NVIDIA RTX Server Reference Architecture

HPE ProLiant DL380 and Teradici for Autodesk Maya 2020 with Arnold Render 6

Application Sizing Guide

SP-09916-001_v01 | April 2020

Document History

SP-09916-001_v01

Version	Date	Authors	Description of Change
01	April 17, 2020	NN, SM	Initial Release

Table of Contents

List of Figures

List of Tables

Chapter 1. Executive Summary

This specification provides insights on how to deploy NVIDIA® Quadro® Virtual Data Center Workstation (Quadro vDWS) software for modern day production pipelines within the media and entertainment (M&E) industry. Recommendations are based on actual customer deployments and sample proof-of-concept (PoC) artistic 3D production pipeline workflows and cover three common questions:

- Which NVIDIA GPU should I use for a 3D Production pipeline?
- How do I select the right profile(s) for the types of users I will have?
- Using sample 3D production pipeline workflows, how many users can be supported (user density) for this server configuration and workflow?

NVIDIA RTX™ Server offers a highly flexible reference design which combines NVIDIA® Quadro RTX™ 6000 or Quadro RTX 8000 graphics processing units (GPUs) with NVIDIA virtual GPU software running on OEM server hardware. NVIDIA RTX Server can be configured to accelerate multiple workloads within the data center. IT administrators can provision multiple, easy-to-manage virtual workstations to tackle various artistic workloads. Since user behavior varies and is a critical factor in determining the best GPU and profile size, the recommendations in this reference architecture are meant to be a guide. The most successful customer deployments start with a proof of concept (PoC) and are "tuned" throughout the lifecycle of the deployment. Beginning with a PoC enables customers to understand the expectations and behavior of their users and optimize their deployment for the best user density while maintaining required performance levels. A PoC also allows administrators to understand infrastructure conditions, such as network, which is a key component to ensure performance within their specific environment. Continued maintenance is important because user behavior can change over the course of a project and as the role of an individual changes in the organization along with potential improvement of displays during refresh cycles. A 3D production artist that was once a light graphics user might become a heavy graphics user when they change teams, assigned to a different project or even receive a display upgrade to a higher resolution monitor. NVIDIA virtual GPU management and monitoring tools enable administrators and IT staff to ensure their deployment is optimized for each user.

Chapter 2. About Autodesk Maya 2020 and Arnold

Autodesk Maya 2020 is one of the most recognizable applications for 3D computer animation, modeling, simulation, and rendering utilized to create expansive worlds, complex characters, and dazzling effects. Creative professionals bring believable characters to life with engaging animation tools, shape 3D objects and scenes with intuitive modeling tools, and create realistic effects - from explosions to cloth simulation all within the Maya software.

Autodesk Arnold is the built-in interactive renderer for Maya and is an advanced Monte Carlo ray tracing renderer. It's designed for artists and for the demands of modern animation and visual effects (VFX) production. Originally co-developed with Sony Pictures Imageworks and now their main renderer, Arnold is used at over 300 studios worldwide including ILM, Framestore, MPC, The Mill and Digic Pictures. Arnold was the primary renderer on dozens of films from *Monster House* and *Cloudy with a Chance of Meatballs* to *Pacific Rim* and *Gravity*. It is available as a standalone renderer on Linux, Windows, and Mac OS, with supported plug-ins for Maya, 3ds Max, Houdini, Cinema 4D, and Katana.

Autodesk works closely with NVIDIA to ensure that creative innovation is never over. Studio drivers are released throughout the year to supercharge your favorite, most demanding applications. Using the same NVIDIA Studio drivers that are deployed on non-virtualized systems, NVIDIA Quadro vDWS software provides virtual machines (VMs) with the same breakthrough performance and versatility that the NVIDIA RTX platform offers to a physical environment. VDI eliminates the need to install Autodesk Arnold and Maya on a local client, which can help reduce IT support and maintenance costs and enables greater mobility and collaboration. This virtual workstation deployment option enhances flexibility and further expands the wide variety of platform choices available to Autodesk customers.

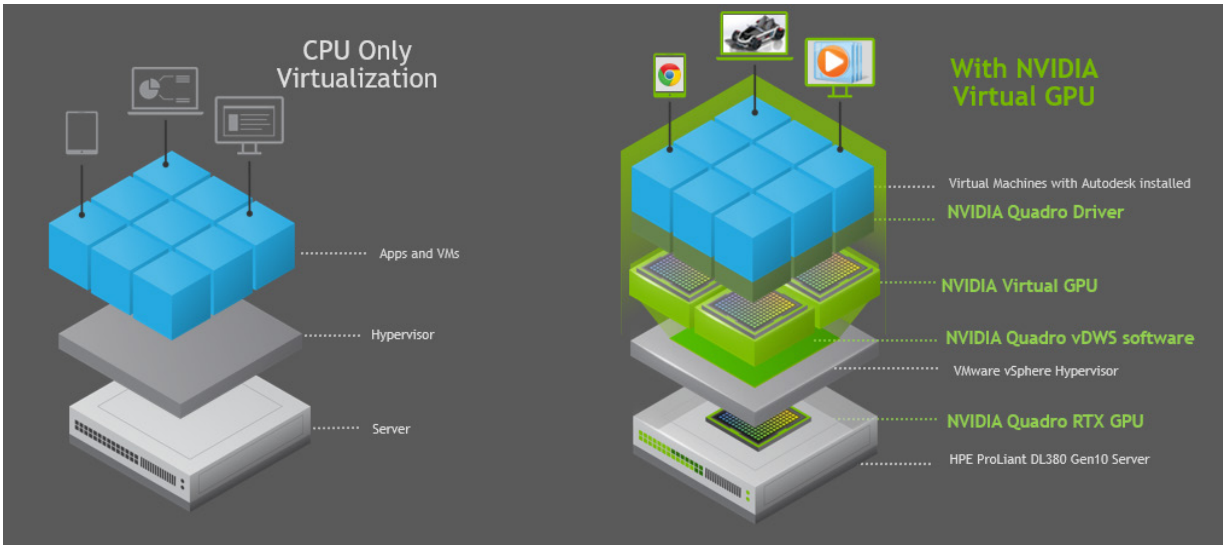
Chapter 3. About NVIDIA RTX Server on HPE Proliant

This reference architecture is comprised of the following components:

- HPE ProLiant DL380 Gen10 server
- NVIDIA Quadro RTX 6000 and/or Quadro RTX 8000 graphics cards
- NVIDIA Quadro vDWS GPU virtualization software
- Autodesk Maya 2020 design software
- Autodesk Arnold 6 rendering software
- Teradici Cloud access software

When combined, this validated NVIDIA RTX Server solution provides unprecedented rendering and compute performance at a fraction of the cost, space, and power consumption of traditional CPU-based render nodes, as well as high performance virtual workstations enabling designers and artists to arrive at their best work, faster.

Figure 3-1. NVIDIA RTX Server Solution



Refer to Appendix A for further details regarding the system configuration used to complete the rigorous NVIDIA NVQual verification for Autodesk Maya, Autodesk Arnold, and Teradici software packages.

3.1 NVIDIA RTX Server

NVIDIA RTX Server is a validated reference design for multiple workloads that are accelerated by Quadro RTX 6000 or Quadro RTX 8000 GPUs. When deployed for high performance virtual workstations, the NVIDIA RTX Server solution delivers a native physical workstation experience from the data center, enabling creative professionals to do their best work from anywhere, using any device. NVIDIA RTX Server can also bring GPU-acceleration and performance to deliver the most efficient end-to-end rendering solution, from interactive sessions in the desktop to final batch rendering in the data center. Content production is undergoing massive growth as render complexity and quality demands increase. Designers and artists across industries continually strive to produce more visually rich content faster than ever before, yet find their creativity and productivity bound by inefficient CPU-based render solutions. NVIDIA RTX Server delivers the performance that all artists need, by allowing them to take advantage of key GPU enhancements to increase interactivity and visual quality, while centralizing GPU resources.

3.2 Quadro RTX GPUs

The NVIDIA Quadro RTX 6000 and Quadro RTX 8000, both powered by the NVIDIA Turing™ architecture and the NVIDIA RTX platform, bring the most significant advancement in computer graphics in over a decade to professional workflows. Designers and artists can now wield the power of hardware-accelerated ray tracing, deep learning, and advanced shading to dramatically boost productivity and create amazing content faster than ever before. The Quadro RTX 6000 has 24 GB of GPU memory, whereas the Quadro RTX 8000 has 48 GB to handle larger animations or visualizations. The artistic workflows covered within our testing for this reference architecture used Quadro RTX 6000 GPUs.

3.3 Quadro Virtual Data Center Workstation Software

NVIDIA virtual GPU (vGPU) software enables the delivery of graphics-rich virtual desktops and workstations accelerated by NVIDIA GPUs. There are three versions of NVIDIA vGPU software available, one being NVIDIA Quadro Virtual Data Center Workstation (Quadro vDWS). NVIDIA Quadro vDWS software includes the Quadro graphics driver required to run professional 3D applications. The Quadro vDWS license enables sharing an NVIDIA GPU across multiple virtual machines, or multiple GPUs can be allocated to a single virtual machine to power the most demanding workflows.

NVIDIA Quadro is the world's preeminent visual computing platform, trusted by millions of creative and technical professionals to accelerate their workflows. With Quadro vDWS software, you can deliver the most powerful virtual workstation from the data center. Designers and artists can work more efficiently, leveraging high performance virtual workstations that perform just like physical workstations. IT has the flexibility to provision render nodes and virtual workstations, scaling resources up or down as needed. An NVIDIA RTX Server solution can be configured to deliver multiple virtual workstations customized for specific tasks. This means that utilization of compute resources can be optimized, and virtual machines can be adjusted to handle workflows that may demand more or less memory.

To deploy an NVIDIA vGPU solution for Autodesk Maya 2020 with Arnold, you will need NVIDIA GPUs and a Quadro vDWS software license for each concurrent user.

3.4 VMware vSphere

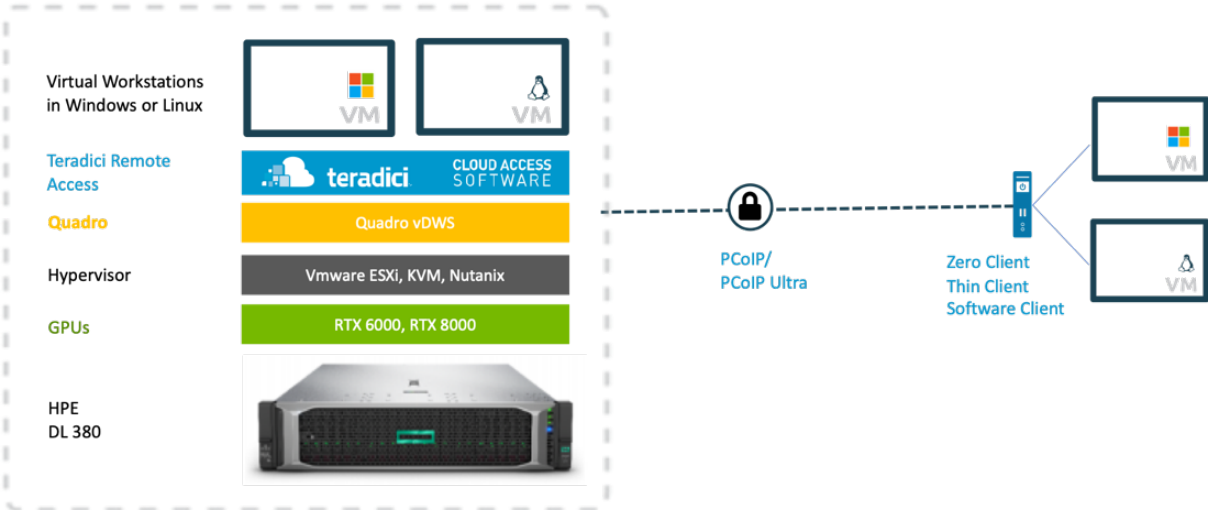
VMware vSphere provides a powerful, flexible, and secure foundation for business agility that accelerates your digital transformation to hybrid cloud and success in the digital economy. With vSphere, you can support new workloads and use cases while keeping pace with the growing needs and complexity of your infrastructure. vSphere is the heart of a secure software defined data center (SDDC). With SDDC securing applications, data, infrastructure, and access has never been easier. Advanced security capabilities fully integrated into the hypervisor and powered by machine learning, provide better visibility, protection and faster response times for security incidents. vSphere helps you run, manage, connect and secure your applications in a common operating environment across the hybrid cloud.

3.5 Teradici Cloud Access Software

Teradici is the creator of the industry-leading PCoIP remoting protocol technology and Cloud Access software. Teradici Cloud Access software enables enterprises to securely deliver high performance graphics-intensive applications and workstations from private data centers, public clouds or hybrid environments with crisp text clarity, true color accuracy and lossless image quality to any endpoint, anywhere.

Teradici PCoIP Ultra with NVIDIA RTX Server can provide virtual machines to multiple artists resulting in virtual machines that are indistinguishable from physical workstations. Artists can enjoy workspaces set up on the latest hardware, and work with confidence in high fidelity with steady frame rates.

Figure 3-2. HPE DL380 NVIDIA RTX Server with Teradici



3.6 HPE ProLiant DL380 Gen10

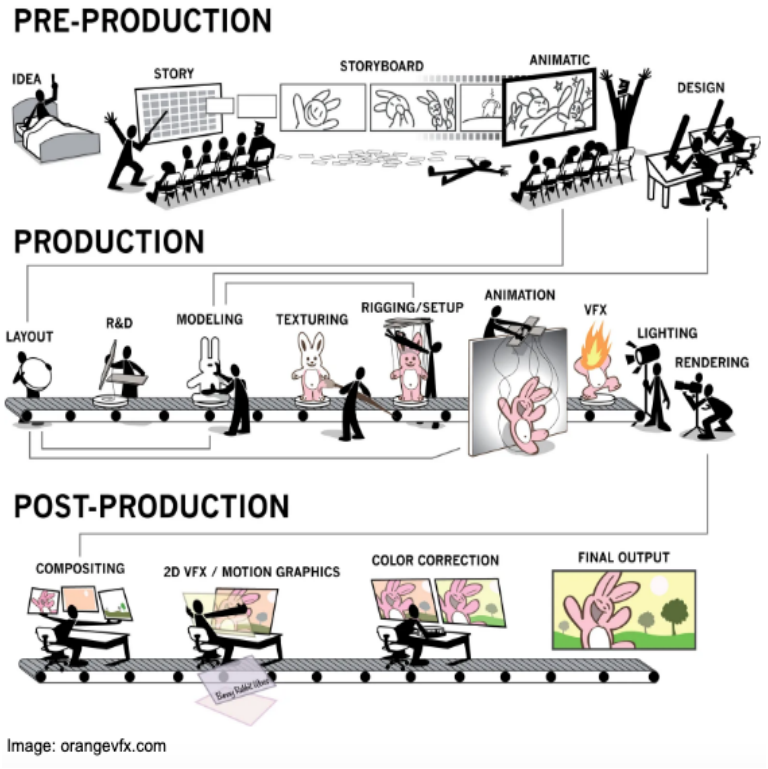
HPE is committed to innovation, quality, and an excellent customer experience. Excellence in innovation and quality is instilled across the product life cycle, from a customer-first approach to design, to supplier selection, quality and management, to world-class manufacturing and rigorous product testing, to global support services and a strong network of channel partners.

The HPE ProLiant DL380 Gen10 Server delivers the latest in security, performance, and expandability. It supports the Intel® Xeon® Processor Scalable Family supporting HPE 2933 MT/s DDR4 SmartMemory. The HPE ProLiant DL380 Gen10 Server has an adaptable chassis, including new HPE modular drive bay configuration options with up to 30 SFF, up to 19 LFF, or up to 20 NVMe drive options along with support for up to 3 double wide GPU options. Along with an embedded 4x1GbE, there is a choice of HPE FlexibleLOM or PCIe standup adapters which offer a choice of networking bandwidth (1 GbE to 40 GbE) and fabric allowing customers to adapt and grow to changing business needs. The HPE ProLiant DL380 Gen10 Server comes with a complete set of HPE technology services, delivering confidence, reducing risk, and helping customers realize agility and stability.

Chapter 4. Autodesk Maya and Arnold PoC Testing

To determine the optimal configuration of Quadro vDWS for Autodesk Maya and Arnold, both user performance and scalability were considered. For comparative purposes, we considered the requirements for a configuration optimized for performance only, and this configuration is based solely on performance using sample artistic workflows. The scenes used within our PoC testing focused on a VFX pipeline where a single shot is the result of several artist specialists working on different pieces. The following illustration shows the entire 3D production pipeline and illustrates the areas where our PoC testing focused.

Figure 4-1. 3D Production Pipeline



Our testing focused on a few of the phases illustrated in Figure 4-1. We executed three GPU-accelerated artistic workflows within 4 VM's:

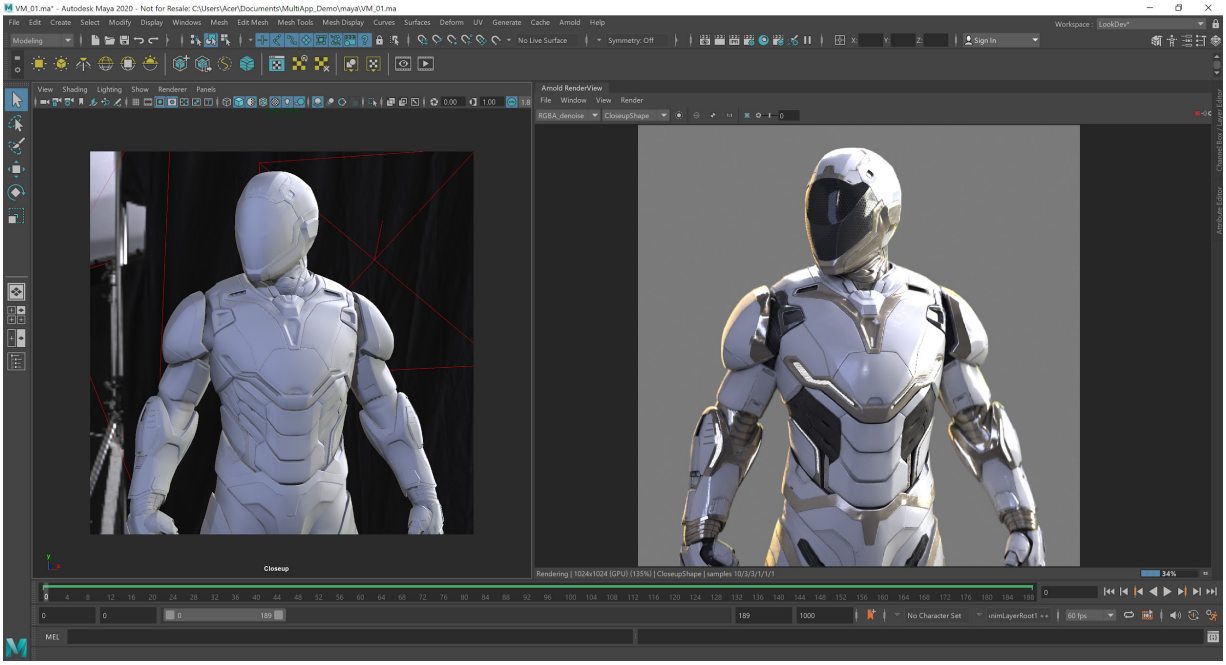
- VM1 and VM2 - Modeling, Texturing and Shading
- VM3 - Animation
- VM4 - Lighting and Rendering

The goal of this testing was to show how four artists from three unique parts of the pipeline can all work at the same time using shared server virtualized resources and be productive. The following paragraphs goes into further detail of each of these workflows.

4.1 VM1 and VM2 - Modeling, Texturing and Shading

For artists to model effectively, they need fast interaction with their models to see different views, quick material changes, and realistic rendering. This workflow takes advantage of the NVIDIA® TensorRT™ cores in the NVIDIA RTX Server to accelerate the rendering process, and artists can view their noiseless assets by leveraging NVIDIA OptiX™ AI Denoising. The GPU memory needed to support this artist would be considered small to medium, therefore a single VM was assigned half of the Quadro RTX 6000 GPU, which equates to a 12Q vGPU profile. Two VM's can share the same GPU on a server. The following screenshot illustrates the artist's work.

Figure 4-2. VM Modeling, Texturing and Shading Example



In order to bring characters to life in film, they need to go through a "Look Development" process. In the example illustration in Figure 4-2, Autodesk's Arnold GPU Renderer utilizes NVIDIA RTX compatible features for performant ray tracing. Look Development involves the following:

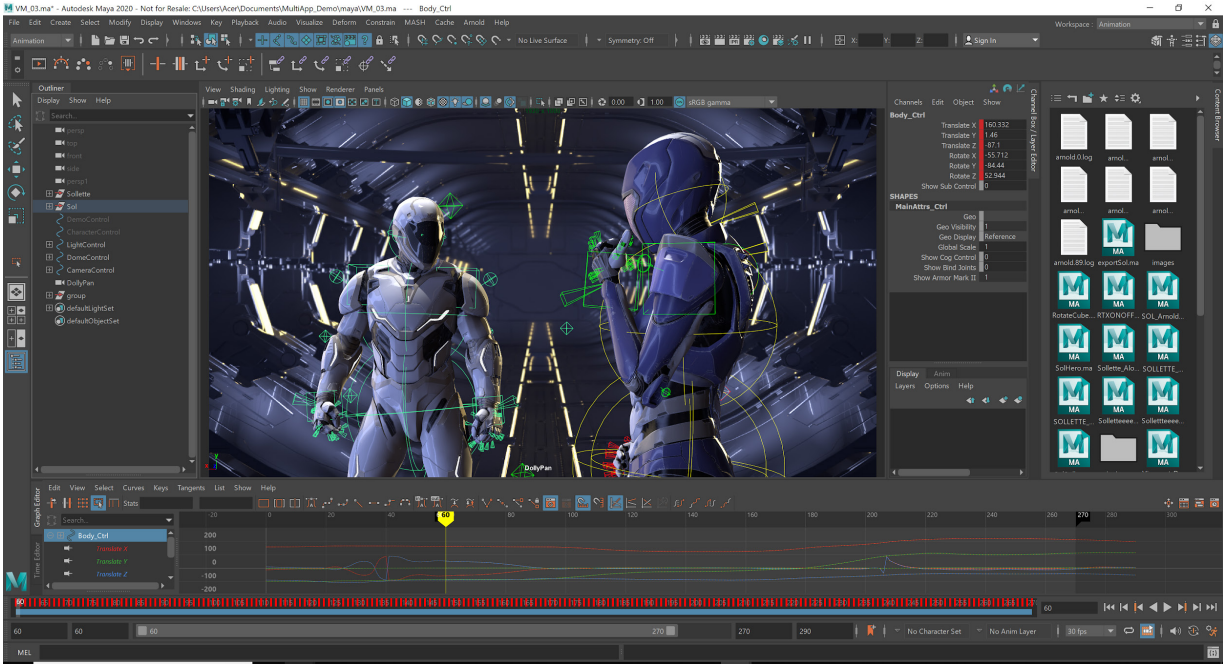
- Refining textures and materials that often result in a time-consuming, back and forth process.
- Real time updates with NVIDIA RTX Server allow for artistic interaction to accurately dial in the look of the character, in-context to the scene.
- NVIDIA RTX AI, employing NVIDIA OptiX Denoiser, provides high-fidelity changes in real time.
- Artists can define and deliver higher quality content in a more intuitive workflow providing an overall increase in production value.

Having a full color range without compression is important to make accurate changes in confidence. Teradici PCoIP Ultra, which takes advantage of NVIDIA RTX GPU encoding, ensures that the virtual machines look indistinguishable from a local display.

4.2 VM3 - Animation

For artists to animate effectively, artists need smooth playback with no pauses or stutters as they make pose changes. Since this artist uses the Maya 2020 GPU animation cache, the GPU memory needed to support this artist would be considered large. Therefore, a single VM was assigned an entire Quadro RTX 6000 GPU, which equates to a 24Q vGPU profile. The following screenshot illustrates the artist's work.

Figure 4-3. VM3 Animation Example



Animation production can place extreme demands on compute hardware. Traditional workflows involve artists outputting time-consuming preview videos. Since Autodesk Maya 2019, real time animation playback and preview is now possible. Furthermore, with Viewport 2.0 enhancements, real-time rendering features are also available. In this scene, we are using the GPU to cache animation, and preview ambient occlusion, shadows, lights and reflections, all in real-time in the viewport. Maya Viewport 2.0 leverages GPU memory to deliver high quality materials, lights, screen space ambient occlusion and more - at interactive speed. Starting in Maya 2019, you can use your GPU to cache animation calculations to memory in a fraction of the time of a CPU cache. With this feature, you can playback your animations in real time, and continue to tweak and update your shots without having to play blast the timeline.

By leveraging NVIDIA RTX GPU encoding with PCoIP Ultra, this VM can deliver interactive, real time animation playback without dropping any frames, which is important to animators who are constantly reviewing their changes. Every frame counts.

4.3 VM4 - Lighting and Rendering

Artists who work with lighting and rendering, need fast resolution of the full image so they can see the impact of their lighting and camera changes. Since this artist is the user who most intensely uses the NVIDIA TensorRT cores in the NVIDIA RTX Server (for accelerating the rendering process), the GPU memory needed to support this artist is the largest of all and may even need acceleration from multiple GPUs. NVIDIA vGPU technology provides administrators the ability to assign up to four shared GPUs to a single VM. The following screenshot illustrates the artist's work.

Figure 4-4. VM4 Lighting and Rendering Example

Lighting and rendering are resource intensive processes that are responsible for the final output of a scene. NVIDIA RTX Server enables artists to work and adjust scenes while utilizing leftover GPU resources to render. This provides for an incredibly efficient use of GPU resources, furthering the production pipeline workflow.

4.4 Evaluating vGPU Frame Buffer

The GPU Profiler is a tool which can be installed within each of the VM's and used for evaluating GPU to CPU utilization rates while executing the aforementioned artistic workflows. The vGPU frame buffer is allocated out of the physical GPU frame buffer at the time the vGPU is assigned to the VM and the NVIDIA vGPU retains exclusive use of that frame buffer. All vGPUs resident on a physical GPU share access to the GPUs engines including the graphic 3D, video decode, and video encode engines. Since user behavior varies and is a critical factor in determining the best GPU and profile size, it is highly recommended to profile your own data and workflows during your PoC to properly size your environment for optional performance.

Chapter 5. Findings

Our testing showed that four artists from three unique parts of the pipeline can all effectively do their 3D production work using VMs. To determine the optimal configuration of Quadro vDWS to support these four artists, both user performance and scalability were considered. To further support this conclusion, NVIDIA collected insights from media and entertainment customers as well, to understand how animation studio customers are deploying Quadro vDWS. The HPE DL380 server configured with three Quadro RTX 6000 GPUs provided the necessary resources so that 3D production artists could work more efficiently, leveraging high-performance virtual workstations which perform just like physical workstations. When sizing a Quadro vDWS deployment for Autodesk Maya and Arnold, NVIDIA recommends conducting your own PoC to fully analyze resource utilization using objective measurements and subjective feedback. It is highly recommended that you install the GPU Profiler within your artist VMs to properly size your VMs.

Chapter 6. Deployment Best Practices

6.1 Run a Proof of Concept

The most successful deployments are those that balance user density (scalability) with performance. This is achieved when Quadro vDWS-powered virtual machines are used in production while objective measurements and subjective feedback from end users is gathered.

We highly recommend a PoC is run prior to doing a full deployment to provide a better understanding of how your users work and how many GPU resources they really need, analyzing the utilization of all resources, both physical and virtual. Consistently analyzing resource utilization and gathering subjective feedback allows for optimizing the configuration to meet the performance requirements of end users while optimizing the configuration for best scale.

Table A-1. Metrics for a Successful PoC Example

Objective Measurements	Subjective Feedback
Loading time of application	Overall user experience
Loading time of dataset	Application performance
Utilization (CPU, GPU, Networking	Zooming and panning experience

6.2 Leverage Management and Monitoring Tools

Quadro vDWS software provides extensive monitoring features enabling IT to better understand usage of the various engines of an NVIDIA GPU. The utilization of the compute engine, the frame buffer, the encoder, and decoder can all be monitored and logged through a command line interface called the NVIDIA System Management Interface (nvidia-smi), accessed on the hypervisor or within the virtual machine. In addition, NVIDIA vGPU metrics are integrated with Windows Performance Monitor (PerfMon) and through management packs like VMware vRealize Operations.

To identify bottlenecks of individual end users or of the physical GPU serving multiple end users, execute the following nvidia-smi commands on the hypervisor.

Virtual Machine Frame Buffer Utilization:

```
nvidia-smi vgpu -q -l 5 | grep -e "VM ID" -e "VM Name" -e "Total" -e "Used"
-e "Free"
```

Virtual Machine GPU, Encoder and Decoder Utilization:

```
nvidia-smi vgpu -q -l 5 | grep -e "VM ID" -e "VM Name" -e "Utilization" -e
"Gpu" -e "Encoder" -e "Decoder"
```

Physical GPU, Encoder and Decoder Utilization:

```
nvidia-smi -q -d UTILIZATION -l 5 | grep -v -e "Duration" -e "Number" -e
"Max" -e "Min" -e "Avg" -e "Memory" -e "ENC" -e "DEC" -e "Samples"
```

6.3 Understand Your Users

Another benefit of performing a PoC prior to deployment is that it enables more accurate categorization of user behavior and GPU requirements for each virtual workstation. Customers often segment their end users into user types for each application and bundle similar user types on a host. Light users can be supported on a smaller GPU and smaller profile size while heavy users require more GPU resources, a large profile size, and may be best supported on a larger GPU like the Quadro RTX 8000 for example.

6.4 Understanding the GPU Scheduler

NVIDIA Quadro vDWS provides three GPU scheduling options to accommodate a variety of QoS requirements of customers.

- **Fixed share scheduling:** Always guarantees the same dedicated quality of service.
  The fixed share scheduling policies guarantee equal GPU performance across all vGPUs sharing the same physical GPU. Dedicated quality of service simplifies a PoC since it allows the use of common benchmarks used to measure physical workstation performance such as SPECviewperf, to compare the performance with current physical or virtual workstations.
- **Best effort scheduling[1]:** Provides consistent performance at a higher scale and therefore reduces the TCO per user. This is the default scheduler.
  The best effort scheduler leverages a round-robin scheduling algorithm which shares GPU resources based on actual demand which results in optimal utilization of resources. This results in consistent performance with optimized user density. The best effort scheduling policy best utilizes the GPU during idle and not fully utilized times, allowing for optimized density and a good QoS.
- **Equal share scheduling:** Provides equal GPU resources to each running VM. As vGPUs are added or removed, the share of GPU processing cycles allocated changes accordingly, resulting in performance to increase when utilization is low, and decrease when utilization is high.

Organizations typically leverage the best effort GPU scheduler policy for their deployment to achieve better utilization of the GPU, which usually results in supporting more users per server with a lower quality of service (QoS) and better TCO per user.

Note:

[1]Available since 2013 when NVIDIA virtual GPU technology was first introduced.

Chapter 7. Summary

The HPE DL380 configured with three Quadro RTX 6000 GPUs provided the necessary resources for 3D production artists to work more efficiently, leveraging high performance virtual workstations which perform just like physical workstations. When sizing a Quadro vDWS deployment for Autodesk Maya and Arnold, NVIDIA recommends conducting your own PoC to fully analyze resource utilization using objective measurements and subjective feedback. NVIDIA RTX Server offers flexibility to IT administrators to size VMs based on workload or workflow needs.

Access NVIDIA vGPU software today by downloading a 90-day free trial evaluation. Or learn more about Quadro vDWS software on our product webpage.

Appendix A. Solution Configuration and Details

Table A-1 outlines the system configuration utilized to complete the rigorous NVIDIA NVQual verification along with the Autodesk Maya, Autodesk Arnold, and Teradici software packages all in line with the NVIDIA RTX Server validation process.

Table A-1. Solution Components

Components	Vendor, Model, and Quantity	Details
System	Hewlett Packard Enterprise ProLiant DL380 Gen10	CPU: 2x Intel Xeon Gold 6142 Memory: 384 GB DDR4-2933 Storage: Detached OS: Windows 10 / CentOS 7.7
Graphics	3x Quadro RTX 6000 Quadro driver release: 430 U2 [430.64] or later	GPU memory: 24 GB NVIDIA® CUDA® cores: 4,608 Tensor cores: 576 TensorRT cores: 72
Graphics software	NVIDIA Quadro Virtual Workstation Software	12GB frame buffer per user example: • GRID_RTX6000-12Q: 2 users 24GB frame buffer per user example: • GRID_RTX6000-24Q: 1 user • GRID_RTX6000-24Q: 1user
Hypervisor	VMware vSphere 6.7U1 or later	Enterprise Plus edition or higher: http://www.vmware.com/products/vsphere/compare.html https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/products/vsphere/vmware-vsphere-desktop-faqs.pdf
Application and software	Teradici Cloud Access Software Autodesk Maya 2020 Autodesk Arnold 6	

A.1 Server Recommendation: Dual Socket, 2U Rack Server

A 2RU, 2-socket server configured with two Intel Xeon Gold 6154 processors is recommended. With a high-frequency 3.0 GHz combined with 18-cores, this CPU is well-suited for optimal performance for each end user while supporting the highest user scale, making it a cost-effective solution for Autodesk Maya.

A.2 Flash Based Storage for Best Performance

The use of flash-based storage, such as solid-state drives (SSDs) are recommended for optimal performance. Flash-based storage is the common choice for users on physical workstations and similar performance can be achieved in similarly configured virtual environments.

A typical configuration for non-persistent virtual machines is to use the direct attached storage (DAS) on the server in a RAID 5 or RAID 10 configuration. For persistent virtual machines, a high performing all-flash storage solution is the preferred option.

A.3 Typical Networking Configuration for Quadro vDWS

There is no typical network configuration for in a Quadro vDWS powered virtual environment since this varies based on multiple factors including choice of hypervisor, persistent versus non-persistent virtual machines, and choice of storage solution. Most customers are using 10 GbE networking for optimal performance.

A.4 Optimizing for Dedicated Quality of Service

For comparative purposes, we considered the requirements for a configuration optimized for performance only. This configuration option does not take into account the need to further optimize for scale, or user density. Additionally, this configuration option is based solely on performance using the aforementioned sample 3D production artistic workflows.

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.

Trademarks

NVIDIA, the NVIDIA logo, CUDA, NVIDIA OptiX, NVIDIA RTX, NVIDIA Turing, Quadro, Quadro RTX, and TensorRT are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. Other company and product names may be trademarks of the respective companies with which they are associated.

Copyright

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