



# DATACENTER GPU MANAGER 2.2

v2.2 | July 2021

**Release Notes**



# TABLE OF CONTENTS

- Changelog..... iii
- Patch Releases..... iii
  - DCGM v2.2.9..... iii
  - DCGM v2.2.8..... iii
- DCGM v2.2 GA..... iv
  - New Features..... iv
  - Improvements..... iv
  - Bug Fixes..... iv
  - Known Issues..... v

# CHANGELOG

This version of **DCGM** (v2.2) requires a minimum R418 driver that can be downloaded from [NVIDIA Drivers](#). On NVSwitch based systems such as DGX A100 or HGX A100, a minimum of Linux R450 ( $\geq 450.80.02$ ) driver is required. If using the new profiling metrics capabilities in DCGM, then a minimum of Linux R418 ( $\geq 418.87.01$ ) driver is required. It is recommended to install the latest datacenter driver from NVIDIA drivers downloads site for use with DCGM.

## Patch Releases

### DCGM v2.2.9

DCGM v2.2.9 released in July 2021.

#### Improvements

- ▶ Added support for the NVIDIA A100 80GB-PCIe product.
- ▶ Added support for the NVIDIA RTX A5000 product.
- ▶ The `--plugin-path` option is no longer supported by DCGM Diagnostics (`dcgmi diag`) and has been removed.

#### Bug Fixes

- ▶ Fixed an issue where a DCGM exception (`DcgmException`) could result in a crash in some environments.

### DCGM v2.2.8

DCGM v2.2.8 released in July 2021.

#### Improvements

- ▶ DCGM now supports multiple independent clients (`libdcgm.so`) to connect and obtain GPU telemetry including profiling metrics.
- ▶ Added support to obtain GPU telemetry on systems that may have different GPU SKUs.
- ▶ Added support for the NVIDIA RTX A6000 product.

- ▶ Improved the error logging for `dcgmproftester` when run in incorrect configurations, such as non-administrator privileges when requesting profiling metrics.

### Bug Fixes

- ▶ Fixed an issue where the DCGM packages did not include the `dcgm_api_export.h` header.

### Known Issues

- ▶ [Profiling metrics](#) are currently not supported on Arm64 server systems. This feature will be enabled in a future release of DCGM.

## DCGM v2.2 GA

DCGM v2.2.3 released in May 2021.

## New Features

### General

- ▶ Added support for NVIDIA A10 and A30 products.
- ▶ Added Python3 support in DCGM for the bindings.
- ▶ Profiling metrics are now supported on all [NVIDIA datacenter/enterprise GPUs](#).

## Improvements

- ▶ Reduced the CPU overhead of Profiling metrics fields 1001-1012.
- ▶ Added `--log-level` and `--log-filename` parameters to `dcgmproftester`.
- ▶ DCGM installer packages no longer include `libnvidia-nscq-dcgm.so.450.51.06` version of NSCQ. Customers are advised to install the latest NSCQ packages from the CUDA network repository for use with DCGM.
- ▶ DCGM diagnostics now checks for row remapping failures on NVIDIA Ampere GPUs.
- ▶ DCGM libraries do not expose symbols other than `dcgm*` API. This should fix issues for users who link with DCGM and protobuf libraries at the same time.
- ▶ DCGM has reduced logging verbosity in some situations, resulting in a smaller log file size footprint.

## Bug Fixes

- ▶ Added better error message reporting when `dcgmi diag` cannot find the NVVS binary at the expected installed locations on the system.
- ▶ Fixed an issue where NVVS error reports would not include the GPU id.
- ▶ Fixed an issue with DCGM detecting GPU brands. DCGM has been updated to be consistent with new brand strings returned by NVML (e.g. "NVIDIA T4" vs. "Tesla T4"). This issue may manifest as errors when trying to obtain profiling

metrics with the following message: **Error setting watches. Result: Profiling is not supported for this group of GPUs or GPU**

- ▶ Fixed an issue with the output of `dcgmi discovery -v -i c` where clocks were reported incorrectly.
- ▶ DCGM now reports an error when attempting to run diagnostics on a system with GPUs where some are in MIG mode.
- ▶ Fixed a bug in the DCGM Diagnostic where some valid parameters were being rejected as invalid when specified in the configuration file.

## Known Issues

- ▶ On DGX-2/HGX-2 systems, ensure that `nv-hostengine` and the Fabric Manager service are started before using `dcgmprof tester` for testing the new profiling metrics. See the Getting Started section in the DCGM User Guide for details on installation.
- ▶ On K80s, `nvidia-smi` may report hardware throttling (`clocks_throttle_reasons.hw_slowdown = ACTIVE`) during DCGM Diagnostics (Level 3). The stressful workload results in power transients that engage the HW slowdown mechanism to ensure that the Tesla K80 product operates within the power capping limit for both long term and short term timescales. For Volta or later Tesla products, this reporting issue has been fixed and the workload transients are no longer flagged as "HW Slowdown". The NVIDIA driver will accurately detect if the slowdown event is due to thermal thresholds being exceeded or external power brake event. It is recommended that customers ignore this failure mode on Tesla K80 if the GPU temperature is within specification.
- ▶ To report NVLINK bandwidth utilization DCGM programs counters in the HW to extract the desired information. It is currently possible for certain other tools a user might run, including `nvprof`, to change these settings after DCGM monitoring begins. In such a situation DCGM may subsequently return errors or invalid values for the NVLINK metrics. There is currently no way within DCGM to prevent other tools from modifying this shared configuration. Once the interfering tool is done a user of DCGM can repair the reporting by running `nvidia-smi nvlink -sc 0bz; nvidia-smi nvlink -sc 1bz`.

## Notice

THE INFORMATION IN THIS GUIDE AND ALL OTHER INFORMATION CONTAINED IN NVIDIA DOCUMENTATION REFERENCED IN THIS GUIDE IS PROVIDED “AS IS.” NVIDIA MAKES NO WARRANTIES, EXPRESSED, IMPLIED, STATUTORY, OR OTHERWISE WITH RESPECT TO THE INFORMATION FOR THE PRODUCT, AND EXPRESSLY DISCLAIMS ALL IMPLIED WARRANTIES OF NONINFRINGEMENT, MERCHANTABILITY, AND FITNESS FOR A PARTICULAR PURPOSE. Notwithstanding any damages that customer might incur for any reason whatsoever, NVIDIA’s aggregate and cumulative liability towards customer for the product described in this guide shall be limited in accordance with the NVIDIA terms and conditions of sale for the product.

THE NVIDIA PRODUCT DESCRIBED IN THIS GUIDE IS NOT FAULT TOLERANT AND IS NOT DESIGNED, MANUFACTURED OR INTENDED FOR USE IN CONNECTION WITH THE DESIGN, CONSTRUCTION, MAINTENANCE, AND/OR OPERATION OF ANY SYSTEM WHERE THE USE OR A FAILURE OF SUCH SYSTEM COULD RESULT IN A SITUATION THAT THREATENS THE SAFETY OF HUMAN LIFE OR SEVERE PHYSICAL HARM OR PROPERTY DAMAGE (INCLUDING, FOR EXAMPLE, USE IN CONNECTION WITH ANY NUCLEAR, AVIONICS, LIFE SUPPORT OR OTHER LIFE CRITICAL APPLICATION). NVIDIA EXPRESSLY DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY OF FITNESS FOR SUCH HIGH RISK USES. NVIDIA SHALL NOT BE LIABLE TO CUSTOMER OR ANY THIRD PARTY, IN WHOLE OR IN PART, FOR ANY CLAIMS OR DAMAGES ARISING FROM SUCH HIGH RISK USES.

NVIDIA makes no representation or warranty that the product described in this guide will be suitable for any specified use without further testing or modification. Testing of all parameters of each product is not necessarily performed by NVIDIA. It is customer’s sole responsibility to ensure the product is suitable and fit for the application planned by customer and to do 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 guide. NVIDIA does not accept any 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 guide, or (ii) customer product designs.

Other than the right for customer to use the information in this guide with the product, no other license, either expressed or implied, is hereby granted by NVIDIA under this guide. Reproduction of information in this guide is permissible only if reproduction is approved by NVIDIA in writing, is reproduced without alteration, and is accompanied by all associated conditions, limitations, and notices.

## 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

© 2013-2021 NVIDIA Corporation. All rights reserved.