



TESLA DRIVER VERSION 445 (LINUX AND WINDOWS)

RN-08625-445 _v01 | February 2020

Release Notes



TABLE OF CONTENTS

| | |
|---|----------|
| Chapter 1. Version Highlights | 1 |
| 1.1. QS Update..... | 1 |
| 1.2. Update 2..... | 3 |
| 1.3. Update 1..... | 3 |
| 1.4. Initial Release..... | 4 |
| 1.4.1. New Features..... | 4 |
| 1.4.2. Known Issues..... | 5 |
| Chapter 2. Hardware and Software Support | 6 |

Chapter 1.

VERSION HIGHLIGHTS

This section provides highlights of the NVIDIA Tesla 445 Driver for Linux. This driver is being released in conjunction with the HGX-3 hardware milestones.

For changes related to the 445 release of the NVIDIA display driver, review the file "NVIDIA_Changelog" available in the .run installer packages.

1.1. QS Update

2/26/2020: For this release, the changelog is listed below.

Software Versions

- ▶ CUDA Toolkit: 11.0.104
- ▶ NVIDIA Tesla Driver: 445.34 (Linux) and 445.61 (Windows)
- ▶ Fabric Manager: 445.34 (Use `nv-fabricmanager -v`)
- ▶ GPU VBIOS: 92.00.01.27.04
- ▶ NVSwitch VBIOS: 92.10.04.00.02
- ▶ NVFlash: 5.601.0

New Features

- ▶ Added support for Multi-Instance GPU (MIG) on A100.
- ▶ Added support for CUDA 11.0. See the CUDA Toolkit 11.0 release notes for more information.
- ▶ Improved the time to train NVLinks on HGX-3 systems
- ▶ Added the ability for NVLink to switch to single lane mode to save power at idle
- ▶ Added the ability to automatically configure the droopy value for A100 SXM4 modules based on the input voltage
- ▶ Added new SMBPBI query GPU power cap values
- ▶ This driver release includes changes to improve security when A100 is in device passthrough

- ▶ Added the ability for CUDA workloads to continue running even when the fabric manager is stopped or terminated
- ▶ Added new SMBPBI output for reporting failures when performing row-remapping updates
- ▶ Added new SMBPBI query for reporting row-remapping pending state
- ▶ Added support to SMBPBI to report GPU XID errors
- ▶ NVML Updates
 - ▶ New error code `NVML_ERROR_INSUFFICIENT_RESOURCES` when the GPU runs out critical resources, other than memory
 - ▶ Added new counters for NVLink, including flit error counters, CRC data errors, replay errors, bandwidth, speed, throughput and row remapper. See the NVML API documentation for more information on these new counters.
 - ▶ Added support for MIG configuration and management. See the NVML APIs for setting up MIG mode, managing GPU and Compute instances.
 - ▶ Added `nvmlDeviceGetRemappedRows` to return the number of rows remapped by the A100 GPU.
 - ▶ The following NVLink APIs are deprecated.
 - ▶ `nvmlDeviceSetNvLinkUtilizationControl()`
 - ▶ `nvmlDeviceGetNvLinkUtilizationControl()`
 - ▶ `nvmlDeviceGetNvLinkUtilizationCounter()`
 - ▶ `nvmlDeviceFreezeNvLinkUtilizationCounter()`
 - ▶ `nvmlDeviceResetNvLinkUtilizationCounter()`
 - ▶ `nvmlDeviceGetNvLinkUtilizationCounter()` is now replaced by `nvmlDeviceGetFieldValues()`. See the NVML API documentation for the new field parameters.

Known Issues

- ▶ Some CUDA sample applications (`cudaGraphs`, `cuSolverDn_LinearSolver`, `conjugateGradientMultiDeviceCG`, `p2pBandwidthLatencyTest`, `cuSolverSP_LinearSolver`, `cuSolverSp_LowlevelCholesky`) may have issues on RHEL 8.1 and Windows Server 2019. These issues manifest as either incorrect results or hangs while running the applications.
- ▶ By default, Fabric Manager runs as a systemd service. If using `DAEMONIZE=0` in the Fabric Manager configuration file, then the following steps may be required.
 1. Disable FM service from auto starting. (`systemctl disable nvidia-fabricmanager`)
 2. Once the system is booted, manually start FM process. (`/usr/bin/nv-fabricmanager -c /usr/share/nvidia/nvswitch/fabricmanager.cfg`).
Note, since the process is not a daemon, the SSH/Shell prompt will not be

returned (use another SSH shell for other activities or run FM as a background task).

1.2. Update 2

1/21/2020: For this release, the changelog is listed below.

Software Versions

- ▶ CUDA Toolkit: 11.0.70
- ▶ NVIDIA Tesla Driver: 445.20 (Linux) and 445.49 (Windows)
- ▶ Fabric Manager: 445.20 (Use `nv-fabricmanager -v`)
- ▶ GPU VBIOS: 92.00.01.27.04
- ▶ NVSwitch VBIOS: 92.10.04.00.02
- ▶ NVFlash: 5.601.0

New Features

- ▶ Added support for additional operating system distributions. See this [section](#) for more information.

Known Issues

- ▶ Some CUDA sample applications (cudaGraphs, cuSolverDn_LinearSolver, conjugateGradientMultiDeviceCG, p2pBandwidthLatencyTest, cuSolverSP_LinearSolver, cuSolverSp_LowlevelCholesky) may have issues on RHEL 8.1 and Windows Server 2019. These issues manifest as either incorrect results or hangs while running the applications.
- ▶ By default, Fabric Manager runs as a systemd service. If using `DAEMONIZE=0` in the Fabric Manager configuration file, then the following steps may be required.
 1. Disable FM service from auto starting. (`systemctl disable nvidia-fabricmanager`)
 2. Once the system is booted, manually start FM process. (`/usr/bin/nv-fabricmanager -c /usr/share/nvidia/nvswitch/fabricmanager.cfg`). Note, since the process is not a daemon, the SSH/Shell prompt will not be returned (use another SSH shell for other activities or run FM as a background task).

1.3. Update 1

1/7/2020: For this release, the changelog is listed below.

Software Versions

- ▶ NVIDIA Driver: 445.16
- ▶ Fabric Manager: 445.16 (Use `nv-fabricmanager -v`)
- ▶ GPU VBIOS: 92.00.01.27.03
- ▶ NVSwitch VBIOS: 92.10.04.00.02
- ▶ NVFlash: 5.601.0

New Features

- ▶ Added support for Fabric Manager service to be run from a non-root account.
- ▶ Added support for CentOS 7.7.
- ▶ Added support power cap configuration changes via SMBPBI.
- ▶ SMBus now supports an increased data rate of 400KHz.

Fixed Issues

- ▶ Fixed a crash with the Fabric Manager when starting up the `nv-fabricmanager` service with the configuration file. The new Fabric Manager software requires an update to the GPU VBIOS, NVSwitch VBIOS and the NVIDIA driver on the system.

Known Issues

- ▶ By default, Fabric Manager runs as a systemd service. If using `DAEMONIZE=0` in the Fabric Manager configuration file, then the following steps may be required.
 1. Disable FM service from auto starting. (`systemctl disable nvidia-fabricmanager`)
 2. Once the system is booted, manually start FM process. (`/usr/bin/nv-fabricmanager -c /usr/share/nvidia/nvswitch/fabricmanager.cfg`).
Note, since the process is not a daemon, the SSH/Shell prompt will not be returned (use another SSH shell for other activities or run FM as a background task).

1.4. Initial Release

12/18/2019: For this release, the software versions are listed below.

- ▶ NVIDIA Driver: 445.09
- ▶ GPU VBIOS: 92.00.01.1B.08
- ▶ NVFlash: 5.599.0

1.4.1. New Features

- ▶ Added a new SMPBI command to read the default value and range of possible max settings for the application and memory clocks.

- ▶ Added a new SMBPBI command to read the current value of the application and memory clocks.
- ▶ Fabric Manager is now available as via standalone installer packages (`.deb/.rpm`) for Linux distributions (`nvidia-fabricmanager-driver-version`)
- ▶ Added a number of performance improvements for out-of-band management using SMBPBI
 - ▶ Merged SMBPBI status and data into one request with 3-bytes (24 bits) of data per read
 - ▶ Bundled GPU telemetry using user selectable commands (such as GPU power, GPU temperature, HBM temperature) thereby reducing the number of reads required
 - ▶ Increased data rate over SMBUS from 100KHz to 400KHz
- ▶ Added a new SMBPBI command to read HBM temperature.
- ▶ Added support for systems that include 2x 8 GPU baseboards (up to 16 GPUs)

1.4.2. Known Issues

- ▶ There may be issues when running the fabric manager service from a non-root account
- ▶ While baseboards with single or two GPUs populated are expected to be functional, this configuration has not been fully validated
- ▶ CentOS 7.7 support has not been fully validated and may have issues
- ▶ Power cap configuration changes via SMBPBI have not been fully validated
- ▶ Using SMBPBI to read the bundled GPU telemetry may sometimes fail with an error
- ▶ Increased data rate of 400KHz over SMBus has not been fully validated and may have issues in this release of the driver

Chapter 2.

HARDWARE AND SOFTWARE SUPPORT

Supported Operating Systems and CPU Configurations

The Release 445 driver is validated with HGX-2 Next on the following operating systems and CPU configurations:

- ▶ Linux 64-bit distributions:
 - ▶ Red Hat Enterprise Linux 8.1 (in 8-GPU configurations)
 - ▶ CentOS Linux 7.7 (in 4-GPU configurations)
 - ▶ Ubuntu 18.04.3 LTS (in 16-GPU configurations)
- ▶ Windows 64-bit distributions:
 - ▶ Windows Server 2019 (in 8-GPU configurations)
- ▶ CPU Configurations:
 - ▶ AMD Rome in PCIe Gen4 mode
 - ▶ Intel Skylake/Cascade Lake in PCIe Gen3 mode

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

© 2020 NVIDIA Corporation. All rights reserved.