

    The first time you log in as **vcp_cli_user**, you are prompted to set a password for this user.

2. Run the `set-dhcp-cli.sh` script with `sudo` user privileges.

    ```
    $ sudo ./set-dhcp-cli.sh
    ```

Confirm that the network settings have been updated as required.

1. Confirm that the vSphere client shows the correct IP address or fully qualified domain name of the NVIDIA GPU Manager for VMware vCenter.

2. Confirm that the **Sync updates** operation in vSphere Lifecycle Management (vLCM) succeeds.

2.11.3. Setting the IP Address of an NVIDIA GPU Manager for VMware vCenter Virtual Appliance from the Hypervisor

1. Use the web console option of the vCenter web client to log in as the user **vcp_cli_user** to the VM that hosts the NVIDIA GPU Manager for VMware vCenter virtual appliance.

    The first time you log in as **vcp_cli_user**, you are prompted to set a password for this user.

2. Run the `set-static-ip-cli.sh` script with `sudo` user privileges.

    ```
    $ sudo ./set-static-ip-cli.sh
    ```

3. When prompted, enter the details of the IP address.

> **Note:** All IP addresses that you enter must be IPv4 addresses. IPv6 addresses are **not** supported.

   a). Enter the IP address that you want to assign to the NVIDIA GPU Manager for VMware vCenter virtual appliance.

   b). Enter the IP address of the NVIDIA GPU Manager for VMware vCenter virtual appliance's default gateway.

   c). Enter the IP address of the first DNS server to be used for name resolution.

d). Enter the IP address of the second DNS server to be used for name resolution.

e). Enter the subnet mask of the NVIDIA GPU Manager for VMware vCenter virtual appliance's network in classless inter-domain routing (CIDR) format **without the leading slash character (/)**.

To get a subnet mask in CIDR format from its decimal equivalent, refer to the table on page 2 of IETF RFC 1878: Variable Length Subnet Table For IPv4.

For example, the subnet mask in CIDR format of the decimal equivalent 255.255.255.0 is 24.

Confirm that the network settings have been updated as required.

1. Confirm that the vSphere client shows the correct IP address or fully qualified domain name of the NVIDIA GPU Manager for VMware vCenter.

2. Confirm that the **Sync updates** operation in vSphere Lifecycle Management (vLCM) succeeds.

2.11.4. Setting the Host Name of an NVIDIA GPU Manager for VMware vCenter Virtual Appliance

1. Use the web console option of the vCenter web client to log in as the user **vcp_cli_user** to the VM that hosts the NVIDIA GPU Manager for VMware vCenter virtual appliance.

   The first time you log in as **vcp_cli_user**, you are prompted to set a password for this user.

2. Run the `set-hostname-cli.sh` script with `sudo` user privileges.

   ```
   $ sudo ./set-hostname-cli.sh
   ```

3. When prompted, enter your choice of new host name for the virtual appliance.

2.11.5. Since 1.2.0: Resetting the SSL Certificate on an NVIDIA GPU Manager for VMware vCenter Virtual Appliance

1. Use the web console option of the vCenter web client to log in as the user **vcp_cli_user** to the VM that hosts the NVIDIA GPU Manager for VMware vCenter virtual appliance.

   The first time you log in as **vcp_cli_user**, you are prompted to set a password for this user.

2. Run the `reset-ssl-cert.sh` script with `sudo` user privileges.

   ```
   $ sudo ./reset-ssl-cert.sh
   ```

Confirm that the NVIDIA GPU Manager for VMware vCenter virtual appliance has been configured with the SSL certificate.

1. From the **vSphere Client** menu, choose **Lifecycle Manager**, click the **Settings** tab, choose **Administration** > **Patch Setup** , and confirm that the download source is set correctly.

2. Confirm that the **Sync updates** operation in vSphere Lifecycle Management (vLCM) succeeds.

2.12. Since 1.2.0: Configuring NVIDIA GPU Manager for VMware vCenter with a Proxy Server

NVIDIA GPU Manager for VMware vCenter supports communication through the Squid proxy server with entities such as NVIDIA Licensing Portal and vCenter Server instances.

NVIDIA GPU Manager for VMware vCenter supports transparent proxy servers and non-transparent proxy servers.

▶ **A transparent proxy server** identifies itself to the server and does **not** modify client requests and responses.

▶ **A non-transparent proxy server** does not reveal the IP address of the client and modifies client requests and responses.

Both authenticated and unauthenticated proxy servers are supported. For authenticated proxy servers, the following authentication methods are supported:

▶ Basic

▶ Kerberos

Since 1.5.0: You can specify hosts for which a configured proxy server is bypassed and disable and enable a configured proxy server.

2.12.1. Verifying the Correct Operation of the Proxy Server

Before configuring NVIDIA GPU Manager for VMware vCenter with a proxy server, ensure that the proxy server is reachable, the port on which the proxy server listens for client requests is open, and the proxy service is running on it.

1. Use the web console option of the vCenter web client to log in as the user `vcp_cli_user` to the VM that hosts the NVIDIA GPU Manager for VMware vCenter virtual appliance.

   The first time you log in as `vcp_cli_user`, you are prompted to set a password for this user.

2. Use the `netcat` (`nc`) command to scan the port on which the proxy server listens for client requests in verbose mode with a connection timeout of 10 seconds.
   ```
   $ nc -zvw10 proxy-server-address port
   ```

proxy-server-address
> The fully qualified domain name or IP address of the proxy server.

port
> The port on which the proxy server listens for client requests.

2.12.2. Since 1.5.0, Optional: Specifying Hosts for which a Configured Proxy Server is Bypassed

If a proxy server is configured, all network communications with the NVIDIA GPU Manager for VMware vCenter appliance are routed through the proxy server by default. However, your network deployment might require communications to some hosts to bypass the proxy server. For example, a proxy server might be required only for communications to external hosts on the public Internet and must be bypassed for communications to hosts within your corporate intranet.

Create a plain-text file that contains a comma-separated list of entries in which each entry defines a single host or a set of related hosts. Each entry must be in one of the following formats:

▶ The complete IPv4 address of a single host, for example, `192.0.2.89`

▶ A range of IPv4 addresses specified by using the wildcard character `*` to replace the decimal digit in the fields at the end of an IPv4 address, for example, `192.0.2.*` or `192.0.*.*`

▶ The fully qualified domain name of a single host, for example, `temp.develop.example.com`

▶ A set of hosts within the same subdomain specified by using the wildcard character `*` at the start of a fully qualified domain name, for example, `*.develop.example.com`

▶ An IPv4 subnet in which the subnet mask is expressed in classless inter-domain routing (CIDR) format, for example, `192.0.2.0/24`

> 📄 **Note:** Wildcard characters are not allowed in a subnet.

To get a subnet mask in CIDR format from its decimal equivalent, refer to the table on page 2 of IETF RFC 1878: Variable Length Subnet Table For IPv4.

For example, the subnet mask in CIDR format of the decimal equivalent 255.255.255.0 is 24.

> 📄 **Note:** The entries in this file are case insensitive.

This file must meet the following requirements:

▶ The size of the file must **not** exceed 5 KB.

▶ The file must **not** be empty.

This example specifies that a configured proxy server is bypassed for communications to hosts with the IPv4 addresses `192.0.2.89` and `192.0.3.101`, and to all hosts in the `develop.example.com` subdomain.

```
192.0.2.89, 192.0.3.101, *.develop.example.com
```

After creating the file, upload it to NVIDIA GPU Manager for VMware vCenter when you configure NVIDIA GPU Manager for VMware vCenter with a proxy server. For more information, refer to the following topics:

▶ Configuring NVIDIA GPU Manager for VMware vCenter with an Unauthenticated Proxy Server

▶ Configuring NVIDIA GPU Manager for VMware vCenter with a Basic-Authenticated Proxy Server

▶ Configuring NVIDIA GPU Manager for VMware vCenter with a Kerberos-Authenticated Proxy Server

If the proxy server settings are removed, any file that has been uploaded is deleted from NVIDIA GPU Manager for VMware vCenter.

2.12.3. Configuring NVIDIA GPU Manager for VMware vCenter with an Unauthenticated Proxy Server

Ensure that you have verified the correct operation of proxy server as explained in Verifying the Correct Operation of the Proxy Server.

Since 1.5.0, Optional: If you require network communications to some hosts to bypass the proxy server, ensure that you have created a plain text file in which these hosts are specified as explained in Since 1.5.0, Optional: Specifying Hosts for which a Configured Proxy Server is Bypassed.

1. Log in to the NVIDIA GPU Manager for VMware vCenter virtual appliance for which you are configuring a proxy server.
2. In the left navigation pane, click **SETTINGS**.
3. On the **SETTINGS** page that opens, in the **Proxy Server Settings** section, click **Update**.
4. In the **Server** section of the **Proxy Server Settings** window that opens, provide the details of the proxy server.
   a). Specify the fully qualified domain name or IP address of the proxy server and the port on which the proxy server listens for client requests.
   b). Select the **NO AUTH** authentication type.
5. **Since 1.5.0:** If you require network communications to some hosts to bypass the proxy server, upload the plain text file in which these hosts are specified.
   a). In the **Exceptions** section of the **Proxy Server Settings** window, click **Choose file**.

b). In the file browser that opens, navigate to the folder that contains the text file in which these hosts are specified and select the file.

> **Note:**
>
> ▶ If a file has previously been uploaded, it is **overwritten**.
>
> ▶ If you change other proxy server settings after uploading a file without uploading a new file, the existing file is retained.

6. **Since 1.5.0:** If you want to configure the proxy server without testing connections through the proxy server, deselect the **Test connection via proxy server** option.

   By default this option is selected. When this option is selected, NVIDIA GPU Manager for VMware vCenter validates the proxy server settings by connecting back to itself through the proxy server. If your network deployment disallows cyclic connections, you can omit the validation step by deselecting the **Test connection via proxy server** option.

7. Click **Update**.

NVIDIA GPU Manager for VMware vCenter verifies the correct operation of proxy server and configures it. A confirmation that the proxy server has been configured appears on the **EVENTS** page.

2.12.4. Configuring NVIDIA GPU Manager for VMware vCenter with a Basic-Authenticated Proxy Server

Ensure that you have verified the correct operation of proxy server as explained in Verifying the Correct Operation of the Proxy Server.

Since 1.5.0, Optional: If you require network communications to some hosts to bypass the proxy server, ensure that you have created a plain text file in which these hosts are specified as explained in Since 1.5.0, Optional: Specifying Hosts for which a Configured Proxy Server is Bypassed.

1. Log in to the NVIDIA GPU Manager for VMware vCenter virtual appliance for which you are configuring a proxy server.
2. In the left navigation pane, click **SETTINGS**.
3. On the **SETTINGS** page that opens, in the **Proxy Server Settings** section, click **Update**.
4. In the **Server** section of the **Proxy Server Settings** window that opens, provide the details of the proxy server.
   a). Specify the fully qualified domain name or IP address of the proxy server and the port on which the proxy server listens for client requests.
   b). Select the **BASIC** authentication type.
   c). Provide the user name and password for authenticating access to the proxy server for using the proxy.

5. **Since 1.5.0:** If you require network communications to some hosts to bypass the proxy server, upload the plain text file in which these hosts are specified.

   a). In the **Exceptions** section of the **Proxy Server Settings** window, click **Choose file**.

   b). In the file browser that opens, navigate to the folder that contains the text file in which these hosts are specified and select the file.

> **Note:**
>
> ▶ If a file has previously been uploaded, it is **overwritten**.
>
> ▶ If you change other proxy server settings after uploading a file without uploading a new file, the existing file is retained.

6. **Since 1.5.0:** If you want to configure the proxy server without testing connections through the proxy server, deselect the **Test connection via proxy server** option.

   By default this option is selected. When this option is selected, NVIDIA GPU Manager for VMware vCenter validates the proxy server settings by connecting back to itself through the proxy server. If your network deployment disallows cyclic connections, you can omit the validation step by deselecting the **Test connection via proxy server** option.

7. Click **Update**.

NVIDIA GPU Manager for VMware vCenter verifies the correct operation of proxy server, verifies the credentials for authenticating access to the proxy server for using the proxy, and configures the proxy server. A confirmation that the proxy server has been configured appears on the **EVENTS** page.

2.12.5. Configuring NVIDIA GPU Manager for VMware vCenter with a Kerberos-Authenticated Proxy Server

Ensure that you have verified the correct operation of proxy server as explained in Verifying the Correct Operation of the Proxy Server.

Since 1.5.0, Optional: If you require network communications to some hosts to bypass the proxy server, ensure that you have created a plain text file in which these hosts are specified as explained in Since 1.5.0, Optional: Specifying Hosts for which a Configured Proxy Server is Bypassed.

1. Ensure that the NVIDIA GPU Manager for VMware vCenter virtual appliance can resolve the fully qualified domain names of the key distribution center (KDC) and the administration server.

   a). Use the web console option of the vCenter web client to log in as the user `vcp_cli_user` to the VM that hosts the NVIDIA GPU Manager for VMware vCenter virtual appliance.

   b). Ensure that the `/etc/hosts` file contains entries for the fully qualified domain names of the KDC and the administration server.

2. Create the [krb5.conf](#) configuration file on the NVIDIA GPU Manager for VMware vCenter virtual appliance.

   If Kerberos is already configured on another server in your network, you can copy the file from that server. The default path to this file is `/etc/krb5.conf`. Otherwise, create the file in the format shown in the following example.

   ```
   [libdefaults]
     default_realm = EXAMPLE.COM

   [realms]
   EXAMPLE.COM = {
        kdc = vc-plugin-dc.EXAMPLE.COM
        admin_server = vc-plugin-dc.EXAMPLE.COM
   }

   [domain_realm]
   .linux.home = EXAMPLE.COM
   .example.com = EXAMPLE.COM
   example.com = EXAMPLE.COM
   ```

3. Log in to the NVIDIA GPU Manager for VMware vCenter virtual appliance for which you are configuring a proxy server.

4. In the left navigation pane, click **SETTINGS**.

5. On the **SETTINGS** page that opens, in the **Proxy Server Settings** section, click **Update**.

6. In the **Server** section of the **Proxy Server Settings** window that opens, provide the details of the proxy server.

   a). Specify the fully qualified domain name or IP address of the proxy server and the port on which the proxy server listens for client requests.

   b). Select the **KERBEROS** authentication type.

   c). Provide the user name and password for authenticating access to the proxy server for using the proxy.

   d). Click **Choose File** adjacent to **Kerberos Configuration File** and, in the file browser that opens, navigate to the folder that contains the `krb5.conf` that you created in Step [2](#) and select the file.

7. **Since 1.5.0:** If you require network communications to some hosts to bypass the proxy server, upload the plain text file in which these hosts are specified.

   a). In the **Exceptions** section of the **Proxy Server Settings** window, click **Choose file**.

   b). In the file browser that opens, navigate to the folder that contains the text file in which these hosts are specified and select the file.

   > **Note:**
   >
   > ▶ If a file has previously been uploaded, it is **overwritten**.
   >
   > ▶ If you change other proxy server settings after uploading a file without uploading a new file, the existing file is retained.

8. Click **Update**.

NVIDIA GPU Manager for VMware vCenter verifies the correct operation of proxy server, verifies the credentials for authenticating access to the proxy server for using the

proxy, and configures the proxy server. A confirmation that the proxy server has been configured appears on the **EVENTS** page.

> **Note:** While connecting to vCenter Server instances, NVIDIA GPU Manager for VMware vCenter does **not** communicate through the Kerberos-authenticated proxy server. For all other communications, the Kerberos-authenticated proxy server is used.

2.12.6. Since 1.5.0: Disabling and Enabling a Configured Proxy Server

By default, a configured proxy server is enabled. To simplify the testing of the behavior of the NVIDIA GPU Manager for VMware vCenter virtual appliance without a proxy server, you can disable the proxy server. When a proxy server is disabled, network communications to all hosts bypass the proxy server. After the testing is complete, you can enable the proxy server again.

1. Log in to the NVIDIA GPU Manager for VMware vCenter virtual appliance for which you want to enable or disable a configured proxy server.
2. In the left navigation pane, click **SETTINGS**.
3. On the **SETTINGS** page that opens, in the **Proxy Server Settings** section, click **Update**.
4. On the **Proxy Server Settings** window that opens, click the button to disable or enable the proxy server.

   ▶ To disable the proxy server, click **DISABLE**.

   ▶ To enable the proxy server, click **ENABLE**.

Chapter 3. Registering and Unregistering NVIDIA GPU Manager for VMware vCenter

By registering NVIDIA GPU Manager for VMware vCenter, you can download and manage NVIDIA GPU drivers from vSphere client. Unregister NVIDIA GPU Manager for VMware vCenter when you no longer require to be able to download and manage NVIDIA GPU drivers from the vSphere client.

3.1. Privileges Required for Registering and Unregistering NVIDIA GPU Manager for VMware vCenter

Registering and unregistering NVIDIA GPU Manager for VMware vCenter requires the privileges for installing and managing extensions, configuring the vSphere Lifecycle Manager service, managing depots and remediation policies, and managing certificates on VMware vCenter.

Users with vCenter Server administrator privileges have all the privileges required for registering and unregistering NVIDIA GPU Manager for VMware vCenter.

Users without full vCenter Server administrator privileges require the following vCenter Server privileges:

▶ Extension Privileges

  ▶ Register extension

  ▶ Unregister extension

  ▶ Update extension

▶ VMware vSphere Lifecycle Manager Configuration Privileges

  ▶ Configure Service

▶ VMware vSphere Lifecycle Manager Settings Privileges

   ▶ Read

   ▶ Write

▶ VMware vSphere Lifecycle Manager Depots Privileges

   ▶ Delete

▶ Certificate Management Privileges

   ▶ Create/Delete (Admins priv)

   ▶ Create/Delete (below Admins priv)

> **Note:** When assigning a role created with these privileges to a user, ensure that the **Propagate to children** option is set.

3.2.   Registering NVIDIA GPU Manager for VMware vCenter

Registering NVIDIA GPU Manager for VMware vCenter creates a repository where downloaded NVIDIA GPU drivers are stored. After registering NVIDIA GPU Manager for VMware vCenter, you can download and manage NVIDIA GPU drivers from the vSphere client.

> **Note:** An NVIDIA GPU Manager for VMware vCenter virtual appliance can be registered with up to **five** vCenter Server instances.

1. Log in to the virtual appliance for the NVIDIA GPU Manager for VMware vCenter that you want to register.
2. In the left navigation pane, click **REGISTRATION**.
3. On the **REGISTRATION** page that opens, click **REGISTER**.
4. In the **Register NVIDIA GPU Manager** window that opens, provide the details of the vCenter Server instance with which you are registering NVIDIA GPU Manager for VMware vCenter, and click **REGISTER**.
   a). In the **vCenter Server** text-entry field, type the fully qualified domain name or IP address of the vCenter Server host.
   b). In the **Username** text-entry field, type the username of the user that will log in to the vCenter Server host.
   c). In the **Password** text-entry field, type the password of the user that will log in to the vCenter Server host.

An entry for the vCenter Server instance with which NVIDIA GPU Manager for VMware vCenter is registered is added to the **REGISTRATION** page.

Use the vSphere client to verify that NVIDIA GPU Manager for VMware vCenter has been registered.

3.3. Verifying the Registration of NVIDIA GPU Manager for VMware vCenter

After registering NVIDIA GPU Manager for VMware vCenter, confirm that it is listed on the **Client Plugins** and **Patch Setup** pages of the vSphere client.

1. Log in to vCenter Server by using the **VMware vSphere Client**.

   For additional information, refer to Log in to vCenter Server by Using the vSphere Web Client on the VMware Docs site.

2. From the **vSphere Client** menu, choose **Administration** > **Solutions** > **Client Plugins** and confirm that **NVIDIA GPU Manager** appears in the **Client Plugins** table.

3. From the **vSphere Client** menu, choose **Lifecycle Manager**, click the **Settings** tab, choose **Administration** > **Patch Setup** , and confirm that the **NVIDIA GPU Manager drivers repository** appears in the table.

After confirming the registration of NVIDIA GPU Manager for VMware vCenter, you can manage NVIDIA GPU drivers as explained in Managing NVIDIA GPU Drivers from the vSphere Client.

3.4. Unregistering NVIDIA GPU Manager for VMware vCenter

Unregister NVIDIA GPU Manager for VMware vCenter when you no longer require to be able to download and manage NVIDIA GPU drivers from the vSphere client. Unregistering NVIDIA GPU Manager for VMware vCenter removes the URL to the repository where downloaded NVIDIA GPU drivers were stored. The driver repository, including and downloaded driver files it contains, remains unchanged and can be accessed again by re-registering NVIDIA GPU Manager for VMware vCenter.

1. Log in to the virtual appliance for the NVIDIA GPU Manager for VMware vCenter that you want to unregister.

2. In the left navigation pane, click **REGISTRATION**.

3. On the **REGISTRATION** page that opens, click **UNREGISTER** for the vCenter Server host from which you want to unregister NVIDIA GPU Manager for VMware vCenter.

4. In the **Unregister NVIDIA GPU Manager** window that opens, provide the credentials of the user that will log in to this vCenter Server host, and click **UNREGISTER**.

   a). In the **Username** text-entry field, type the username of the user that will log in to the vCenter Server host.

b). In the **Password** text-entry field, type the password of the user that will log in to the vCenter Server host.

The entry for the vCenter Server instance from which NVIDIA GPU Manager for VMware vCenter is unregistered is removed from the **REGISTRATION** page.
Use the vSphere client to verify that NVIDIA GPU Manager for VMware vCenter has been unregistered.

3.5. Verifying the Unregistration of NVIDIA GPU Manager for VMware vCenter

After unregistering NVIDIA GPU Manager for VMware vCenter, confirm that it is no longer listed on the **Client Plugins** and **Patch Setup** pages of the vSphere client.

1. Log in to vCenter Server by using the **VMware vSphere Client**.

   For additional information, refer to <u>Log in to vCenter Server by Using the vSphere Web Client</u> on the VMware Docs site.
2. From the **vSphere Client** menu, choose **Administration** > **Solutions** > **Client Plugins** and confirm that **NVIDIA GPU Manager** no longer appears in the **Client Plugins** table.
3. From the **vSphere Client** menu, choose **Lifecycle Manager**, click the **Settings** tab, choose **Administration** > **Patch Setup** , and confirm that the **NVIDIA GPU Manager drivers repository** no longer appears in the table.

If the **NVIDIA GPU Manager drivers repository** still appears in the table, you must delete it manually as explained in <u>Modify a Download Source</u> on the VMware Docs site.

Chapter 4. Managing NVIDIA GPU Drivers from the vSphere Client

After registering NVIDIA GPU Manager for VMware vCenter, you can use the vSphere client to download drivers from NVIDIA Licensing Portal to the NVIDIA GPU Manager drivers repository. After synchronizing the downloaded drivers with vSphere Lifecycle Management (vLCM), you can then install a downloaded driver on an ESXi host. To free storage space in the NVIDIA GPU Manager drivers repository, delete any downloaded drivers that are no longer used.

> **Note:** When you log in to the vSphere client to manage NVIDIA GPU drivers, specify a fully qualified domain name or an IP address in the URL to the vSphere client. To avoid the possibility of a `Plugin Access Check Failed` error, do not specify the host name of the VMware vCenter Server host. This error occurs if **NVIDIA GPU Manager** cannot resolve the host name.

4.1. Privileges Required for Managing NVIDIA GPU Drivers from the vSphere Client

Managing NVIDIA GPU drivers from the vSphere client requires the privileges for configuring the vSphere Lifecycle Manager service, managing depots and remediation policies, and managing baselines and baseline groups on VMware vCenter.

Users with vCenter Server administrator privileges have all the privileges required for managing NVIDIA GPU drivers from the vSphere client.

Users without full vCenter Server administrator privileges require the following vCenter Server privileges:

▶ VMware vSphere Lifecycle Manager Configuration Privileges

    ▶ Configure Service

▶ VMware vSphere Lifecycle Manager Settings Privileges

▶ Read

▶ Write

▶ VMware vSphere Lifecycle Manager Manage Baseline Privileges

   ▶ Attach Baseline

> **Note:** When assigning a role created with these privileges to a user, ensure that the **Propagate to children** option is set.

4.2. Creating a Software Downloads API Key

Automated downloads of NVIDIA software from the NVIDIA Licensing Portal require a software downloads API key for authentication. This key is required for downloading NVIDIA GPU drivers or NVIDIA GPU Manager for VMware vCenter virtual appliance updates.

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 **API KEYS**.
3. In the **API Key Management** window that opens, click **CREATE API KEY**.
4. In the **Create API Key** window that opens, provide the name, type of access, and lifetime of the key, and click **CREATE API KEY**.
    a). In the **Key name** text-entry field, type your choice of name for the key.
    b). From the **Access type** drop-down list, select **Software Downloads**.
    c). Accept the NVIDIA Software License Agreement for use of the vGPU software.
    d). In the **Days until expiration** text-entry field, type the lifetime of the key in days or, if you don't want the key to expire, deselect the **Expire this key after a specified number of days?** option.
    The key is added to the table of API keys in the **API Key Management** window.
5. When you're ready to use the software downloads API key follow the **view api key** link and copy the key to the clipboard.

Configure the vSphere client with the API key that you created as explained in Configuring the vSphere Client with a Software Downloads API Key.

4.3. Configuring the vSphere Client with a Software Downloads API Key

After creating a software downloads API key, configure the vSphere client with the key before using the client to download and manage NVIDIA GPU drivers. For any NVIDIA GPU

Manager for VMware vCenter virtual appliance, you need to configure the API key only once. The key is applied to the vSphere client of every vCenter Server instance with which the appliance is registered.

Ensure that you have created a software downloads API key as explained in Creating a Software Downloads API Key.

1. Log in to vCenter Server by using the **VMware vSphere Client**.

    For additional information, refer to Log in to vCenter Server by Using the vSphere Web Client on the VMware Docs site.
2. From the **vSphere Client** menu, choose **NVIDIA GPU Manager**.
3. Ensure that you have copied the software downloads API key to the clipboard.
4. On the **Driver Download** tab, paste the copied API key into the **NVIDIA Licensing Portal API Key** text-entry field and click **Apply**.

If the **Driver Download** tab displays the message `API key is invalid or expired`, create another software downloads API key and configure the vSphere client with it.

4.4.   Downloading and Synchronizing NVIDIA GPU Drivers

The NVIDIA GPU drivers that you can download from the NVIDIA Licensing Portal depend on the terms of your NVIDIA software entitlements. After downloading the drivers and synchronizing them with vSphere Lifecycle Management (vLCM), you can install a downloaded driver on an ESXi host.

1. Log in to vCenter Server by using the **VMware vSphere Client**.

    For additional information, refer to Log in to vCenter Server by Using the vSphere Web Client on the VMware Docs site.
2. From the **vSphere Client** menu, choose **NVIDIA GPU Manager**.
3. On the **Driver Download** tab, select each driver that you want to download and click **Download Selected**.

    If necessary, click **Refresh Catalog** to update the list of drivers that you can download from the NVIDIA Licensing Portal.

> 🗩 **Note:** You can select up to a maximum of 10 drivers for a single download.
>
> Because each download is large, you cannot start a new download while another download is in progress.

4. When the download is complete, synchronize the downloaded drivers with vLCM.
    a). From the **vSphere Client** menu, choose **Lifecycle Manager**.
    b). On the **Lifecycle Manager** page, from the **ACTIONS** menu, choose **Sync updates**.
5. Confirm that the downloaded drivers have been synchronized with vLCM.
    a). From the **vSphere Client** menu, choose **Lifecycle Manager**.

b). On the **Image Depot** tab of the **Lifecycle Manager** page, click **COMPONENTS** and confirm that the downloaded drivers appear in the table.

After downloading the drivers and synchronizing them with vLCM, install a downloaded driver on an ESXi host as explained in the following topics on the VMware Docs site:

▶ Using vSphere Lifecycle Manager Images

▶ Using vSphere Lifecycle Manager Baselines and Baseline Groups

4.5. Deleting Downloaded NVIDIA GPU Drivers

On the NVIDIA GPU Manager for VMware vCenter virtual appliance, 15 GB of storage is reserved for the NVIDIA GPU Manager drivers repository. To free some storage space, delete unused drivers from the repository.

1. Log in to vCenter Server by using the **VMware vSphere Client**.

   For additional information, refer to Log in to vCenter Server by Using the vSphere Web Client on the VMware Docs site.

2. From the **vSphere Client** menu, choose **NVIDIA GPU Manager**.

3. Click the **Manage Drivers** tab, select each downloaded driver that you want to delete, and click **DELETE SELECTED DOWNLOADS**.

   The **Manage Drivers** tab shows the total number of drivers downloaded to the NVIDIA GPU Manager drivers repository and the total size of the drivers in GB.

> **Note:** Deleting a driver from the NVIDIA GPU Manager drivers repository does not remove the entry for the driver from VLCM.

Chapter 5. Maintaining the NVIDIA GPU Manager for VMware vCenter Virtual Appliance

Maintaining the NVIDIA GPU Manager for VMware vCenter virtual appliance involves updating the virtual appliance in place with minor changes and upgrading the virtual appliance with major changes when a new version of the virtual appliance software is released.

5.1. Updating the NVIDIA GPU Manager for VMware vCenter Virtual Appliance

NVIDIA periodically releases updates to the NVIDIA GPU Manager for VMware vCenter that contain minor changes such as security updates and bug fixes. You install these updates in place on the existing virtual appliance. The NVIDIA GPU Manager for VMware vCenter automatically checks for new updates on the NVIDIA Licensing Portal.

Ensure that an API key has been configured for the vSphere client of a vCenter Server instance with which the appliance is registered. For more information, refer to Configuring the vSphere Client with a Software Downloads API Key.

To prevent loss of data if an update fails, create a snapshot of the virtual appliance before updating it as explained in Take Snapshots of a Virtual Machine. If an update fails, you can restore the virtual appliance from the snapshot as explained in Revert a Virtual Machine Snapshot.

1. Log in to the virtual appliance for the NVIDIA GPU Manager for VMware vCenter that you want to update.
2. In the left navigation bar, click **MAINTENANCE**.
3. If any updates are available, on the **MAINTENANCE** page that opens, click **INSTALL UPDATES**.

If the virtual appliance is up to date, the **MAINTENANCE** page displays a message that no new updates are available. The **INSTALL UPDATES** button is deactivated and dimmed.

> **Note:** During the installation of the updates, the **NVIDIA GPU Manager** application on the appliance and its web-based management interface are unavailable.

After the updates are installed, confirmation of the installation appears on the **EVENTS** page.
Confirm that the NVIDIA GPU Manager for VMware vCenter virtual appliance has been updated.

1. Confirm that the vSphere client shows the correct version of the NVIDIA GPU Manager for VMware vCenter virtual appliance.

2. Confirm that the **Sync updates** operation in vSphere Lifecycle Management (vLCM) succeeds.

5.2. Upgrading the NVIDIA GPU Manager for VMware vCenter Virtual Appliance

NVIDIA distributes major upgrades, including new features, in new releases of the NVIDIA GPU Manager for VMware vCenter virtual appliance software. You install these upgrades side by side on a new virtual appliance and then migrate data from the existing virtual appliance to the new virtual appliance.

The following data is migrated from the existing virtual appliance to the new virtual appliance:

▶ Configuration files

▶ API keys

▶ Registered vCenter Server instances

▶ Downloaded drivers

1. Prepare the new NVIDIA GPU Manager for VMware vCenter virtual appliance for a side-by-side upgrade.
   a). Install the new release of the NVIDIA GPU Manager for VMware vCenter virtual appliance software on a new virtual machine.

   For instructions, refer to Installing the NVIDIA GPU Manager for VMware vCenter Virtual Appliance.
   b). Register the NVIDIA GPU Manager for VMware vCenter administrator user on the NVIDIA GPU Manager for VMware vCenter virtual appliance that was created in the previous step.

For instructions, refer to Registering the NVIDIA GPU Manager for VMware vCenter Administrator User.

2. Log in to the existing virtual appliance that you want to upgrade.

3. In the left navigation bar, click **MAINTENANCE**.

4. On the **MAINTENANCE** page that opens, click **MIGRATE DATA**.

5. In the **Migrate Data** window that opens, provide the details of the new virtual appliance and click **MIGRATE**.

   a). In the **Virtual Appliance** text-entry field, type the fully qualified domain name or IP address of the new virtual appliance.

   b). In the **Password** text-entry field, type the password of the **vcp_admin** user that you registered on the new virtual appliance.

   Migration of data to the new virtual appliance can take between 10 and 30 minutes. During the migration, driver-download operations are not available from **NVIDIA GPU Manager** on the vSphere client.

   > ⚠ **CAUTION:** Do not use **NVIDIA GPU Manager** during the migration.

6. Log in to the new NVIDIA GPU Manager for VMware vCenter virtual appliance and confirm that it is registered with all vCenter Server instances with which the old virtual appliance was registered.

7. Use the vSphere client to verify that the new virtual appliance and its **NVIDIA GPU Manager drivers repository** have been registered.

   For instructions, refer to Verifying the Registration of NVIDIA GPU Manager for VMware vCenter.

   If the fully qualified domain name or IP address of the **NVIDIA GPU Manager drivers repository** for the old virtual appliance appears in the table, delete it.

8. Use the standard interfaces of your hypervisor software to delete the old virtual appliance from the hypervisor host.

After the upgrade is complete, the NVIDIA GPU Manager for VMware vCenter is registered and ready for use. Confirmation of the upgrade appears on the **EVENTS** page. Confirm that the NVIDIA GPU Manager for VMware vCenter virtual appliance has been upgraded.

1. Confirm that the vSphere client shows the correct version and IP address or fully qualified domain name of the NVIDIA GPU Manager for VMware vCenter virtual appliance.

2. Confirm that the **Sync updates** operation in vSphere Lifecycle Management (vLCM) succeeds.

Chapter 6. Troubleshooting NVIDIA GPU Manager for VMware vCenter

6.1. Downloading Log Files from the NVIDIA GPU Manager for VMware vCenter Virtual Appliance

To help NVIDIA Enterprise Support with troubleshooting, you can download the log files in the `/var/log` directory from an NVIDIA GPU Manager for VMware vCenter virtual appliance.

1. Log in to the NVIDIA GPU Manager for VMware vCenter virtual appliance from which you want to download log files.
2. In the left navigation bar, click **EVENTS**.
3. On the **EVENTS** page that opens, click **EXPORT EVENTS**.
   A file named `log_bundle.zip` is saved to your default downloads folder.

6.2. The **NVIDIA GPU Manager** Management Interface Cannot be Accessed with Firefox or Safari

The NVIDIA GPU Manager for VMware vCenter virtual appliance provides a web-based management interface to the **NVIDIA GPU Manager** application on the appliance. The appliance's date and time are synchronized with the date and time on the ESXi host on which the appliance is deployed. If the ESXi host's date and time are not set accurately, the management interface cannot be accessed with the Firefox or Safari web browsers because of their strict security requirements on the expiration of cookies.

Ensure that the ESXi host's date and time are set accurately, for example, by using NTP. For information about how to configure NTP on an ESXi host, refer to VMware Knowledge Base Article: Configuring Network Time Protocol (NTP) on an ESXi host using the vSphere Client (57147).

6.3. The Virtual Appliance VM Was Deleted While NVIDIA GPU Manager for VMware vCenter Was Still Registered

NVIDIA GPU Manager for VMware vCenter is unregistered from the **NVIDIA GPU Manager** application on the virtual appliance. If the virtual appliance VM has already been deleted, use the vCenter Managed Object Browser (MOB) interface in a web browser to delete the NVIDIA GPU Manager for VMware vCenter extension manually.

1. Open a web browser and connect to the URL `https://vcenter-server-ip-address/mob/?moid=ExtensionManager&method=unregisterExtension`.

   vcenter-server-ip-address
   The IP address or, if defined, the fully qualified domain name of your vCenter Server instance.

2. When prompted, sign in with the username and password of a vCenter Server administrator user on the vCenter Server instance.
   The **void UnregisterExtension** page opens.

3. In the text box for the **extensionKey** value, type **com.nvidia.gpu.manager**.

4. Follow the **Invoke Method** link.

5. Confirm that NVIDIA GPU Manager for VMware vCenter is no longer listed on the **Client Plugins** page of the vSphere client.

   a). Log in to vCenter Server by using the **VMware vSphere Client**.

      For additional information, refer to Log in to vCenter Server by Using the vSphere Web Client on the VMware Docs site.

   b). From the **vSphere Client** menu, choose **Administration** > **Solutions** > **Client Plugins** and confirm that **NVIDIA GPU Manager** no longer appears in the **Client Plugins** table.

6. Delete the entry for the **NVIDIA GPU Manager drivers repository** manually as explained in Modify a Download Source on the VMware Docs site.

6.4. NVIDIA GPU Manager for VMware vCenter Could Not Be Registered

To communicate with **NVIDIA GPU Manager**, VMware vCenter Server must trust the SSL certificate on the NVIDIA GPU Manager for VMware vCenter appliance. When you register NVIDIA GPU Manager for VMware vCenter, VMware vCenter Server first imports the SSL certificate into its truststore. If the SSL certificate cannot be imported, registration of NVIDIA GPU Manager for VMware vCenter might fail.

1. Import the certificate manually.

   For instructions, refer to <u>VMware Knowledge Base Article: Adding a new online depot to vCenter Server 7.0 U2C and above vCenter Life Cycle Manager fails due to "self signed certificate" error (85819)</u>.
2. Try to register NVIDIA GPU Manager for VMware vCenter again.

   For instructions, refer to <u>Registering NVIDIA GPU Manager for VMware vCenter</u>.