Accelerate Compute-Intensive Workloads with NVIDIA-Certified Systems

Whitepaper

Table of Contents

WP-10326-001_v01

Accelerate Compute-Intensive Workloads with NVIDIA-Certified Systems

Confidently choose performance-optimized hardware solutions.

Infrastructure for Modern Applications

Accelerated workloads abound across all industries, from the use of artificial intelligence for better customer engagement, data analytics for business forecasting, and advanced visualization for quicker product innovation. With the drive towards remote and flexible workplaces, the need for virtual desktops to be as powerful as physical desktops is also growing. Scientific researchers are innovating ways to solve problems that go beyond traditional CPU-only computing.

GPUs have eclipsed their initial role of accelerating graphics-related processes such as rendering and ray tracing and are now being used to deliver revolutionary speedups to a wide range of compute algorithms in areas such as machine learning, deep learning, high-performance computing (HPC), and data analytics. In fact, GPUs are now the leading compute accelerator, and provide better performance than any other technology for both training and inference of AI algorithms.

Network traffic within the data center is also rising. The digital transformation of the modern age has dramatically increased the volume of data available for developing actionable insights. Additionally, as cameras and other remote sensors proliferate at edge locations, secure, reliable, and high-speed networking is essential for moving the large amounts of data processed by these workloads.

To run these modern applications, enterprises need an easy way to deploy accelerated infrastructure. When a particular line of business is interested in purchasing an accelerated server or workstation, they care about the factors that directly impact user productivity and immediate business needs. These factors include performance and the ability to use a large set of developer tools and frameworks.

On the other hand, the team who's responsible for taking care of these systems cares about management, security, and how to scale out as demand grows. One of their biggest challenges is to ensure the systems are configured optimally while delivering performance. Some customers simply want to deploy an accelerated system that's validated for performance, scalability, and security so they can get up and running quickly, without additional tuning. Other administrators want to adjust the configuration for their specific use case in a manner that's consistent with best practices.

The NVIDIA-Certified Systems™ program was created to answer this need. Systems from leading system manufacturers are equipped with NVIDIA GPUs and network adapters and are subjected to rigorous testing. They're stamped as NVIDIA-Certified if they meet specific criteria for the best performance and scalability as well as proper functionality for security and management capabilities. This designation enables customers to get started quickly by simply deploying a certified system with the configuration that resulted in successful certification. NVIDIA provides guidance for how to tune the configuration of these systems to perform best on specific types of workloads. With an NVIDIA-Certified System, enterprises can confidently choose performance-optimized hardware solutions to power their accelerated computing workloads—from desktop to data center to edge.

Certified System Categories

The NVIDIA-Certified Systems program encompasses several system categories across data center, edge, and workstation. Each category is designed to be optimal for specific workloads, as shown in Table 1.

Table 1. Certified System Categories

System Class	System Category	Primary Use Case	Eligible NVIDIA GPUs
Data Center Servers	Compute	AI Training and Inferencing, Data Analytics, HPC	Recommender Systems, Natural Language Processing
	General Purpose	Visualization, Rendering, Deep Learning	Offline batch rendering, accelerating desktop rendering
	High Density Virtualization	Virtual Desktop, Virtual Workstation	Office Productivity, Remote Work, Visualization, Collaboration,
Edge	Enterprise Edge	Edge Inferencing in controlled environments	Image and video analytics, Virtual Radio Access Network (vRAN), Multi-access Edge Computing (MEC)
	Industrial Edge	Edge Inferencing in industrial or rugged environments	Robotics, Visual inspection systems, On-vehicle systems, medical instruments, Field deployed Telco Equipment
Workstations	Desktop Workstation	Design, Content Creation, Data Science	Product, building design, M&E content creation
	Mobile Workstation	Design, Content Creation, Data Science, Software Development	Data analytics data, feature exploration, solution development, Software design, development, build and testing

Server-class systems are also tested with NVIDIA networking hardware to ensure remote data access and multi-node scalability by running workloads that involve coordination and data transfer between hosts. The use of these devices provides key benefits for customers in performance, manageability, scalability, and security.

- Speeds of up to 200Gb/sec and the option to use InfiniBand provides the accelerated networking that's required to support today's advanced workloads.
- GPUDirect® technology allows efficient, zero-copy data transfer between GPUs and increased transfer capacity to storage, unlocking high-throughput, low-latency network connectivity and alleviating data bottlenecks in data center-scale computing clusters.
- Transport Layer Security (TLS) and Internet Protocol Security (IPsec) in-line cryptography are offloaded from the CPU, allowing these operations to be accelerated without impacting the high-bandwidth and low-latency communications needed for data-intensive workloads.
- Hardware root of trust enables security at the platform layer with secure boot and secure firmware updates.
- NVIDIA® ASAP2 - Accelerated Switching and Packet Processing® technology (ASAP2) handles a large portion of the packet processing operations in hardware, freeing up the host's CPU and providing high-throughput connectivity. Connection tracking (CT) offload capability enables stateful connection-based filtering, allowing unmatched performance, scale, and efficiency.

The complete list of certified systems is maintained on the NVIDIA-Certified Systems documentation page System certification status is also called out on the Qualified System Catalog.

Certification Process

The certification test suite is designed to exercise the performance and functionality of the candidate server by running a set of software that represents a wide range of real-world applications. This includes deep learning training; AI inference; end-to-end AI frameworks, including NVIDIA Riva and NVIDIA Clara™; data science, including Spark; intelligent video analytics (IVA); HPC and CUDA® functions; and rendering. Security features such as Trusted Platform Module (TPM) are also tested. Server-class systems are additionally subjected to tests of multi-node scalability by running workloads that involve coordination and data transfer between hosts using GPUDirect. Other tests cover network performance and remote management capabilities with Redfish.

Each of the tests in the suite is carefully chosen to exercise the hardware of the system in a unique and thorough manner so that as many potential configuration issues as possible can be exposed. Some of the tests focus on a single aspect of the hardware, such as the GPU decoder, while others stress multiple components, both simultaneously as well as in a multi-step workflow. It's important to realize that the test suite doesn't certify any particular

software application; the software is used only to exercise the hardware in a thorough manner. An effort is made to avoid having multiple tests that overlap in what they exercise.

Each category of the system runs a subset of the tests that are relevant to the use cases for it. For example, multi-node tests are not performed on workstations, and graphics tests are not run on compute-only GPUs. Table 2 summarizes the test coverage by category.

Table 2. Certification Process

System Category	Training, Data Analytics, HPC	Networking and Multi-node Training	Inference and AI Frameworks	Visualization and Rendering	Remote Management	Security
Compute Server	X	X	X		X	X
General-purpose Server	X	X	X	X	X	X
High Density Virtualization			X	X	X	X
Enterprise Edge	X		X	X	X	X
Industrial Edge	X		X	X		X
Desktop Workstation	X		X	X		X

The tests for each candidate system are performed by the system manufacturer in their labs. NVIDIA provides the following to partners who are looking to certify a system:

- Software containers loaded with application software along with associated datasets for sample runs
- Scripts that can automate the partner's entire test suite
- Various guides for setup and configuration of test systems

NVIDIA partners configure candidate systems according to the guidelines and then execute the test suite. For server certifications, two identical systems are used simultaneously so that multi-node tests can be run. These are additionally equipped with the latest networking smart network interface cards (SmartNICs) and data processing units (DPUs) from NVIDIA. Options include NVIDIA ConnectX®-6 and ConnectX-6 Dx, and NVIDIA BlueField®-2, and tests can be run with either Ethernet or InfiniBand networking. NVIDIA Spectrum™ or Quantum switches and LinkX® cables are used to create the test network so that there's a consistent baseline by which to assess results.

The system test suite takes between one to two days to run, based on the number and type of GPUs in the configuration. NVIDIA works with the partner to resolve any issues that are

highlighted during the test process and helps them determine what configuration changes are needed to achieve the best results.

Optimizing System Configuration

Over the course of the NVIDIA-Certified Systems program, NVIDIA has studied hundreds of test results across many server models and compiled a unique database of detailed test results. NVIDIA architects use this database to determine best practices for configuring systems to maximize performance and create a baseline of performance standards.

NVIDIA-Certified Systems are designed to maximize the performance of modern applications with both compute and input/output (IO) acceleration. The design objective is a well-balanced system in which performance bottlenecks caused by any single component are minimized and which a wide variety of algorithms and workflows can be run as fast as possible for the specific hardware. The certification process identifies the best ways to configure components such as:

- RAM size
- PCI port selection
- Temperature and fan speed curve
- NUMA topology
- BIOS / firmware settings
- Network switch settings (for multi-node performance)
- Versions of OS, libraries, drivers

Misconfiguration of these components can lead to poor performance and the inability to function properly or complete tasks. Some examples of these configuration issues are listed below.

High Operating Temperature

The performance of a GPU can be affected by the operating temperature. Although NVIDIA GPUs have a maximum temperature below which their usage is supported, certification testing has shown that operating at a lower temperature can, in some cases, greatly improve performance.

A typical system has multiple fans to provide air cooling, but the amount of cooling for each device in the enclosure depends greatly on the physical layout of all the components, especially the location of GPUs with respect to fans, baffles, dividers, risers, etc. Many enterprise systems have programmable fan curves, which specify the fan speed as a function of GPU temperature for each fan. Oftentimes, the default fan curve is based on a generic base system and does not account for the presence of GPUs and similar devices that can produce a lot of heat.

In one example of a system with four GPUs, certification testing revealed that one of the GPUs was operating at a much higher temperature than the other three. This was simply due to the specific internal layout of components and airflow characteristics in that particular model. It

wasn't something that could have been anticipated. By adjusting the fan curve, the hot spot was eliminated, and the overall performance of the system was improved.

Since systems can vary greatly in their physical design, there's no universal fan curve profile that can be recommended. Instead, the certification process is invaluable for identifying potential performance issues due to temperature and can validate which fan curves result in the best results for each server being tested. These profiles are documented for each certified system.

Non-optimal BIOS and Firmware Settings

Both BIOS settings and firmware versions can impact performance as well as functionality. This is particularly the case for NUMA-based systems. The certification process determines the optimal BIOS settings for best performance and identifies the best values for other configurations, such as NIC PCI settings and boot grub settings. Multi-node testing has also determined the optimal settings for the network switch. In one example, a system was achieving RDMA communication close to 300Gb/s, and TCP at 120Gb/s. After the settings were properly configured, the performance for RDMA increased to 360Gb/s and TCP to 180Gb/s, which were both nearly line rate.

Improper PCI Slot Configuration

Rapid transfer of data to and from the GPU is critical to getting the best performance on accelerated workloads. In addition to the need to move large amounts of data to the GPU for both training and inferencing, as described above, movement of data between GPUs during the so-called all-reduce phase of multi-GPU training can become a bottleneck. This is also the case for the network interface since data is often loaded from remote storage, or in the case of multi-node algorithms, transferred between systems. Because GPUs and NICs are installed on a system via the PCI bus, improper placement can result in suboptimal performance.

NVIDIA GPUs use 16 PCIe lanes (referred to as x16), which enable 16 parallel data transfer channels. NVIDIA NICs can use 8, 16, or 32 lanes, depending upon the particular model. In a typical server or workstation, the PCI bus is divided into slots with differing numbers of lanes to account for the needs of different peripheral devices. In some cases, this is further affected by the use of a PCI riser card, and the number of slots can also be adjusted in the BIOS. If a GPU or NIC is installed in the motherboard without taking these factors into account, the full capacity of the device might not be used. For example, an x16 device might be installed in an x8 slot, or the slot might be limited to x8 or less by a BIOS setting. The certification process exposes these issues, and the optimal PCI slot configuration is documented when a system is certified.

Lack of NUMA Topology Awareness

NUMA (Non-Uniform Memory Architecture) is a particular design for configuring microprocessors in a multi-CPU system used by certain chip architectures. In these types of systems, devices such as GPUs and NICs have an affinity to one specific CPU because they're

connected to the bus associated with that CPU (in a so-called NUMA node). When running applications that involve communication between GPUs or between GPU and NIC, the performance can be greatly reduced if the devices are not optimally paired. In the worst-case scenario, data will need to traverse between NUMA nodes, resulting in high latency.

The certification process ensures that the design of the system is balanced, with GPUs spread evenly across CPU sockets and PCIe root ports and the NIC under the same PCIe switch or PCIe root complex as the GPUs.

Impact of Insufficient System Memory (RAM)

Lack of sufficient memory is often a source of underperformance in numerous applications, and this is especially true for machine learning, both training and inference. In training, an algorithm typically analyzes huge batches of data, and the system should be able to hold enough data in memory to keep the training algorithm running. For inferencing, the memory requirement depends upon the use case. With batch inferencing, the more data that can be held in memory, the quicker it can be processed. However, with streaming inferencing, the data is typically analyzed as it comes in, and therefore the amount of memory needed might not be as great.

As a result of analyzing the results of many certification tests, NVIDIA has been able to come up with memory-sizing guidelines based on the number of GPUs and amount of GPU memory. In one case, a system with four GPUs that had 128GB RAM installed didn't pass the certification test. Upon increasing the RAM to 384GB, the overall performance increased by 45%, and the system was able to pass certification.

Certified Server Configurations

Once a system has passed the certification test suite, the configuration is documented as the base configuration. Customers can purchase this system and use it with this base configuration, or they can further tune it to their particular use case. It's important to take into account not only the applications to be run on the system but also the environment in which it will be installed. A system used for large model AI training would benefit from having multiple high-performance NVIDIA A100 Tensor Core GPUs and higher-speed networking provided by NVIDIA ConnectX-6 Dx adapters. By contrast, a system to be used for inferencing in a remote data center might be limited both by form factor and power consumption, as well as network infrastructure. In this case, a system with a lower-powered GPU, such as an NVIDIA T4 Tensor Core GPU, and networking with the NVIDIA ConnectX-6 Lx adapter would be an appropriate selection.

NVIDIA provides guidance for how to configure NVIDIA-Certified Systems for various workload classifications in a manner that conforms to the best practices determined by extensive testing. For example, if a system has one GPU per CPU, then a PCIe switch is not necessary. However, with two or more GPUs per CPU, a PCIe switch is likely necessary to achieve optimal performance, except in the case of edge inferencing, where a low-cost server without a PCIe switch provides sufficient capabilities.

Note that the certification applies to the specific combination of server model and GPU model family. Certain NVIDIA GPUs can be grouped together into a model family because they differ from each other in only limited ways. A server that has been certified with one member of the family is considered certified with the other members. Here are some examples of GPU model families.

- A100 PCIe 40GB, A100 PCIe 80GB, A30
- HGX A100 4-GPU 40GB, HGX A100 4-GPU 80GB
- HGX A100 8-GPU 40GB, HGX A100 8-GPU 80GB

The Essential Platform for Enterprise IT

The NVIDIA-Certified Systems program has assembled the industry's most complete set of accelerated workload performance tests to deliver the highest performing systems. Configurations are validated for optimum performance, reliability, and scalability for a diverse range of workloads.

With NVIDIA-Certified Systems, enterprises can confidently choose and configure performance-optimized servers and workstations to power their accelerated computing workloads, both in smaller configurations and at scale. They create the essential platform for the evolution of enterprise IT, and they provide the easiest way for customers to be successful with all their accelerated computing needs. A wide variety of system types are available, including popular data center and edge server models, as well as desktop and mobile workstations from NVIDIA's vast ecosystem of partners.

To learn more, refer to the following resources:

- NVIDIA-Certified Systems webpage
- NVIDIA-Configuration Guide
- Qualified System Catalog

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 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, ASAP2 – Accelerated Switch and Packet Processing, BlueField, Clara, ConnectX, CUDA, GPUDirect, HGX, LinkX, NVIDIA-Certified Systems, RTX, Spectrum, and Turing are trademarks and/or registered trademarks of NVIDIA Corporation and affiliates in the U.S. and other countries. Other company and product names may be trademarks of the respective companies with which they are associated.

DOCUMENT

Copyright

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