



NVIDIA RTX SERVER FOR VIRTUAL WORKSTATION WITH TERADICI CLOUD ACCESS SOFTWARE DESIGN GUIDE

VERSION 1.0

VERSION 1.0

TABLE OF CONTENTS

- Chapter 1. SOLUTION OVERVIEW 1**
- 1.1 RTX Server Solution Overview 1
- Chapter 2. SOLUTION DETAIL..... 2**
- 2.1 VALIDATED SERVER Configurations 4

Chapter 1.

SOLUTION OVERVIEW

NVIDIA RTX Server™ is a validated reference architecture consisting of qualified OEM server hardware with NVIDIA® Quadro® RTX™ 6000 or RTX 8000 GPUs and Quadro® Virtual Data Center Workstation (Quadro vDWS) software to provide virtualized design workstations where artists and designers can create professional, photorealistic images for the Media & Entertainment, Architecture, Engineering & Construction (AEC), and Manufacturing & Design industries.

1.1 RTX SERVER SOLUTION OVERVIEW

In today's media rich world, there is a massive surge of content production. Designers and artists are looking for ways to create faster and more efficiently anywhere on the device of their choice. IT looks to enable this in a secure, easily managed, cost-effective and flexible way. NVIDIA RTX Server is an ideal solution to increase user productivity while increasing data center utilization and reducing cost.

The audience for this document includes, but is not limited to, sales engineers, field consultants, professional services, partner engineers, and IT Managers who wish to take advantage of a solution that is purpose built and optimized to deliver a virtualized design and creation workflow.

Chapter 2.

SOLUTION DETAIL

NVIDIA RTX Server solution is a reference design comprised of (a) NVIDIA Quadro RTX 8000 or Quadro RTX 6000 graphics cards; (b) VMware vSphere Hypervisor with vCenter; (c) NVIDIA Quadro vDWS software; (d) Qualified OEM server system; (e) Thin Client device; (f) Teradici Cloud Access Software with PCoIP® Ultra. This validated solution provides unprecedented graphics performance in a virtualized environment on professional applications at a fraction of the cost, space, and power consumption to individual local workstations.

NVIDIA Quadro RTX 8000 and RTX 6000, powered by the NVIDIA Turing™ architecture and the NVIDIA RTX Server solutions, 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.

NVIDIA Quadro vDWS enables delivery of the most powerful virtual workstations from the data center or cloud to any device, anywhere. It lets IT virtualize any application from the data center with a native workstation user experience, eliminate constrained workflows, and flexibly scale GPU resources to run multiple workloads. Efficiently centralize applications and data for a dramatically lower IT operating expense and focus IT resources on managing strategic projects versus individual workstations – all while enabling a more secure, work-from-anywhere environment with reduced threat of data loss or leakage.

VMware vSphere provides a powerful, flexible, and secure foundation for business agility that accelerates your digital transformation. 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), securing applications, data, infrastructure, and access. Advanced security capabilities fully integrated into the hypervisor and powered by machine learning, provide better visibility, protection and faster response time for security incidents. vSphere helps you run, manage, connect and secure your applications in a common operating environment across the hybrid cloud.

Teradici Cloud Access Software with PCoIP® Ultra enhancements enables NVIDIA RTX servers to deliver highly interactive graphics to remote users without compromise. Leveraging NVIDIA NVENC encoding technology with security and performance benefits of the Teradici PCoIP protocol, graphic designers, artists, and CAD/CAM users can work productively over virtually any network.

RTX Servers, built by our OEM Partners, undergo NVIDIA's Qualification test suite. Among systems that qualify as an RTX Server there is a subset that has gone through additional testing and validation for VMware/Teradici workflow. These RTX Server Validated systems capture best practices from NVIDIA and its ecosystem partners.

Configurations for the Validated RTX Servers are listed in the below segment.

2.1 VALIDATED SERVER CONFIGURATIONS

For Media and Entertainment industry designers, frame buffer allocation can be changed depending on their workload size, by configuring the NVIDIA virtual GPU (vGPU) type on virtual machine.

Example: 16GB FB per user - GRID_RTX8000-16Q: 3 users [max. 9 users supported] or 12GB FB per user - GRID_RTX8000-12Q: 4 users [max. 12 users supported] / GRID_RTX6000-12Q: 2 users [max. 6 users supported]

Recommended Server Configuration:

- For High Performance: Dual Intel Xeon Gold 6126 processors: 2.6-3.7GHz; 12 Cores
For Higher User Density: Dual Intel Xeon Gold 6148 processors: 2.4-3.7GHz; 20 Cores
- 1.5TB Memory (128GB recommended for each user), system memory size might change based on customer's deployment configuration
- Network attached storage, SSD recommended for higher performance
- 10GbE network ports

VMWare/Teradici tests were run using:

- VMware vSphere 6.7U1 or later (requires enterprise license)
- vGPU 8.0+ drivers.
- Thin Client with i3-8100 or higher CPU and 8-32 GB DDR4 2 x SO-DIMM
- [Teradici Cloud Access Plus](#) (subscription based) Remote Access software
 - [Windows Graphics Agent](#)
 - [Linux Graphics Agent](#)
 - [Software clients](#)

Table 1: Validated Server Configurations

| Server Model | Graphics | Configuration |
|-----------------------------------|---|--|
| Advanced HPC Mercury RM408 | 4x Quadro RTX 8000 or RTX 6000 2x Quadro RTX NVLink High Bandwidth Bridge 2-slot | Dual Intel® Xeon® Gold 6126 processor: 2.6-3.7GHz; 12 Cores, 24 Threads 512 GB Memory 1.9 TB SSD |

SOLUTION DETAIL

| | | |
|--------------------------------------|---|--|
| AMAX QR-240 2U | 4x Quadro RTX 8000 or RTX 6000 2x Quadro RTX NVLink High Bandwidth Bridge 2-slot | Dual Intel® Xeon® Gold 6126 processor: 2.6-3.7GHz; 12 Cores, 24 Threads 512 GB Memory 1.9 TB SSD |
| ASUS ESC4000 G4 Series | 4x Quadro RTX 8000 or RTX 6000 2x Quadro RTX NVLink High Bandwidth Bridge 2-slot | Dual Intel® Xeon® Gold 6126 processor: 2.6-3.7GHz; 12 Cores, 24 Threads 1.5TB Memory 1.9 TB SSD |
| Exact TensorEX TS2-673917-RTX | 4x Quadro RTX 8000 or RTX 6000 2x Quadro RTX NVLink High Bandwidth Bridge 2-slot | Dual Intel® Xeon® Gold 6126 processor: 2.6-3.7GHz; 12 Cores, 24 Threads 512 GB Memory 1.9 TB SSD |
| Gigabyte G191-H44 | 4x Quadro RTX 6000 | Dual Intel Platinum 8176 processors 1.5TB Memory Dual 1 GbE network ports |
| Gigabyte W42G-P08R | 4x Quadro RTX 6000 | Dual Intel® Xeon® Platinum 8176 processor: 2.1GHz DDR4 up to 1.5TB Memory |
| HPE ProLiant DL380 Gen10 | 2x Quadro RTX 8000 or RTX 6000 | Dual Intel® Xeon® Gold 6226 processor: 2.7 GHz 512 GB HPE 2600 DDR4 2TB HPE SSD |
| Penguin Relion XE2114GT | 4x Quadro RTX 8000 or RTX 6000 2x Quadro RTX NVLink High Bandwidth Bridge 2-slot | Dual Intel® Xeon® Gold 6126 processor: 2.6-3.7GHz; 12 Cores, 24 Threads 512 GB Memory 1.9 TB SSD |
| RAVE-AS-RTX6000-4 | 4x Quadro RTX 6000 2x Quadro RTX NVLink High Bandwidth Bridge 2-slot | Dual Intel® Xeon® Gold 6126 processor: 2.6-3.7GHz; 12 Cores, 24 Threads 512 GB Memory 1.9 TB SSD |

SOLUTION DETAIL

| | | |
|---------------------------------|---|---|
| <p>RAVE-AS-RTX8000-4</p> | <p>4x Quadro RTX 8000 2x Quadro RTX NVLink High Bandwidth Bridge 2-slot</p> | <p>Dual Intel® Xeon® Gold 6126 processor: 2.6-3.7GHz; 12 Cores, 24 Threads 512 GB Memory 1.9 TB SSD</p> |
|---------------------------------|---|---|

Notice

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.

Information furnished is believed to be accurate and reliable. However, NVIDIA Corporation assumes no responsibility for the consequences of use of such information or for any infringement of patents or other rights of third parties that may result from its use. No license is granted by implication of otherwise under any patent rights of NVIDIA Corporation. Specifications mentioned in this publication are subject to change without notice. This publication supersedes and replaces all other information previously supplied. NVIDIA Corporation products are not authorized as critical components in life support devices or systems without express written approval of NVIDIA Corporation.

Trademarks

NVIDIA, the NVIDIA logo, and RTX Server, Turing, and Quadro 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

© 2019 NVIDIA Corporation. All rights reserved.