



TECHNICAL
WHITE PAPER

CONTENTS & NAVIGATION

2

Introduction

2

Software Considerations

2-3

Hardware Considerations

3-4

Operating System Configuration

4

Monitoring NVIDIA® NVLink™ Status



USING HP WORKSTATIONS WITH
NVIDIA® NVLINK™ FOR NVIDIA®
QUADRO® GRAPHICS CARDS



TECHNICAL WHITE PAPER

CONTENTS & NAVIGATION

2

Introduction

2

Software Considerations

2-3

Hardware Considerations

3-4

Operating System Configuration

4

Monitoring NVIDIA® NVLink™ Status

INTRODUCTION



HP Workstations are fully equipped to allow users to take advantage of a new NVIDIA® Quadro® technology – NVIDIA® NVLink™. In a nutshell, NVIDIA® NVLink™ allows two matching graphics cards to communicate directly via a high-speed interconnection effectively allowing the available memory of both GPUs to be combined. This allows for applications and workloads designed with NVIDIA® NVLink™ in mind to benefit from significant performance increases.

For more information, please review NVIDIA's documentation and examples:

- https://www.nvidia.com/object/NVIDIA_NVLink.html
- <https://blogs.nvidia.com/blog/2015/03/17/pascal/>

SOFTWARE CONSIDERATIONS

To take advantage of NVIDIA® NVLink™ software must be explicitly designed to do so. Unlike some other multi-GPU methods, NVIDIA® NVLink™ is different in that simply connecting NVIDIA® NVLink™ bridges does not automatically benefit a workload or use case; software must be designed with NVIDIA® NVLink™ in mind and written in a way to take full advantage of it.

NVIDIA provides free SDKs and documentation which enable a software developer to include NVIDIA® NVLink™ support in their applications. Specifically, NVIDIA® CUDA® SDK and NCCL are minimum requirements for NVIDIA® NVLink™ enablement in software. These resources can be found here:

- <https://developer.nvidia.com/cuda-toolkit>
- <https://docs.nvidia.com/cuda/>
- <https://developer.nvidia.com/nccl>

In addition to having compatible software, NVIDIA® NVLink™-capable hardware must be correctly installed and configured to take advantage of NVIDIA® NVLink™.

HARDWARE CONSIDERATIONS

NVIDIA® NVLink™ is supported with HP Z8 G4, HP Z4 G4, and HP Z840 Workstations that have been properly configured.

To prepare your system for use with NVIDIA® NVLink™ the following hardware must be installed in a supported system:

- **2x NVIDIA® NVLink™-capable NVIDIA® Quadro® cards, such as NVIDIA® Quadro® GP100 or NVIDIA® Quadro® GV100**
 - Please note, GPU generation mixing is not supported; the cards must be exact matches (ie. 2x NVIDIA® Quadro® GV100)
 - NVIDIA® Quadro® NVIDIA® NVLink™ only works between exactly 2 cards (ie. 3-way NVIDIA® NVLink™ is not possible)



CONTENTS & NAVIGATION

2

Introduction

2

Software Considerations

2-3

Hardware Considerations

3-4

Operating System Configuration

4

Monitoring NVIDIA® NVLink™ Status

- **2x NVIDIA® NVLink™ bridges that match the NVIDIA® Quadro® cards installed**
 - Please note, bridge generation mixing is **not** supported; the bridges must be exact matches (ie. 2x NVIDIA® NVLink™ bridges for NVIDIA® Quadro® GV100 connected to 2x NVIDIA® Quadro® GV100 cards)
 - NVIDIA® NVLink™ is not available or functional unless exactly **two** NVIDIA® NVLink™ bridges are properly connected (ie. NVIDIA® NVLink™ is not available in a single-bridge configuration, neither is it available without any bridges)
 - NVIDIA® NVLink™ bridges are sold separately from HP systems and NVIDIA® Quadro® cards, and are available as a kit (sold in pairs) from HP
- **1x NVIDIA® Quadro® card of the same GPU generation for display connectivity**
 - Please note, GPU generation mixing is **not** supported; all GPUs in the system must be of the same generation
 - The only exception is with NVIDIA® Quadro® GV100 (Volta), which is supported with NVIDIA® Quadro® Pxxxx (Pascal) cards. However, NVIDIA® Quadro® GP100 (Pascal) is only supported with other Pascal cards (NVIDIA® Quadro® Pxxxx)
 - Since the NVIDIA® NVLink™-capable cards will be dedicated to compute purposes (more information in Section 4), a 3rd card is needed to connect a display
 - **Note:** in some Linux® configurations, this 3rd card is optional

OPERATING SYSTEM CONFIGURATION WINDOWS

In **Windows**, both NVIDIA® NVLink™-capable NVIDIA® Quadro® cards must be **manually set to TCC mode** to take full advantage of NVIDIA® NVLink™. By default, all graphics cards are configured in WDDM mode. Please note that all display outputs are disabled when a card is in TCC mode and will remain disabled until the card is manually switched to WDDM mode. For more information on the difference between these modes and why TCC mode is required please review the following documentation:

- <https://docs.nvidia.com/cuda/cuda-c-best-practices-guide/index.html#unified-virtual-addressing>
- <https://docs.nvidia.com/cuda/cuda-installation-guide-microsoft-windows/index.html#driver-model>

To configure the NVIDIA® NVLink™-capable NVIDIA® Quadro® cards in TCC mode:

1. **Install** the latest **HP-provided NVIDIA graphics driver**
 - The latest drivers for your platform can be downloaded here: <https://support.hp.com/us-en/drivers>
2. Restart the computer and **press F10** to enter F10 BIOS Setup, then navigate to the Advanced tab, then to the Graphics Configuration link, and press Enter. From the drop-down box, select the slot in which the non- NVIDIA® NVLink™-capable NVIDIA® Quadro® card is located (ie. the card to be used for display purposes). Press F10 again, then select the option to **save changes and restart**
3. **Connect a display** to the card in the slot selected in F10 Setup during the previous step
4. Using a **command line interface**, such as Command Prompt or PowerShell, **locate the NVSMI tool**
 - The NVSMI tool can be used to toggle TCC mode and view other status information. For more information on NVSMI, run `nvidia-smi.exe -h` or the following documentation: <https://developer.nvidia.com/nvidia-system-management-interface>
 - By default, the utility is located in `C:\Program Files\NVIDIA Corporation\NVSMI`
5. Run `nvidia-smi.exe -l` and **take note of the ID numbers for both** of the NVIDIA® NVLink™-capable NVIDIA® Quadro® cards
 - At this point, you may also notice that both NVIDIA® NVLink™-capable NVIDIA® Quadro® cards and the display-dedicated 3rd card should all be identified as being in WDDM mode
6. Run `nvidia-smi.exe -i # -dm TCC` for **each** NVIDIA® NVLink™-capable NVIDIA® Quadro® card, **replacing # with the correct ID numbers** (for example, 0, 1, etc.) obtained in the previous step, **one at a time**
 - Note: Ensure that nothing is connected to any of the display connectors on the NVIDIA® NVLink™-capable cards or TCC mode will not enable **correctly**
 - Warning: Do attempt to set the display-dedicated card (3rd card) to TCC mode as you may lose the ability to output a display on that card
7. **Restart** the computer
8. Run `nvidia-smi.exe -l` and **verify** that **both** NVIDIA® NVLink™-capable NVIDIA® Quadro® cards are identified as being in **TCC mode**
9. At this point, your system is fully configured for NVIDIA® NVLink™ and is ready for NVIDIA® NVLink™-capable software!



TECHNICAL WHITE PAPER

CONTENTS & NAVIGATION

2

Introduction

2

Software Considerations

2-3

Hardware Considerations

3-4

Operating System Configuration

4

Monitoring NVIDIA® NVLink™ Status

LINUX®

In Linux®, generally no special configuration is required to enable the OS to be ready for NVIDIA® NVLink™; it simply depends on the graphics driver and the target application. In any case, the latest HP-provided NVIDIA graphics driver is required for NVIDIA® NVLink™ to function properly. The latest drivers for your platform can be downloaded here: <https://support.hp.com/us-en/drivers>.

NVIDIA® Quadro® cards properly enabled by NVIDIA® NVLink™ hardware and drivers will reflect their supported status by offering an additional SLI Mosaic mode through the *nvidia-settings* GUI application. Most Linux® desktop managers will benefit immediately from the multiple-card layouts (rows and columns) possible through X server configuration. See NVIDIA graphic driver documentation and manual pages for *nvidia-settings* for more details.

MONITORING NVIDIA® NVLINK™ STATUS

Included with the NVIDIA graphics driver for both Windows and Linux®, the NVSMI tool (introduced in Section 4, above) can be used to monitor NVIDIA® NVLink™ status at any time using the following commands:

- *nvidia-smi.exe NVIDIA NVLink -s*
 - This command displays the current state of all possible NVIDIA® NVLink™ links
 - Each bridge provides each card with 2 NVIDIA® NVLink™ links, so a total of 4 links per card will be listed along with its status (either a speed rating or <inactive>)
 - If <inactive> is listed, then either the bridge is missing/defective or the system is not correctly configured
- *nvidia-smi.exe NVIDIA NVLink -c*
 - This command displays the current status of each link capability for each NVIDIA® NVLink™-capable NVIDIA® Quadro® card
- *nvidia-smi.exe NVIDIA NVLink -h*
 - This command provides additional information on other options for controlling/monitoring NVIDIA® NVLink™ using the NVSMI tool

LET US HELP YOU CREATE AMAZING BUSINESS
SOLUTIONS TODAY

CONTACT US

© Copyright 2018 HP Development Company, L.P. The information contained herein is subject to change without notice. The only warranties for HP products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. HP shall not be liable for technical or editorial errors or omissions contained herein.

NVIDIA, Quadro and NVLink are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries.

4AA7-3070ENW, May 2018

