Deploy NGC on On-Premises Servers

Documentation Introduction
Table of Contents

Chapter 1. NGC Software is Certified on On-Premises Systems and the Cloud....................1
Chapter 2. NVIDIA-Certified Systems................................................................................. 2
Chapter 3. NGC-Ready Servers......................................................................................... 3
Chapter 1.  NGC Software is Certified on On-Premises Systems and the Cloud

NVIDIA certification programs validate the performance of AI, ML and DL workloads using NVIDIA GPUs on data center and edge servers and public clouds. It allows server manufacturers and public clouds to qualify their NVIDIA GPU equipped systems on a wide variety of AI workloads ranging from training to inference on on-premise servers and public cloud infrastructures.
Chapter 2. NVIDIA-Certified Systems

NVIDIA is working with its partners to certify the industry's most advanced AI systems running the latest NVIDIA GPUs—both as single-node servers and in cluster configurations using Mellanox smart network adapters and Mellanox Quantum HDR InfiniBand and Mellanox Spectrum® Ethernet switches.

The NVIDIA-Certified Systems program is the evolution of the NGC-Ready program that tests single node servers against a diverse range of workloads. The NVIDIA-Certified Systems program introduces new tests of multi-node deep learning training, distributed training and inference workloads, high-speed remote storage access, and cluster health, performance, and security tests.

- For the current list of NVIDIA-Certified systems, visit [NVIDIA Certified Systems](http://docs.nvidia.com/ngc/ngc-deploy-on-premises/nvidia-certified-systems/index.html)
- For guidance on system configurations needed to be considered NVIDIA-Certified, visit [NVIDIA-Certified Systems Configuration Guide](http://docs.nvidia.com/ngc/ngc-deploy-on-premises/nvidia-certified-configuration-guide/index.html)
Chapter 3. NGC-Ready Servers

The NGC-Ready program validates the performance of AI, ML and DL workloads using NVIDIA GPUs on leading on-premises servers. It allows server manufacturers to qualify their NVIDIA GPU equipped systems on a wide variety of AI workloads ranging from training to inference on on-premise servers, cloud infrastructure and edge devices.

Servers certified as NGC-Ready for Edge undergo additional testing for the security and remote system management capabilities required for remote deployment at the edge.

Performance-validated NGC-Ready systems accelerate time to solution, increase business agility, and are backed by enterprise-grade support, reducing the risk of deploying complex AI solutions.

‣ For the current list of NGC-Ready systems, visit NGC-Ready Systems

DGX Systems

All NGC content is tested on NVIDIA DGX systems, NVIDIA’s best-in-class portfolio of purpose-built AI systems. To learn more about NVIDIA DGX systems visit: https://www.nvidia.com/en-us/data-center/dgx-systems/.

Deploying NGC on a Public Cloud

For documentation on deploying NGC on a public cloud, visit Deploying NGC on a Public Cloud
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 and the NVIDIA logo are trademarks and/or registered trademarks of NVIDIA Corporation in the United States and other countries. Other company and product names may be trademarks of the respective companies with which they are associated.

Copyright

© 2022 NVIDIA Corporation & affiliates. All rights reserved.

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