Chapter 1. Creating and Using an NVIDIA Quadro Virtual Workstation Instance from the Amazon Web Services Elastic Compute Cloud Marketplace ......................................................... 1
  1.1. Creating a GPU-Accelerated Virtual Workstation ........................................................ 1
  1.2. Connecting to a GPU-Accelerated Virtual Workstation on a Windows VM Instance .......... 4
  1.3. Verifying the Creation of your GPU-Accelerated Virtual Workstation ............................... 5
    1.3.1. Verifying the Creation of a GPU-Accelerated Virtual Workstation on a Windows VM Instance .................................................................................................................. 5
    1.3.2. Verifying the Creation of a GPU-Accelerated Virtual Workstation on a Linux VM Instance ...................................................................................................................... 7
Appendix A. Resources for NVIDIA Quadro Virtual Workstation on Amazon Web Services Elastic Compute Cloud ........................................................................................................... 9
NVIDIA® Quadro® Virtual Workstation is an NVIDIA Virtual Machine Image (VMI) preconfigured with Quadro Virtual Workstation software and NVIDIA GPU hardware. The NVIDIA Quadro driver is preinstalled on the VMI and NVIDIA ensures that the image is always up to date with the latest Quadro ISV certifications, patches, and upgrades. Support and technical information to help you get started are available on the NVIDIA Quadro vWS on CSP Marketplace community forum and from additional resources.

1.1. Creating a GPU-Accelerated Virtual Workstation

Prepare for creating a GPU-accelerated virtual workstation from the Amazon Web Services Elastic Compute Cloud marketplace as follows:

- Ensure that you have a Amazon Web Services Elastic Compute Cloud account and an active subscription.
- Decide the EC2 instance type on which you want to deploy your GPU-accelerated virtual workstation.

For information about the EC2 instance types that support Quadro Virtual Workstation, see Quadro Virtual Workstation on Amazon Web Services Elastic Compute Cloud Release Notes.

1. In a web browser, log on to Amazon EC2 console (https://console.aws.amazon.com/ec2/).
2. Go to the NVIDIA Quadro Virtual Workstation listing for the guest OS that you want to use:

- NVIDIA Quadro Virtual Workstation - WinServer 2019
- NVIDIA Quadro Virtual Workstation - WinServer 2016
- NVIDIA Quadro Virtual Workstation - Amazon Linux 2
- NVIDIA Quadro Virtual Workstation - CentOS 7
- NVIDIA Quadro Virtual Workstation - Ubuntu 18.04

3. Follow the View Details link to view hourly pricing for the Quadro Virtual Workstation software.

   There are also additional Amazon Web Services Elastic Compute Cloud infrastructure costs for the GPU, memory, and storage.
4. Select the most cost effective region and the EC2 instance type on which you want to deploy your GPU-accelerated virtual workstation.

The only fulfillment option available for NVIDIA Quadro Virtual Workstation on Amazon Web Services Elastic Compute Cloud is 64-bit (x86) Amazon Machine Image (AMI).

5. Click Continue to Subscribe.

6. In the Subscribe to this software page that opens, click Accept Terms.

7. When the subscription is complete, click Continue to Configuration.

8. In the Configure this software page that opens, select the following options:

- The software version of the NVIDIA vGPU software graphics driver that you want to run in your instance
- The region where you want the instance to be deployed

The only fulfillment option available for NVIDIA Quadro Virtual Workstation on Amazon Web Services Elastic Compute Cloud is 64-bit (x86) Amazon Machine Image (AMI)

9. Click Continue to Launch.

10. On the Launch this software page that opens, from the Choose Action drop-down list, select the action for launching the instance.

- To launch the instance from the Launch this software page, select Launch from Website.
  
  If you select this option, the other options remain on this page to enable you to configure the instance from this page.

- To launch the instance from the Amazon EC2 console page, select Launch through EC2.
  
  If you select this option, the other options are removed from this page to enable you to configure the instance by using the Launch Instance wizard.

11. Configure and launch the instance.

- If you selected Launch from Website, use the options on the Launch this software page to configure the instance and click Launch.
- If you selected Launch through EC2, click Launch and use the Launch Instance wizard to configure and launch the wizard.

For more information, see the instructions for your chosen OS in the Amazon Elastic Compute Cloud documentation:

- Launching an Instance Using the Launch Instance Wizard (Windows)
- Launching an Instance Using the Launch Instance Wizard (Linux)

When prompted, ensure that you create or add a key pair to the set of keys authorized for this instance. If you proceed without a key pair, you will not be able to log in to the instance.

When the instance is launched, a confirmation page is displayed with a link to the instance on the EC2 Console.
12. In the launch confirmation page, follow the link to your instance on the EC2 Console to confirm that the instance is running and that all status checks are completed.

13. If you are creating a Windows instance and want to connect to the instance through RDP, edit the instance’s security group to add an RDP security rule to open port 3389 for TCP connections.

Follow the instructions for your chosen OS to connect to the instance:

- Connecting to a GPU-Accelerated Virtual Workstation on a Windows VM Instance
- Connect to Your Linux Instance in Amazon EC2 User Guide for Linux Instances

After connecting to the instance, verify that it was created properly as explained in Verifying the Creation of your GPU-Accelerated Virtual Workstation

1.2. Connecting to a GPU-Accelerated Virtual Workstation on a Windows VM Instance

For instructions for a Linux VM, see Connect to Your Linux Instance in Amazon EC2 User Guide for Linux Instances.

1. In the list of instances on the EC2 Dashboard, select the instance and click Connect.
2. In the Connect to Your Instance window that opens, download the RDP file and get the password for connecting to your instance.
   a) Click Download Remote Desktop File.
   b) Click Get Password.
   c) When prompted, specify the path to the key pair file that is associated with this instance and click Decrypt Password.
3. When the remote desktop file is downloaded, double-click the file to start a Remote Desktop Connection session on the VM.

4. If you are warned that the publisher of the remote connection cannot be identified and are asked about whether to connect anyway, click Connect.

5. When you are prompted, log in to the VM with the user name Administrator and the password that you got earlier.

6. If you are warned that the publisher of the remote connection cannot be identified and are asked about whether to connect anyway, click Yes.

7. Close the Connect to Your Instance window.

After connecting to your GPU-accelerated virtual workstation, verify that it was created properly as explained in Verifying the Creation of your GPU-Accelerated Virtual Workstation.

1.3. Verifying the Creation of your GPU-Accelerated Virtual Workstation

After connecting to your GPU-accelerated virtual workstation, verify that it was created properly by listing its GPUs. On a Windows VM instance you can also use NVIDIA Control Panel to verify that the NVIDIA driver is running.

For details, see:

- Verifying the Creation of a GPU-Accelerated Virtual Workstation on a Windows VM Instance
- Verifying the Creation of a GPU-Accelerated Virtual Workstation on a Linux VM Instance

After verifying the creation of your GPU-accelerated virtual workstation, you are now ready to run your design and engineering software.

1.3.1. Verifying the Creation of a GPU-Accelerated Virtual Workstation on a Windows VM Instance

1. Open a command prompt window and change to the folder that contains the nvidia-smi command.
2. List the GPUs in your GPU-accelerated virtual workstation by running the `nvidia-smi` command without any options.

The following example shows the output from `nvidia-smi` for a Windows VM instance configured with four NVIDIA V100 GPUs.

```
C:\Program Files\NVIDIA Corporation\NVSMI>nvidia-smi
Fri Jul 12 01:33:40 2019
+-----------------------------------------------------------------------------
| NVIDIA-SMI 425.31       Driver Version: 425.31       CUDA Version: 10.1 |
|-------------------------------+----------------------|
| GPU  Name            TCC/WDDM | Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
|===============================|======================|
| 0  Tesla V100-SXM2... WDDM | 00000000:00:1B.0 Off | N/A  42C    P0  37W / 300W | 343MiB / 16384MiB |      1% Default |
|-------------------------------+----------------------|
| 1  Tesla V100-SXM2... WDDM | 00000000:00:1C.0 Off | N/A  41C    P0  38W / 300W | 239MiB / 16384MiB |      0% Default |
|-------------------------------+----------------------|
| 2  Tesla V100-SXM2... WDDM | 00000000:00:1D.0 Off | N/A  43C    P0  40W / 300W | 239MiB / 16384MiB |      0% Default |
|-------------------------------+----------------------|
| 3  Tesla V100-SXM2... WDDM | 00000000:00:1E.0 Off | N/A  42C    P0  38W / 300W | 239MiB / 16384MiB |      0% Default |
+-----------------------------------------------------------------------------
+ Processes:                     GPU Memory |
| GPU       PID   Type   Process name                             Usage |
|-------------------------------|-------------------|------------------|
| 0 1256    C+G   Insufficient Permissions                   N/A |
| 0 1536    C+G   ...t_cw5n1h2txyewy\ShellExperienceHost.exe N/A |
| 0 4244    C+G   C:\Windows\explorer.exe                    N/A |
| 0 4340    C+G   Insufficient Permissions                   N/A |
| 0 5724    C+G   ...dows.Cortana_cw5n1h2txyewy\SearchUI.exe N/A |
```
3. Start **NVIDIA Control Panel** to verify that the NVIDIA driver is running.
   a) Right-click on the desktop.
   b) From the menu that opens, choose **NVIDIA Control Panel**.

4. In the **NVIDIA Control Panel**, from the **Help** menu, choose **System Information** to get information about the GPU.

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

1.3.2. **Verifying the Creation of a GPU-Accelerated Virtual Workstation on a Linux VM Instance**

List the GPUs in your GPU-accelerated virtual workstation by running the `nvidia-smi` command without any options.

The following example shows the output from `nvidia-smi` for a Linux VM instance configured with four NVIDIA V100 GPUs.

```
ubuntu@ip-172-31-16-135:~$ nvidia-smi
Fri Jul 12 01:47:40 2019
+-----------------------------------------------------------------------------+
| NVIDIA-SMI 425.31       Driver Version: 425.31       CUDA Version: 10.1     |
|-------------------------------+----------------------+----------------------+
| GPU  Name            TCC/WDDM | Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
|===============================+======================+======================|
|   0  Tesla V100-SXM2... WDDM  | 00000000:00:1B.0 Off |                    0 |
| N/A   42C    P0    37W / 300W |    343MiB / 16384MiB |      1%      Default |
|-------------------------------+----------------------+----------------------+
|   1  Tesla V100-SXM2... WDDM  | 00000000:00:1C.0 Off |                    0 |
| N/A   41C    P0    38W / 300W |    239MiB / 16384MiB |      0%      Default |
```

---

**www.nvidia.com**

Quadro Virtual Workstation on Amazon Web Services
Elastic Compute Cloud
<table>
<thead>
<tr>
<th>Process ID</th>
<th>Type</th>
<th>Process Name</th>
<th>GPU Memory</th>
<th>Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**GPU Memory**

<table>
<thead>
<tr>
<th>GPU</th>
<th>PID</th>
<th>Type</th>
<th>Process Name</th>
<th>Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**Processes:**

| No running processes found |

---

**GPU Memory:**

<table>
<thead>
<tr>
<th>GPU</th>
<th>PID</th>
<th>Type</th>
<th>Process Name</th>
<th>Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix A.
RESOURCES FOR NVIDIA QUADRO VIRTUAL WORKSTATION ON AMAZON WEB SERVICES ELASTIC COMPUTE CLOUD

- Quadro Virtual Workstation on Amazon Web Services Elastic Compute Cloud Release Notes
- NVIDIA Quadro Virtual Workstation listings on Amazon Web Services Elastic Compute Cloud marketplace:
  - NVIDIA Quadro Virtual Workstation - WinServer 2019
  - NVIDIA Quadro Virtual Workstation - WinServer 2016
  - NVIDIA Quadro Virtual Workstation - Amazon Linux 2
  - NVIDIA Quadro Virtual Workstation - CentOS 7
  - NVIDIA Quadro Virtual Workstation - Ubuntu 18.04
- Amazon Web Services Elastic Compute Cloud documentation:
  - Launching an Instance Using the Launch Instance Wizard (Windows)
  - Launching an Instance Using the Launch Instance Wizard (Linux)
  - Connecting to Your Windows Instance
  - Connect to Your Linux Instance
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.

www.nvidia.com
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, 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
© 2020 NVIDIA Corporation. All rights reserved.