



NVIDIA RTX™ SERVER FOR VIRTUAL DATA SCIENCE WORKSTATION DESIGN GUIDE

VERSION: 1.0



TABLE OF CONTENTS

- Chapter 1. SOLUTION OVERVIEW 1**
- 1.1 NVIDIA RTX Server Overview..... 1
- Chapter 2. SOLUTION DETAILS 2**
- 2.1 VALIDATED SERVER Configurations 3

Chapter 1.

SOLUTION OVERVIEW

Designed and tested through multi-vendor cooperation between NVIDIA and its system partners, NVIDIA RTX™ Server combines the power of Quadro RTX GPUs with accelerated CUDA-X AI data science software to turn massive amounts of data into valuable predictions.

1.1 NVIDIA RTX SERVER OVERVIEW

Introduction:

Data is fundamentally changing the way companies do business. While this data can provide companies with valuable insights, processing and extracting the right information is a challenge. The time it takes to wrangle, prepare, and clean data from multiple data stores can also be significant. NVIDIA RTX® Server helps data scientists transform massive amounts of information into insights faster than ever before by bringing GPU acceleration to the entire data science workflow within a virtualized environment—from data preparation and model training to visualization.

Audience:

The audience for this document include, but not limited to: Sales Engineers, Field Consultants, Professional Services, Data Scientists, Partner Engineers, IT Managers and Customers who wish to take advantage of an appliance that is built and optimized to deliver on Data Science workflows.

Chapter 2.

SOLUTION DETAILS

NVIDIA RTX Server for Virtual Data Science Workstation is a reference design comprised of (a) NVIDIA Quadro RTX 8000 or RTX 6000 graphics cards; (b) VMware vSphere Hypervisor with vCenter; (c) NVIDIA Quadro vDWS software; (d) Qualified OEM server system. Combined, this validated solution provides powerful Data Science Workstations in a virtualized environment.

Bring the power of RTX to your data science workflow with NVIDIA Quadro RTX 8000 and RTX 6000. Get up to 96 GB of ultra-fast local memory to handle the largest datasets and compute-intensive workloads. Leverage the latest in ray-tracing technology for high performance local visualization. Maximize productivity, reduce time to insight and lower the cost of data science projects with virtualized workstations.

NVIDIA is bringing massive acceleration to the data science ecosystem through our CUDA-X AI architecture. CUDA-X AI is a collection of software-acceleration libraries built on top of CUDA, NVIDIA's groundbreaking parallel programming model, that provides essential optimizations for deep learning, machine learning, and compute. These libraries unleash the power of NVIDIA GPUs, giving developers the power to increase productivity while benefiting from continuous application performance gains in the underlying infrastructure. The ecosystem for data science built on CUDA-X AI is extensive, spanning across frameworks, cloud machine learning services, and deployments.

As such, NVIDIA is helping to create data science breakthroughs that completely change the timescales on which data science operates.

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 Data Science 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

Table 1 outlines the servers utilized to complete the NVIDIA RTX Server validation process.

Software Requirement:

- VMware vSphere ESXi hypervisor 6.7U3 or later (requires enterprise license)
- VMware vCenter (6.7 U3 or newer)
- vGPU 9.1+ drivers
- Linux (Ubuntu 18.04 LTS or RHEL Workstation 7.5/7.6) for guest VMs
- NVIDIA Quadro vDWS software for guest VMs

Table 1: Validated Server Configurations

| Server Model | Graphics | Configuration |
|----------------------------|--|--|
| ASUS ESC4000 G4 | 4x Quadro RTX 8000 or RTX 6000 2x Quadro RTX NVLink High Bandwidth Bridge 2-slot 96 GB FB per VM 2x GRID_RTX8000-48Q or 48 GB FB per VM 2x GRID_RTX6000-24Q per VM | Dual Intel® Xeon® Gold 6226 processor: 2.6-3.7GHz; 12 Cores, 24 Threads 1.5TB Memory 1.9 TB SSD 10GbE network ports |

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 DGX 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.