Installation Guide for Linux

Release 12.5

NVIDIA Corporation

Jul 19, 2024
Contents

1 System Requirements ................................................. 3
2 OS Support Policy .................................................. 5
3 Host Compiler Support Policy ....................................... 7
   3.1 Supported C++ Dialects ........................................ 7
4 About This Document ................................................ 9
5 Pre-installation Actions ........................................... 11
   5.1 Verify You Have a CUDA-Capable GPU ....................... 11
   5.2 Verify You Have a Supported Version of Linux ............... 12
   5.3 Verify the System Has gcc Installed .......................... 12
   5.4 Verify the System has the Correct Kernel Headers and Development Packages Installed 12
   5.5 Install GPUDirect Storage .................................... 13
   5.6 Choose an Installation Method ................................. 13
   5.7 Download the NVIDIA CUDA Toolkit ......................... 14
   5.8 Address Custom xorg.conf, If Applicable ................... 14
   5.9 Handle Conflicting Installation Methods .................... 14
6 Package Manager Installation ..................................... 17
   6.1 Overview ...................................................... 17
   6.2 RHEL 8 / Rocky 8 ............................................... 18
      6.2.1 Prepare RHEL 8 / Rocky 8 ................................. 18
      6.2.2 Local Repo Installation for RHEL 8 / Rocky 8 .......... 18
      6.2.3 Network Repo Installation for RHEL 8 / Rocky 8 ....... 19
      6.2.4 Common Instructions for RHEL 8 / Rocky 8 ............. 19
   6.3 RHEL 9 / Rocky 9 ............................................... 20
      6.3.1 Prepare RHEL 9 / Rocky 9 ................................. 20
      6.3.2 Local Repo Installation for RHEL 9 / Rocky 9 .......... 20
      6.3.3 Network Repo Installation for RHEL 9 / Rocky 9 ....... 21
      6.3.4 Common Instructions for RHEL 9 / Rocky 9 ............. 21
   6.4 KylinOS 10 ...................................................... 22
      6.4.1 Prepare KylinOS 10 ........................................ 22
      6.4.2 Local Repo Installation for KylinOS ...................... 22
      6.4.3 Network Repo Installation for KylinOS ................... 22
      6.4.4 Common Instructions for KylinOS 10 ..................... 22
   6.5 Fedora .......................................................... 23
      6.5.1 Prepare Fedora ............................................. 23
      6.5.2 Local Repo Installation for Fedora ....................... 23
      6.5.3 Network Repo Installation for Fedora .................... 23
      6.5.4 Common Installation Instructions for Fedora ............ 24
   6.6 SLES .......................................................... 25
      6.6.1 Prepare SLES .............................................. 25
6.6.2 Local Repo Installation for SLES ......................................................... 25
6.6.3 Network Repo Installation for SLES ...................................................... 26
6.6.4 Common Installation Instructions for SLES ........................................ 26
6.7 OpenSUSE ......................................................................................... 27
   6.7.1 Prepare OpenSUSE ................................................................. 27
   6.7.2 Local Repo Installation for OpenSUSE ........................................... 27
   6.7.3 Network Repo Installation for OpenSUSE ...................................... 27
   6.7.4 Common Installation Instructions for OpenSUSE .............................. 28
6.8 WSL ............................................................................................... 28
   6.8.1 Prepare WSL ......................................................................... 28
   6.8.2 Local Repo Installation for WSL ................................................ 29
   6.8.3 Network Repo Installation for WSL ............................................. 29
   6.8.4 Common Installation Instructions for WSL .................................. 29
6.9 Ubuntu ............................................................................................ 30
   6.9.1 Prepare Ubuntu ................................................................. 30
   6.9.2 Local Repo Installation for Ubuntu ............................................. 30
   6.9.3 Network Repo Installation for Ubuntu ....................................... 30
   6.9.4 Common Installation Instructions for Ubuntu .................................. 32
6.10 Debian ............................................................................................ 32
   6.10.1 Prepare Debian ................................................................. 32
   6.10.2 Local Repo Installation for Debian ............................................. 33
   6.10.3 Network Repo Installation for Debian ....................................... 33
   6.10.4 Common Installation Instructions for Debian ............................... 33
6.11 Amazon Linux 2023 ............................................................... 34
   6.11.1 Prepare Amazon Linux 2023 .................................................. 34
   6.11.2 Local Repo Installation for Amazon Linux .................................. 34
   6.11.3 Network Repo Installation for Amazon Linux ............................... 34
   6.11.4 Common Installation Instructions for Amazon Linux ....................... 35
6.12 Additional Package Manager Capabilities ........................................ 35
   6.12.1 Available Packages .......................................................... 35
   6.12.2 Meta Packages ............................................................... 36
   6.12.3 Optional 32-bit Packages for Linux x86_64 .deb/.rpm ..................... 37
   6.12.4 Package Upgrades ............................................................. 37

7 Driver Installation ................................................................. 39

8 NVIDIA Open GPU Kernel Modules ................................................ 41
   8.1 CUDA Runfile ................................................................. 41
   8.2 Debian .................................................................................. 42
   8.3 Fedora ...................................................................................... 42
   8.4 KylinOS 10 .............................................................................. 42
   8.5 RHEL 9 and Rocky 9 ............................................................. 43
   8.6 RHEL 8 and Rocky 8 ............................................................. 43
   8.7 OpenSUSE and SLES ............................................................. 43
   8.8 Ubuntu .................................................................................. 44

9 Precompiled Streams .......................................................... 45
   9.1 Precompiled Streams Support Matrix ......................................... 46
   9.2 Modularity Profiles ............................................................... 47

10 Kickstart Installation ............................................................. 49
   10.1 RHEL 8 / Rocky Linux 8 ........................................................... 49
   10.2 RHEL 9 / Rocky Linux 9 ........................................................... 49

11 Runfile Installation ............................................................. 51
11.1 Runfile Overview ........................................... 51
11.2 Installation .................................................. 51
11.3 Disabling Nouveau ........................................... 53
  11.3.1 Fedora ...................................................... 53
  11.3.2 RHEL / Rocky and KylinOS ............................... 53
  11.3.3 OpenSUSE .................................................. 53
  11.3.4 SLES ......................................................... 54
  11.3.5 WSL ......................................................... 54
  11.3.6 Ubuntu ..................................................... 54
  11.3.7 Debian ....................................................... 54
11.4 Device Node Verification ................................... 54
11.5 Advanced Options ............................................ 56
11.6 Uninstallation .................................................. 57
12 Conda Installation ............................................. 59
  12.1 Conda Overview ............................................. 59
  12.2 Installing CUDA Using Conda ............................. 59
  12.3 Uninstalling CUDA Using Conda ......................... 59
  12.4 Installing Previous CUDA Releases ..................... 59
  12.5 Upgrading from cudatoolkit Package .................... 60
13 Pip Wheels ....................................................... 61
14 Tarball and Zip Archive Deliverables ...................... 63
  14.1 Parsing Redistrib JSON ..................................... 64
  14.2 Importing Tarballs into CMake ........................... 65
  14.3 Importing Tarballs into Bazel ........................... 65
15 CUDA Cross-Platform Environment ......................... 67
  15.1 CUDA Cross-Platform Installation ....................... 67
  15.2 CUDA Cross-Platform Samples ............................ 68
16 Post-installation Actions ..................................... 69
  16.1 Mandatory Actions ......................................... 69
    16.1.1 Environment Setup ...................................... 69
  16.2 Recommended Actions ....................................... 70
    16.2.1 Install Persistence Daemon ............................ 70
    16.2.2 Install Writable Samples .............................. 70
    16.2.3 Verify the Installation ................................. 70
      16.2.3.1 Verify the Driver Version ........................ 70
      16.2.3.2 Running the Binaries .............................. 71
    16.2.4 Install Nsight Eclipse Plugins ...................... 72
    16.2.5 Local Repo Removal .................................... 72
  16.3 Optional Actions ........................................... 73
    16.3.1 Install Third-party Libraries ....................... 73
    16.3.2 Install the Source Code for cuda-gdb ............... 74
    16.3.3 Select the Active Version of CUDA .................. 74
17 Advanced Setup .................................................. 75
18 Frequently Asked Questions .................................. 77
  18.1 How do I install the Toolkit in a different location? .... 77
  18.2 Why do I see "nvcc: No such file or directory" when I try to build a CUDA application? 77
18.3 Why do I see “error while loading shared libraries: <lib name>: cannot open shared object file: No such file or directory” when I try to run a CUDA application that uses a CUDA library? 78
18.4 Why do I see multiple “404 Not Found” errors when updating my repository meta-data on Ubuntu? 78
18.5 How can I tell X to ignore a GPU for compute-only use? 78
18.6 Why doesn’t the cuda-repo package install the CUDA Toolkit and Drivers? 79
18.7 How do I get CUDA to work on a laptop with an iGPU and a dGPU running Ubuntu14.04? 79
18.8 What do I do if the display does not load, or CUDA does not work, after performing a system update? 79
18.9 How do I install a CUDA driver with a version less than 367 using a network repo? 80
18.10 How do I install an older CUDA version using a network repo? 80
18.11 Why does the installation on SUSE install the Mesa-dri-nouveau dependency? 80
18.12 How do I handle “Errors were encountered while processing: glx-diversions”? 81

19 Additional Considerations 83

20 Switching between Driver Module Flavors 85

21 Removing CUDA Toolkit and Driver 87

22 Notices 89
22.1 Notice 89
22.2 OpenCL 90
22.3 Trademarks 90

23 Copyright 91
NVIDIA CUDA Installation Guide for Linux

The installation instructions for the CUDA Toolkit on Linux.

CUDA® is a parallel computing platform and programming model invented by NVIDIA®. It enables dramatic increases in computing performance by harnessing the power of the graphics processing unit (GPU).

CUDA was developed with several design goals in mind:

▶ Provide a small set of extensions to standard programming languages, like C, that enable a straightforward implementation of parallel algorithms. With CUDA C/C++, programmers can focus on the task of parallelization of the algorithms rather than spending time on their implementation.

▶ Support heterogeneous computation where applications use both the CPU and GPU. Serial portions of applications are run on the CPU, and parallel portions are offloaded to the GPU. As such, CUDA can be incrementally applied to existing applications. The CPU and GPU are treated as separate devices that have their own memory spaces. This configuration also allows simultaneous computation on the CPU and GPU without contention for memory resources.

CUDA-capable GPUs have hundreds of cores that can collectively run thousands of computing threads. These cores have shared resources including a register file and a shared memory. The on-chip shared memory allows parallel tasks running on these cores to share data without sending it over the system memory bus.

This guide will show you how to install and check the correct operation of the CUDA development tools.
Chapter 1. System Requirements

To use NVIDIA CUDA on your system, you will need the following installed:

- CUDA-capable GPU
- A supported version of Linux with a gcc compiler and toolchain

The CUDA development environment relies on tight integration with the host development environment, including the host compiler and C runtime libraries, and is therefore only supported on distribution versions that have been qualified for this CUDA Toolkit release.

The following table lists the supported Linux distributions. Please review the footnotes associated with the table.

Table 1: Native Linux Distribution Support in CUDA 12.5 Update

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Kernel</th>
<th>Default GCC</th>
<th>GLIBC</th>
</tr>
</thead>
<tbody>
<tr>
<td>x86_64</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RHEL 9.y (y &lt;= 4)</td>
<td>5.14.0-427</td>
<td>11.4.1</td>
<td>2.34</td>
</tr>
<tr>
<td>RHEL 8.y (y &lt;= 10)</td>
<td>4.18.0-553</td>
<td>8.5.0</td>
<td>2.28</td>
</tr>
<tr>
<td>OpenSUSE Leap 15.y (y &lt;= 5)</td>
<td>5.14.21-150500</td>
<td>7.5.0</td>
<td>2.31</td>
</tr>
<tr>
<td>Rocky Linux 8.y (y&lt;=10)</td>
<td>4.18.0-553</td>
<td>8.5.0</td>
<td>2.28</td>
</tr>
<tr>
<td>Rocky Linux 9.y (y&lt;=4)</td>
<td>5.14.0-427</td>
<td>11.4.1</td>
<td>2.34</td>
</tr>
<tr>
<td>SUSE SLES 15.y (y &lt;= 5)</td>
<td>5.14.21-150500</td>
<td>7.5.0</td>
<td>2.31</td>
</tr>
<tr>
<td>Ubuntu 24.04 LTS</td>
<td>6.8.0-31</td>
<td>13.2.0</td>
<td>2.39</td>
</tr>
<tr>
<td>Ubuntu 22.04.z (z &lt;= 4) LTS</td>
<td>6.5.0-27</td>
<td>12.3.0</td>
<td>2.35</td>
</tr>
<tr>
<td>Ubuntu 20.04.z (z &lt;= 6) LTS</td>
<td>5.15.0-67</td>
<td>9.4.0</td>
<td>2.31</td>
</tr>
<tr>
<td>Debian 12.x (x&lt;=5)</td>
<td>6.1.76-1</td>
<td>12.2.0</td>
<td>2.36</td>
</tr>
<tr>
<td>Debian 11.y (y&lt;=9)</td>
<td>5.10.209-2</td>
<td>10.2.1</td>
<td>2.31</td>
</tr>
<tr>
<td>Debian 10.z (z&lt;=13)</td>
<td>4.19.0-21</td>
<td>8.3.0</td>
<td>2.28</td>
</tr>
<tr>
<td>Fedora 39</td>
<td>6.5.6-300</td>
<td>13.2.1</td>
<td>2.38</td>
</tr>
<tr>
<td>KylinOS V10 SP2</td>
<td>4.19.90-25.14.v2101.ky10</td>
<td>7.3.0</td>
<td>2.28</td>
</tr>
</tbody>
</table>

continues on next page
### Table 1 – continued from previous page

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Kernel¹</th>
<th>Default GCC</th>
<th>GLIBC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amazon Linux 2023</td>
<td>6.1.82-99.168</td>
<td>11.4.1</td>
<td>2.34</td>
</tr>
<tr>
<td><strong>Arm64 sbsa</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RHEL 9.y (y &lt;= 4)</td>
<td>5.14.0-427</td>
<td>11.4.1</td>
<td>2.34</td>
</tr>
<tr>
<td>RHEL 8.y (y &lt;= 10)</td>
<td>4.18.0-553</td>
<td>8.5.0</td>
<td>2.28</td>
</tr>
<tr>
<td>SUSE SLES 15.y (y &lt;= 5)</td>
<td>5.14.21-150500</td>
<td>7.5.0</td>
<td>2.32</td>
</tr>
<tr>
<td>Ubuntu 24.04 LTS</td>
<td>6.8.0-31</td>
<td>13.2.0</td>
<td>2.39</td>
</tr>
<tr>
<td>Ubuntu 22.04 LTS (z &lt;= 5) LTS</td>
<td>5.15.0-102</td>
<td>11.4.0</td>
<td>2.35</td>
</tr>
<tr>
<td>Ubuntu 20.04.z (z &lt;= 5) LTS</td>
<td>5.4.0-174</td>
<td>9.4.0</td>
<td>2.31</td>
</tr>
<tr>
<td><strong>Arm64 sbsa Jetson (dGPU)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20.04.06 LTS Rel35 JP 5.x</td>
<td>5.10.192-tegra</td>
<td>9.4.0</td>
<td>2.31</td>
</tr>
<tr>
<td>22.04.4 LTS Rel36 - JP6.x</td>
<td>5.15.136-tegra</td>
<td>11.4.0</td>
<td>2.35</td>
</tr>
<tr>
<td><strong>Aarch64 Jetson (IGPU)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L4T Ubuntu 22.04 Rel36 - JP6.x</td>
<td>6.1.80-tegra</td>
<td>11.4.0</td>
<td>2.35</td>
</tr>
</tbody>
</table>

(1) The following notes apply to the kernel versions supported by CUDA:

- For specific kernel versions supported on Red Hat Enterprise Linux (RHEL), visit [https://access.redhat.com/articles/3078](https://access.redhat.com/articles/3078).

- A list of kernel versions including the release dates for SUSE Linux Enterprise Server (SLES) is available at [https://www.suse.com/support/kb/doc/?id=000019587](https://www.suse.com/support/kb/doc/?id=000019587).

Chapter 2. OS Support Policy

- CUDA support for Ubuntu 20.04.x, Ubuntu 22.04.x, RHEL 8.x, RHEL 9.x, Rocky Linux 8.x, Rocky Linux 9.x, SUSE SLES 15.x and OpenSUSE Leap 15.x will be until the standard EOS as defined for each OS. Please refer to the support lifecycle for these OSes to know their support timelines.

- CUDA supports the latest Fedora release version. For Fedora release timelines, visit https://docs.fedoraproject.org/en-US/releases/.

- CUDA supports a single KylinOS release version. For details, visit https://www.kylinos.cn/.

Refer to the support lifecycle for these supported OSes to know their support timelines and plan to move to newer releases accordingly.
Chapter 3. Host Compiler Support Policy

In order to compile the CPU “Host” code in the CUDA source, the CUDA compiler NVCC requires a compatible host compiler to be installed on the system. The version of the host compiler supported on Linux platforms is tabulated as below. NVCC performs a version check on the host compiler’s major version and so newer minor versions of the compilers listed below will be supported, but major versions falling outside the range will not be supported.

Table 2: Supported Compilers

<table>
<thead>
<tr>
<th>Distribution</th>
<th>GCC</th>
<th>Clang</th>
<th>NVHPC</th>
<th>XLC</th>
<th>ArmC/C++</th>
<th>ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td>x86_64</td>
<td>6.x - 13.2</td>
<td>7.x - 17.0</td>
<td>24.3</td>
<td>No</td>
<td>No</td>
<td>2021.7</td>
</tr>
<tr>
<td>Arm64 sbsa</td>
<td>6.x - 13.2</td>
<td>7.x - 17.0</td>
<td>24.3</td>
<td>No</td>
<td>23.04.1</td>
<td>No</td>
</tr>
</tbody>
</table>

For GCC and Clang, the preceding table indicates the minimum version and the latest version supported. If you are on a Linux distribution that may use an older version of GCC toolchain as default than what is listed above, it is recommended to upgrade to a newer toolchain CUDA 11.0 or later toolkit. Newer GCC toolchains are available with the Red Hat Developer Toolset for example. For platforms that ship a compiler version older than GCC 6 by default, linking to static or dynamic libraries that are shipped with the CUDA Toolkit is not supported. We only support libstdc++ (GCC’s implementation) for all the supported host compilers for the platforms listed above.

3.1. Supported C++ Dialects

NVCC and NVRTC (CUDA Runtime Compiler) support the following C++ dialect: C++11, C++14, C++17, C++20 on supported host compilers. The default C++ dialect of NVCC is determined by the default dialect of the host compiler used for compilation. Refer to host compiler documentation and the CUDA Programming Guide for more details on language support.

C++20 is supported with the following flavors of host compiler in both host and device code.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>GCC</th>
<th>Clang</th>
<th>NVHPC</th>
<th>Arm C/C++</th>
</tr>
</thead>
<tbody>
<tr>
<td>x86_64</td>
<td>&gt;=10.x</td>
<td>&gt;=11.x</td>
<td>&gt;=22.x</td>
<td>&gt;=22.x</td>
</tr>
</tbody>
</table>
Chapter 4. About This Document

This document is intended for readers familiar with the Linux environment and the compilation of C programs from the command line. You do not need previous experience with CUDA or experience with parallel computation. Note: This guide covers installation only on systems with X Windows installed.

**Note:** Many commands in this document might require *superuser* privileges. On most distributions of Linux, this will require you to log in as root. For systems that have enabled the sudo package, use the sudo prefix for all necessary commands.
Chapter 5. Pre-installation Actions

Some actions must be taken before the CUDA Toolkit and Driver can be installed on Linux:

▶ Verify the system has a CUDA-capable GPU.
▶ Verify the system is running a supported version of Linux.
▶ Verify the system has gcc installed.
▶ Verify the system has the correct kernel headers and development packages installed.
▶ Download the NVIDIA CUDA Toolkit.
▶ Handle conflicting installation methods.

Note: You can override the install-time prerequisite checks by running the installer with the -override flag. Remember that the prerequisites will still be required to use the NVIDIA CUDA Toolkit.

5.1. Verify You Have a CUDA-Capable GPU

To verify that your GPU is CUDA-capable, go to your distribution's equivalent of System Properties, or, from the command line, enter:

```
lspci | grep -i nvidia
```

If you do not see any settings, update the PCI hardware database that Linux maintains by entering `update-pciids` (generally found in `/sbin`) at the command line and rerun the previous `lspci` command.

If your graphics card is from NVIDIA and it is listed in https://developer.nvidia.com/cuda-gpus, your GPU is CUDA-capable.

The Release Notes for the CUDA Toolkit also contain a list of supported products.
5.2. Verify You Have a Supported Version of Linux

The CUDA Development Tools are only supported on some specific distributions of Linux. These are listed in the CUDA Toolkit release notes.

To determine which distribution and release number you’re running, type the following at the command line:

```bash
uname -m && cat /etc/*release
```

You should see output similar to the following, modified for your particular system:

```
x86_64
Red Hat Enterprise Linux Workstation release 6.0 (Santiago)
```

The x86_64 line indicates you are running on a 64-bit system. The remainder gives information about your distribution.

5.3. Verify the System Has gcc Installed

The gcc compiler is required for development using the CUDA Toolkit. It is not required for running CUDA applications. It is generally installed as part of the Linux installation, and in most cases the version of gcc installed with a supported version of Linux will work correctly.

To verify the version of gcc installed on your system, type the following on the command line:

```bash
gcc --version
```

If an error message displays, you need to install the development tools from your Linux distribution or obtain a version of gcc and its accompanying toolchain from the Web.

5.4. Verify the System has the Correct Kernel Headers and Development Packages Installed

The CUDA Driver requires that the kernel headers and development packages for the running version of the kernel be installed at the time of the driver installation, as well whenever the driver is rebuilt. For example, if your system is running kernel version 3.17.4-301, the 3.17.4-301 kernel headers and development packages must also be installed.

While the Runfile installation performs no package validation, the RPM and Deb installations of the driver will make an attempt to install the kernel header and development packages if no version of these packages is currently installed. However, it will install the latest version of these packages, which may or may not match the version of the kernel your system is using. Therefore, it is best to
manually ensure the correct version of the kernel headers and development packages are installed prior to installing the CUDA Drivers, as well as whenever you change the kernel version.

The version of the kernel your system is running can be found by running the following command:

```
uname -r
```

This is the version of the kernel headers and development packages that must be installed prior to installing the CUDA Drivers. This command will be used multiple times below to specify the version of the packages to install. Note that below are the common-case scenarios for kernel usage. More advanced cases, such as custom kernel branches, should ensure that their kernel headers and sources match the kernel build they are running.

**Note:** If you perform a system update which changes the version of the Linux kernel being used, make sure to rerun the commands below to ensure you have the correct kernel headers and kernel development packages installed. Otherwise, the CUDA Driver will fail to work with the new kernel.

### 5.5. Install GPUDirect Storage

If you intend to use GPUDirectStorage (GDS), you must install the CUDA package and MLNX_OFED package.

GDS packages can be installed using the CUDA packaging guide. Follow the instructions in MLNX_OFED Requirements and Installation.

GDS is supported in two different modes: GDS (default/full perf mode) and Compatibility mode. Installation instructions for them differ slightly. Compatibility mode is the only mode that is supported on certain distributions due to software dependency limitations.

Full GDS support is restricted to the following Linux distros:

- Ubuntu 20.04, Ubuntu 22.04
- RHEL 8.3, RHEL 8.4, RHEL 9.0

Starting with CUDA toolkit 12.2.2, GDS kernel driver package nvidia-gds version 12.2.2-1 (provided by nvidia-fs-dkms 2.17.5-1) and above is only supported with the NVIDIA open kernel driver. Follow the instructions in Removing CUDA Toolkit and Driver to remove existing NVIDIA driver packages and then follow instructions in NVIDIA Open GPU Kernel Modules to install NVIDIA open kernel driver packages.

### 5.6. Choose an Installation Method

The CUDA Toolkit can be installed using either of two different installation mechanisms: distribution-specific packages (RPM and Deb packages), or a distribution-independent package (runfile packages).

The distribution-independent package has the advantage of working across a wider set of Linux distributions, but does not update the distribution’s native package management system. The distribution-specific packages interface with the distribution’s native package management system. It is recommended to use the distribution-specific packages, where possible.
5.7. Download the NVIDIA CUDA Toolkit

The NVIDIA CUDA Toolkit is available at https://developer.nvidia.com/cuda-downloads. Choose the platform you are using and download the NVIDIA CUDA Toolkit.

The CUDA Toolkit contains the CUDA driver and tools needed to create, build and run a CUDA application as well as libraries, header files, and other resources.

Download Verification

The download can be verified by comparing the MD5 checksum posted at https://developer.download.nvidia.com/compute/cuda/12.5.1/docs/sidebar/md5sum.txt with that of the downloaded file. If either of the checksums differ, the downloaded file is corrupt and needs to be downloaded again.

To calculate the MD5 checksum of the downloaded file, run the following:

```
md5sum <file>
```

5.8. Address Custom xorg.conf, If Applicable

The driver relies on an automatically generated `xorg.conf` file at `/etc/X11/xorg.conf`. If a custom-built `xorg.conf` file is present, this functionality will be disabled and the driver may not work. You can try removing the existing `xorg.conf` file, or adding the contents of `/etc/X11/xorg.conf.d/00-nvidia.conf` to the `xorg.conf` file. The `xorg.conf` file will most likely need manual tweaking for systems with a non-trivial GPU configuration.

5.9. Handle Conflicting Installation Methods

Before installing CUDA, any previous installations that could conflict should be uninstalled. This will not affect systems which have not had CUDA installed previously, or systems where the installation method has been preserved (RPM/Deb vs. Runfile). See the following charts for specifics.

```
<table>
<thead>
<tr>
<th>Installing Toolkit Version</th>
<th>RPM/Deb</th>
<th>No Action</th>
<th>Uninstall Run</th>
<th>RPM/Deb</th>
<th>No Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Installed Toolkit Version == X.Y</td>
<td>run</td>
<td>Uninstall Run</td>
<td>Uninstall Run</td>
<td>No Action</td>
<td>No Action</td>
</tr>
<tr>
<td>RPM/Deb</td>
<td>run</td>
<td>Uninstall Run</td>
<td>Uninstall Run</td>
<td>No Action</td>
<td>No Action</td>
</tr>
</tbody>
</table>
```

Table 3: CUDA Toolkit Installation Compatibility Matrix
### Table 4: NVIDIA Driver Installation Compatibility Matrix

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>RPM/Deb</td>
<td>run</td>
<td>RPM/Deb</td>
</tr>
<tr>
<td>run</td>
<td>No Action</td>
<td>Uninstall Run</td>
</tr>
<tr>
<td>Uninstall RPM/Deb</td>
<td>No Action</td>
<td>RPM/Deb</td>
</tr>
<tr>
<td>Run</td>
<td>Uninstall</td>
<td>No Action</td>
</tr>
</tbody>
</table>

Use the following command to uninstall a Toolkit runfile installation:

```
sudo /usr/local/cuda-X.Y/bin/cuda-uninstaller
```

Use the following command to uninstall a Driver runfile installation:

```
sudo /usr/bin/nvidia-uninstall
```

Use the following commands to uninstall an RPM/Deb installation:

```
sudo dnf remove <package_name>       # RHEL 8 / Rocky Linux 8
sudo dnf remove <package_name>       # Fedora
sudo zypper remove <package_name>    # OpenSUSE / SLES
sudo apt-get --purge remove <package_name> # Ubuntu
```
Chapter 6. Package Manager Installation

Basic instructions can be found in the Quick Start Guide. Read on for more detailed instructions.

6.1. Overview

Installation using RPM or Debian packages interfaces with your system’s package management system. When using RPM or Debian local repo installers, the downloaded package contains a repository snapshot stored on the local filesystem in /var/. Such a package only informs the package manager where to find the actual installation packages, but will not install them.

If the online network repository is enabled, RPM or Debian packages will be automatically downloaded at installation time using the package manager: apt-get, dnf, yum, or zypper.

Distribution-specific instructions detail how to install CUDA:

- RHEL 8 / Rocky Linux 8
- RHEL 9 / Rocky Linux 9
- KylinOS 10
- Fedora
- SLES
- OpenSUSE
- WSL
- Ubuntu
- Debian
- Amazon Linux 2023

Finally, some helpful package manager capabilities are detailed.

These instructions are for native development only. For cross-platform development, see the CUDA Cross-Platform Environment section.

Note: Optional components such as nvidia-fs, libnvidia_nscq, and fabricmanager are not installed by default and will have to be installed separately as needed.
6.2. RHEL 8 / Rocky 8

6.2.1. Prepare RHEL 8 / Rocky 8

1. Perform the pre-installation actions.

2. The kernel headers and development packages for the currently running kernel can be installed with:

   ```
   sudo dnf install kernel-devel-$(uname -r) kernel-headers-$(uname -r)
   ```

   If matching kernel-headers and kernel-devel packages are not available for the currently running kernel version, you may need to use the previously shipped version of these packages. See https://bugzilla.redhat.com/show_bug.cgi?id=1986132 for more information.

3. Satisfy third-party package dependency:

   - **Satisfy DKMS dependency:** The NVIDIA driver RPM packages depend on other external packages, such as DKMS and libvdpau. Those packages are only available on third-party repositories, such as EPEL. Any such third-party repositories must be added to the package manager repository database before installing the NVIDIA driver RPM packages, or missing dependencies will prevent the installation from proceeding.

     To enable EPEL:

     ```
     ```

     - **Enable optional repos:**

       On **RHEL 8 Linux** only, execute the following steps to enable optional repositories.

       - **On x86_64 systems:**

         ```
         subscription-manager repos --enable=rhel-8-for-x86_64-appstream-rpms
         subscription-manager repos --enable=rhel-8-for-x86_64-baseos-rpms
         subscription-manager repos --enable=codeready-builder-for-rhel-8-x86_64-rpms
         ```

4. Remove Outdated Signing Key:

   ```
   sudo rpm --erase gpg-pubkey-7fa2af80*
   ```

5. Choose an installation method: local repo or network repo.

6.2.2. Local Repo Installation for RHEL 8 / Rocky 8

1. Install local repository on file system:

   ```
   sudo rpm --install cuda-repo-<distro>-X-Y-local-<version>*.<arch>.rpm
   ```
6.2.3. Network Repo Installation for RHEL 8 / Rocky 8

1. **Enable the network repo:**

   ```
   ```

   where $distro/$arch should be replaced by one of the following:
   - rhel8/cross-linux-sbsa
   - rhel8/sbsa
   - rhel8/x86_64

2. **Install the new CUDA public GPG key:**

   The new GPG public key for the CUDA repository (RPM-based distros) is d42d0685.

   On a fresh installation of RHEL, the dnf package manager will prompt the user to accept new keys when installing packages the first time. Indicate you accept the change when prompted.

   For upgrades, you must also also fetch an updated .repo entry:

   ```
   ```

3. **Clean Yum repository cache:**

   ```
   sudo dnf clean expire-cache
   ```

6.2.4. Common Instructions for RHEL 8 / Rocky 8

These instructions apply to both local and network installation.

1. **Install CUDA SDK:**

   ```
   sudo dnf module install nvidia-driver:latest-dkms
   sudo dnf install cuda-toolkit
   ```

2. **Install GPUDirect Filesystem:**

   ```
   sudo dnf install nvidia-gds
   ```

3. **Add libcuda.so symbolic link, if necessary**

   The libcuda.so library is installed in the /usr/lib{,64}/nvidia directory. For pre-existing projects which use libcuda.so, it may be useful to add a symbolic link from libcuda.so in the /usr/lib{,64} directory.

4. **Reboot the system:**

   ```
   sudo reboot
   ```

5. **Perform the post-installation actions.**
6.3. RHEL 9 / Rocky 9

6.3.1. Prepare RHEL 9 / Rocky 9

1. Perform the pre-installation actions.

2. The kernel headers and development packages for the currently running kernel can be installed with:

```
sudo dnf install kernel-devel-$(uname -r) kernel-headers-$(uname -r)
```

3. **Satisfy third-party package dependency:**
   - **Satisfy DKMS dependency:** The NVIDIA driver RPM packages depend on other external packages, such as DKMS and libvdpau. Those packages are only available on third-party repositories, such as EPEL. Any such third-party repositories must be added to the package manager repository database before installing the NVIDIA driver RPM packages, or missing dependencies will prevent the installation from proceeding.
     
     To enable EPEL:
     ```
```

   - **Enable optional repos:**
     On RHEL 9 Linux only, execute the following steps to enable optional repositories.
     
     - **On x86_64 systems:**
     ```
     subscription-manager repos --enable=rhel-9-for-x86_64-appstream-rpms
     subscription-manager repos --enable=rhel-9-for-x86_64-baseos-rpms
     subscription-manager repos --enable=codeready-builder-for-rhel-9-x86_64-rpms
     ```

4. **Remove Outdated Signing Key:**

```
sudo rpm --erase gpg-pubkey-7fa2af80*
```

5. Choose an installation method: local repo or network repo.

6.3.2. Local Repo Installation for RHEL 9 / Rocky 9

1. **Install local repository on file system:**

```
sudo rpm --install cuda-repo-<distro>-<architecture>-<version>-.<architecture>.rpm
```
6.3.3. Network Repo Installation for RHEL 9 / Rocky 9

1. **Enable the network repo:**
   
   ```bash
   ```
   
   where $distro/$arch should be replaced by one of the following:
   
   - rhel9/cross-linux-sbsa
   - rhel9/sbsa
   - rhel9/x86_64

2. **Install the new CUDA public GPG key:**
   
   The new GPG public key for the CUDA repository (RPM-based distros) is **d42d0685**.
   
   On a fresh installation of RHEL, the dnf package manager will prompt the user to accept new keys when installing packages the first time. Indicate you accept the change when prompted.
   
   For upgrades, you must also fetch an updated .repo entry:
   
   ```bash
   ```

3. **Clean Yum repository cache:**
   
   ```bash
   sudo dnf clean expire-cache
   ```

6.3.4. Common Instructions for RHEL 9 / Rocky 9

These instructions apply to both local and network installation.

1. **Install CUDA SDK:**
   
   ```bash
   sudo dnf module install nvidia-driver:latest-dkms
   sudo dnf install cuda-toolkit
   ```

2. **Install GPUDirect Filesystem:**
   
   ```bash
   sudo dnf install nvidia-gds
   ```

3. **Add libcuda.so symbolic link, if necessary**
   
   The libcuda.so library is installed in the /usr/lib{,64}/nvidia directory. For pre-existing projects which use libcuda.so, it may be useful to add a symbolic link from libcuda.so in the /usr/lib{,64} directory.

4. **Reboot the system:**
   
   ```bash
   sudo reboot
   ```

5. **Perform the post-installation actions.**
6.4. KylinOS 10

6.4.1. Prepare KylinOS 10

1. Perform the pre-installation actions.
2. The kernel headers and development packages for the currently running kernel can be installed with:

   ```bash
   sudo dnf install kernel-devel-$(uname -r) kernel-headers-$(uname -r)
   ```
3. Choose an installation method: local repo or network repo.

6.4.2. Local Repo Installation for KylinOS

1. Install local repository on file system:

   ```bash
   sudo rpm --install cuda-repo-kylin10-X-Y-local-<version>*.<arch>.rpm
   ```

6.4.3. Network Repo Installation for KylinOS

1. Enable the network repo:

   ```bash
   ```

2. Install the new CUDA public GPG key:

   The new GPG public key for the CUDA repository (RPM-based distros) is `d42d0685`.

   On a fresh installation of RHEL, the dnf package manager will prompt the user to accept new keys when installing packages the first time. Indicate you accept the change when prompted.
3. Clean Yum repository cache:

   ```bash
   sudo dnf clean expire-cache
   ```

6.4.4. Common Instructions for KylinOS 10

These instructions apply to both local and network installation.

1. Install CUDA SDK:

   ```bash
   sudo dnf module install nvidia-driver:latest-dkms
   sudo dnf install cuda-toolkit
   ```

2. Install GPUDirect Filesystem:
Installation Guide for Linux, Release 12.5

3. Add libcuda.so symbolic link, if necessary

The libcuda.so library is installed in the /usr/lib{,64}/nvidia directory. For pre-existing projects which use libcuda.so, it may be useful to add a symbolic link from libcuda.so in the /usr/lib{,64} directory.

4. Reboot the system:

```
sudo reboot
```

5. Perform the post-installation actions.

6.5. Fedora

6.5.1. Prepare Fedora

1. Perform the pre-installation actions.

2. The kernel headers and development packages for the currently running kernel can be installed with:

```
sudo dnf install kernel-devel-$(uname -r) kernel-headers-$(uname -r)
```

3. Remove Outdated Signing Key:

```
sudo rpm --erase gpg-pubkey-7fa2af80*
```

4. Choose an installation method: local repo or network repo.

6.5.2. Local Repo Installation for Fedora

1. Install local repository on file system:

```
sudo rpm --install cuda-repo-<distro>-X-Y-local-<version>*.x86_64.rpm
```

   where distro is fedora37 or fedora39, for example.

6.5.3. Network Repo Installation for Fedora

1. Enable the network repo:

```
```

   where $distro should be replaced by one of the following:

   ▶ fedora37
2. **Install the new CUDA public GPG key:**

   The new GPG public key for the CUDA repository (RPM-based distros) is `d42d0685`. 

   On a fresh installation of Fedora, the dnf package manager will prompt the user to accept new keys when installing packages the first time. Indicate you accept the change when prompted.

   For upgrades, you must also fetch an updated `.repo` entry:

   ```bash
   ```

3. **Clean DNF repository cache:**

   ```bash
   sudo dnf clean expire-cache
   ```

### 6.5.4. Common Installation Instructions for Fedora

These instructions apply to both local and network installation for Fedora.

1. **Install CUDA SDK:**

   ```bash
   sudo dnf module install nvidia-driver:latest-dkms
   sudo dnf install cuda-toolkit
   ```

   **Note:** The CUDA driver installation may fail if the RPMFusion non-free repository is enabled. In this case, CUDA installations should temporarily disable the RPMFusion non-free repository.

   ```bash
   sudo dnf --disablerepo="rpmfusion-nonfree*" install cuda
   ```

   It may be necessary to rebuild the grub configuration files, particularly if you use a non-default partition scheme. If so, then run this below command, and reboot the system:

   ```bash
   sudo grub2-mkconfig -o /boot/grub2/grub.cfg
   ```

2. **Reboot the system:**

   ```bash
   sudo reboot
   ```

3. **Add libcuda.so symbolic link, if necessary:**

   The `libcuda.so` library is installed in the `/usr/lib{,64}/nvidia` directory. For pre-existing projects which use `libcuda.so`, it may be useful to add a symbolic link from `libcuda.so` in the `/usr/lib{,64}` directory.

4. **Perform the post-installation actions.**
6.6. SLES

6.6.1. Prepare SLES

1. Perform the pre-installation actions.
2. The kernel development packages for the currently running kernel can be installed with:
   ```bash
   sudo zypper install -y kernel-<variant>-devel=<version>
   ```
   To run the above command, you will need the variant and version of the currently running kernel. Use the output of the `uname` command to determine the currently running kernel's variant and version:
   ```bash
   $ uname -r
   3.16.6-2-default
   ```
   In the above example, the variant is `default` and version is `3.16.6-2`.
   The kernel development packages for the default kernel variant can be installed with:
   ```bash
   sudo zypper install -y kernel-default-devel=$(uname -r | sed 's∕\-default∕∕')
   ```
3. The kernel headers and development packages for the currently running kernel can be installed with:
   ```bash
   sudo zypper install -y kernel-<variant>-devel=<version>
   ```
4. On SLES12 SP4, install the Mesa-libgl-devel Linux packages before proceeding. See Mesa-libGL-devel.
5. Add the user to the video group:
   ```bash
   sudo usermod -a -G video <username>
   ```
6. Remove Outdated Signing Key:
   ```bash
   sudo rpm --erase gpg-pubkey-7fa2af80* 
   ```
7. Choose an installation method: local repo or network repo.

6.6.2. Local Repo Installation for SLES

1. Install local repository on file system:
   ```bash
   sudo rpm --install cuda-repo-sles15-X-Y-local-<version>*.x86_64.rpm
   ```
6.6.3. Network Repo Installation for SLES

1. Enable the network repo:

   ```
```

   where $distro/$arch should be replaced by one of the following:
   - sles15/cross-linux-sbsa
   - sles15/sbsa
   - sles15/x86_64

2. **Install the new CUDA public GPG key:**

   The new GPG public key for the CUDA repository (RPM-based distros) is d42d0685.

   On a fresh installation of SLES, the zypper package manager will prompt the user to accept new keys when installing packages the first time. Indicate you accept the change when prompted.

   For upgrades, you must also also fetch an updated .repo entry:

   ```
sudo zypper removerepo cuda-$distro-$arch
```

3. **Refresh Zypper repository cache:**

   ```
sudo SUSEConnect --product PackageHub/15/<architecture>
sudo zypper refresh
```

6.6.4. Common Installation Instructions for SLES

These instructions apply to both local and network installation for SLES.

1. **Install CUDA SDK:**

   ```
sudo zypper install cuda-toolkit
```

2. **Install CUDA Samples GL dependencies:**

   Refer to CUDA Cross-Platform Samples.

3. **Reboot the system:**

   ```
sudo reboot
```

4. **Perform the post-installation actions.**
6.7. OpenSUSE

6.7.1. Prepare OpenSUSE

1. Perform the pre-installation actions.
2. The kernel development packages for the currently running kernel can be installed with:
   
   ```
   sudo zypper install -y kernel-<variant>-devel=<version>
   ```

   To run the above command, you will need the variant and version of the currently running kernel. Use the output of the `uname` command to determine the currently running kernel's variant and version:

   ```
   $ uname -r
   3.16.6-2-default
   ```

   In the above example, the variant is `default` and version is `3.16.6-2`.

   The kernel development packages for the default kernel variant can be installed with:

   ```
   sudo zypper install -y kernel-default-devel=$(uname -r | sed 's∕-default∕∕')
   ```

3. Add the user to the video group:

   ```
   sudo usermod -a -G video <username>
   ```

4. Remove Outdated Signing Key:

   ```
   sudo rpm --erase gpg-pubkey-7fa2af80*
   ```

5. Choose an installation method: local repo or network repo.

6.7.2. Local Repo Installation for OpenSUSE

1. Install local repository on file system:

   ```
   sudo rpm --install cuda-repo-opensuse15-<version>.x86_64.rpm
   ```

6.7.3. Network Repo Installation for OpenSUSE

1. Enable the network repo:

   ```
   ```

2. Install the new CUDA public GPG key:

   The new GPG public key for the CUDA repository (RPM-based distros) is `d42d0685`. On fresh installation of openSUSE, the zypper package manager will prompt the user to accept new keys when installing packages the first time. Indicate you accept the change when prompted.
For upgrades, you must also fetch an updated .repo entry:

```
sudo zypper removerepo cuda-opensuse15-x86_64
  opensuse15/x86_64/cuda-opensuse15.repo
```

3. **Refresh Zypper repository cache:**

```
sudo zypper refresh
```

### 6.7.4. Common Installation Instructions for OpenSUSE

These instructions apply to both local and network installation for OpenSUSE.

1. **Install CUDA SDK:**

```
sudo zypper install cuda-toolkit
```

2. **Reboot the system:**

```
sudo reboot
```

3. Perform the post-installation actions.

### 6.8. WSL

These instructions must be used if you are installing in a WSL environment. Do not use the Ubuntu instructions in this case; it is important to not install the cuda-drivers packages within the WSL environment.

#### 6.8.1. Prepare WSL

1. Perform the pre-installation actions.

2. **Remove Outdated Signing Key:**

```
sudo apt-key del 7fa2af80
```

3. Choose an installation method: local repo or network repo.
6.8.2. Local Repo Installation for WSL

1. **Install local repository on file system:**
   
   ```
   sudo dpkg -i cuda-repo-wsl-ubuntu-X-Y-local_<version>*_x86_64.deb
   ```

2. **Enroll ephemeral public GPG key:**
   
   ```
   sudo cp /var/cuda-repo-wsl-ubuntu-X-Y-local/cuda-*keyring.gpg /usr/share/keyrings/
   ```

6.8.3. Network Repo Installation for WSL

The new GPG public key for the CUDA repository (Debian-based distros) is 3bf863cc. This must be enrolled on the system, either using the cuda-keyring package or manually; the apt-key command is deprecated and not recommended.

1. **Install the new cuda-keyring package:**
   
   ```
   wget https://developer.download.nvidia.com/compute/cuda/repos/wsl-ubuntu/x86_64/cuda-keyring_1.1-1_all.deb
   sudo dpkg -i cuda-keyring_1.1-1_all.deb
   ```

   Or if you are unable to install the cuda-keyring package, you can optionally:
   
   a. **Enroll the new signing key manually:**
      
      ```
      sudo mv cuda-archive-keyring.gpg /usr/share/keyrings/cuda-archive-keyring.gpg
      ```
   
   b. **Enable the network repository:**
      
      ```
      echo "deb [signed-by=/usr/share/keyrings/cuda-archive-keyring.gpg] https://developer.download.nvidia.com/compute/cuda/repos/wsl-ubuntu/x86_64/ /" | sudo tee /etc/apt/sources.list.d/cuda-wsl-ubuntu-x86_64.list
      ```
   
   c. **Add pin file to prioritize CUDA repository:**
      
      ```
      sudo mv cuda-wsl-ubuntu.pin /etc/apt/preferences.d/cuda-repository-pin-600
      ```

6.8.4. Common Installation Instructions for WSL

These instructions apply to both local and network installation for WSL.

1. **Update the Apt repository cache:**
   
   ```
   sudo apt-get update
   ```

2. **Install CUDA SDK:**
   
   ```
   sudo apt-get install
   ```
3. Perform the post-installation actions.

6.9. Ubuntu

6.9.1. Prepare Ubuntu

1. Perform the pre-installation actions.
2. The kernel headers and development packages for the currently running kernel can be installed with:
   ```bash
   sudo apt-get install linux-headers-$(uname -r)
   ```
3. Remove Outdated Signing Key:
   ```bash
   sudo apt-key del 7fa2af80
   ```
4. Choose an installation method: local repo or network repo.

6.9.2. Local Repo Installation for Ubuntu

1. Install local repository on file system:
   ```bash
   sudo dpkg -i cuda-repo-<distro>_<version>_<architecture>.deb
   ```
2. Enroll ephemeral public GPG key:
   ```bash
   sudo cp /var/cuda-repo-<distro>-X-Y-local/cuda-*-keyring.gpg /usr/share/keyrings/
   ```
3. Add pin file to prioritize CUDA repository:
   ```bash
   sudo mv cuda-<distro>.pin /etc/apt/preferences.d/cuda-repository-pin-600
   ```

6.9.3. Network Repo Installation for Ubuntu

The new GPG public key for the CUDA repository is 3bf863cc. This must be enrolled on the system, either using the cuda-keyring package or manually; the apt-key command is deprecated and not recommended.

1. Install the new cuda-keyring package:
   ```bash
   sudo dpkg -i cuda-keyring_1.1-1_all.deb
   ```
where $distro/$arch should be replaced by one of the following:

▶ ubuntu1604/x86_64
▶ ubuntu1804/cross-linux-sbsa
▶ ubuntu1804/sbsa
▶ ubuntu1804/x86_64
▶ ubuntu2004/cross-linux-aarch64
▶ ubuntu2004/arm64
▶ ubuntu2004/cross-linux-sbsa
▶ ubuntu2004/sbsa
▶ ubuntu2004/x86_64
▶ ubuntu2204/sbsa
▶ ubuntu2204/x86_64

**Note:** arm64-Jetson repos:

▶ native: $distro/arm64
▶ cross: $distro/cross-linux-aarch64

```
sudo dpkg -i cuda-keyring_1.1-1_all.deb
```

2. Or if you are unable to install the cuda-keyring package, you can optionally:

   a. **Enroll the new signing key manually:**

   ```
   --cuda-archive-keyring.gpg
   sudo mv cuda-archive-keyring.gpg /usr/share/keyrings/cuda-archive-keyring.gpg
   ```

   b. **Enable the network repository:**

   ```
   echo "deb [signed-by=/usr/share/keyrings/cuda-archive-keyring.gpg] https://
   ...developer.download.nvidia.com/compute/cuda/repos/<distro>/<arch>/ /" | sudo
   tee /etc/apt/sources.list.d/cuda-<distro>-<arch>.list
   ```

   c. **Add pin file to prioritize CUDA repository:**

   ```
   --cuda-<distro>.pin
   sudo mv cuda-<distro>.pin /etc/apt/preferences.d/cuda-repository-pin-600
   ```

6.9. Ubuntu
6.9.4. Common Installation Instructions for Ubuntu

These instructions apply to both local and network installation for Ubuntu.

1. **Update the Apt repository cache:**
   
   ```bash
   sudo apt-get update
   ```

2. **Install CUDA SDK:**
   
   **Note:** These two commands must be executed separately.
   
   ```bash
   sudo apt-get install cuda-toolkit
   To include all GDS packages:
   sudo apt-get install nvidia-gds
   ```

3. **Reboot the system**
   
   ```bash
   sudo reboot
   ```

4. Perform the Post-installation Actions

---

6.10. Debian

6.10.1. Prepare Debian

1. Perform the pre-installation actions.

2. The kernel headers and development packages for the currently running kernel can be installed with:
   
   ```bash
   sudo apt-get install linux-headers-$(uname -r)
   ```

3. **Enable the contrib repository:**
   
   ```bash
   sudo add-apt-repository contrib
   ```

4. **Remove Outdated Signing Key:**
   
   ```bash
   sudo apt-key del 7fa2af80
   ```

5. Choose an installation method: local repo or network repo.
6.10.2. Local Repo Installation for Debian

1. **Install local repository on file system:**

   ```bash
   sudo dpkg -i cuda-repo-<distro>-X-Y-local_<version>*_x86_64.deb
   ```

2. **Enroll ephemeral public GPG key:**

   ```bash
   sudo cp /var/cuda-repo-<distro>-X-Y-local/cuda-*keyring.gpg /usr/share/keyrings/
   ```

6.10.3. Network Repo Installation for Debian

The new GPG public key for the CUDA repository (Debian-based distros) is 3bf863cc. This must be enrolled on the system, either using the cuda-keyring package or manually; the apt-key command is deprecated and not recommended.

1. **Install the new cuda-keyring package:**

   ```bash
   cuda-keyring_1.1-1_all.deb
   ```

   where `$distro/$arch` should be replaced by one of the following:
   - debian10/x86_64
   - debian11/x86_64

   ```bash
   sudo dpkg -i cuda-keyring_1.1-1_all.deb
   ```

2. Or if you are unable to install the cuda-keyring package, you can optionally:
   a. **Enroll the new signing key manually:**

      ```bash
      wget https://developer.download.nvidia.com/compute/cuda/repos/<distro>/x86_64/
      cuda-archive-keyring.gpg
      sudo mv cuda-archive-keyring.gpg /usr/share/keyrings/cuda-archive-keyring.gpg
      ```

   b. **Enable the network repository:**

      ```bash
      echo "deb [signed-by=/usr/share/keyrings/cuda-archive-keyring.gpg] https://
      developer.download.nvidia.com/compute/cuda/repos/<distro>/x86_64/ /" | sudo
      tee /etc/apt/sources.list.d/cuda-<distro>-x86_64.list
      ```

6.10.4. Common Installation Instructions for Debian

These instructions apply to both local and network installation for Debian.

1. **Update the Apt repository cache:**

   ```bash
   sudo apt-get update
   ```

   **Note:** If you are using Debian 10, you may instead need to run:
Installation Guide for Linux, Release 12.5

sudo apt-get --allow-releaseinfo-change update

2. **Install CUDA SDK:**
   
   sudo apt-get -y install cuda

3. **Reboot the system:**
   
   sudo reboot

4. Perform the post-installation actions.

---

### 6.11. Amazon Linux 2023

#### 6.11.1. Prepare Amazon Linux 2023

1. **Perform the pre-installation actions.**

2. The kernel headers and development packages for the currently running kernel can be installed with:
   
   ```
   sudo dnf install kernel-devel-$(uname -r) kernel-headers-$(uname -r) kernel-
   modules-extra-$(uname -r)
   ```

3. Choose an installation method: **local repo** or **network repo**.

---

#### 6.11.2. Local Repo Installation for Amazon Linux

1. **Install local repository on file system:**
   
   ```
   sudo rpm --install cuda-repo-amzn2023-X-Y-local-<version>*.x86_64.rpm
   ```

---

#### 6.11.3. Network Repo Installation for Amazon Linux

1. **Enable the network repository:**
   
   ```
   cuda/repos/amzn2023/x86_64/cuda-amzn2023.repo
   ```

2. **Clean DNF repository cache:**
   
   ```
   sudo dnf clean expire-cache
   ```
6.11.4. Common Installation Instructions for Amazon Linux

These instructions apply to both local and network installation for Amazon Linux.

1. **Install CUDA SDK:**
   
   ```
   sudo dnf module install nvidia-driver:latest-dkms
   sudo dnf install cuda-toolkit
   ```

2. **Install GPUDirect Filesystem:**
   
   ```
   sudo dnf install nvidia-gds
   ```

3. **Add libcuda.so symbolic link, if necessary:**
   
   The `libcuda.so` library is installed in the `/usr/lib{,64}/nvidia` directory. For pre-existing projects which use `libcuda.so`, it may be useful to add a symbolic link from `libcuda.so` in the `/usr/lib{,64}` directory.

4. **Reboot the system:**
   
   ```
   sudo reboot
   ```

5. **Perform the post-installation actions.**

6.12. Additional Package Manager Capabilities

Below are some additional capabilities of the package manager that users can take advantage of.

6.12.1. Available Packages

The recommended installation package is the `cuda` package. This package will install the full set of other CUDA packages required for native development and should cover most scenarios.

The `cuda` package installs all the available packages for native developments. That includes the compiler, the debugger, the profiler, the math libraries, and so on. For x86_64 platforms, this also includes Nsight Eclipse Edition and the visual profilers. It also includes the NVIDIA driver package.

On supported platforms, the `cuda-cross-aarch64` and `cuda-cross-sbsa` packages install all the packages required for cross-platform development to arm64-Jetson and arm64-Server, respectively. The libraries and header files of the target architecture’s display driver package are also installed to enable the cross compilation of driver applications. The `cuda-cross-<arch>` packages do not install the native display driver.

**Note:** 32-bit compilation native and cross-compilation is removed from CUDA 12.0 and later Toolkit. Use the CUDA Toolkit from earlier releases for 32-bit compilation. CUDA Driver will continue to support running existing 32-bit applications on existing GPUs except Hopper. Hopper does not support 32-bit applications. Ada will be the last architecture with driver support for 32-bit applications.
The packages installed by the packages above can also be installed individually by specifying their names explicitly. The list of available packages can be obtained with:

```
ym -d disablenrepo="*" --enablerepo="cuda*" list available  # RedHat
```

```
dnf -d disablenrepo="*" --enablerepo="cuda*" list available  # Fedora
```

```
dnpp packages -r cuda  # OpenSUSE & SLES
```

```
cat /var/lib/apt/lists/*/cuda*Packages | grep "Package:"  # Ubuntu
```

### 6.12.2. Meta Packages

Meta packages are RPM/Deb/Conda packages which contain no (or few) files but have multiple dependencies. They are used to install many CUDA packages when you may not know the details of the packages you want. The following table lists the meta packages.

<table>
<thead>
<tr>
<th>Meta Package</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>cuda</td>
<td>Installs all CUDA Toolkit and Driver packages. Handles upgrading to the next version of the cuda package when it’s released.</td>
</tr>
<tr>
<td>cuda-12-5</td>
<td>Installs all CUDA Toolkit and Driver packages. Remains at version 12.5 until an additional version of CUDA is installed.</td>
</tr>
<tr>
<td>cuda-toolkit-12-5</td>
<td>Installs all CUDA Toolkit packages required to develop CUDA applications. Does not include the driver.</td>
</tr>
<tr>
<td>cuda-toolkit-15</td>
<td>Installs all CUDA Toolkit packages required to develop applications. Will not upgrade beyond the 12.x series toolkits. Does not include the driver.</td>
</tr>
<tr>
<td>cuda-toolkit</td>
<td>Installs all CUDA Toolkit packages required to develop applications. Handles upgrading to the next 12.x version of CUDA when it’s released. Does not include the driver.</td>
</tr>
<tr>
<td>cuda-tools-12-5</td>
<td>Installs all CUDA command line and visual tools.</td>
</tr>
<tr>
<td>cuda-runtime-12-5</td>
<td>Installs all CUDA Toolkit packages required to run CUDA applications, as well as the Driver packages.</td>
</tr>
<tr>
<td>cuda-compiler-12-5</td>
<td>Installs all CUDA compiler packages.</td>
</tr>
<tr>
<td>cuda-libraries-12-5</td>
<td>Installs all runtime CUDA Library packages.</td>
</tr>
<tr>
<td>cuda-libraries-dev-12-5</td>
<td>Installs all development CUDA Library packages.</td>
</tr>
<tr>
<td>cuda-drivers</td>
<td>Installs all NVIDIA Driver packages with proprietary kernel modules. Handles upgrading to the next version of the Driver packages when they’re released.</td>
</tr>
<tr>
<td>cuda-drivers-555</td>
<td>Installs all NVIDIA Driver packages with proprietary kernel modules. Will not upgrade beyond the 555 branch drivers.</td>
</tr>
</tbody>
</table>
6.12.3. Optional 32-bit Packages for Linux x86_64 .deb/.rpm

These packages provide 32-bit driver libraries needed for things such as Steam (popular game app store/launcher), older video games, and some compute applications.

For Debian 10 and Debian 11:

```
sudo dpkg --add-architecture i386
sudo apt-get update
sudo apt-get install libcudal1-i386 nvidia-driver-libs-i386
```

For Debian 12:

```
sudo dpkg --add-architecture i386
sudo apt-get update
apt install nvidia-driver-libs:i386
```

For Ubuntu:

```
sudo dpkg --add-architecture i386
sudo apt-get update
sudo apt-get install libnvidia-compute-<branch>:i386 libnvidia-decode-<branch>:i386 \ libnvidia-encode-<branch>:i386 libnvidia-extra-<branch>:i386 libnvidia-fbc1-<branch> \ :i386 \ libnvidia-gl-<branch>:i386
```

Where <branch> is the driver version, for example 495.

For Fedora and RHEL8+:

```
sudo dnf install nvidia-driver-cuda-libs.i686 nvidia-driver-devel.i686 \ nvidia-driver-libs.i686 nvidia-driver-NvFBCOpenGL.i686 nvidia-driver-NVML.i686
```

**Note:** There is no modularity profile support.

For openSUSE/SLES:

```
sudo zypper install nvidia-compute-G06-32bit nvidia-gl-G06-32bit nvidia-video-G06- \ .32bit
```

6.12.4. Package Upgrades

The cuda package points to the latest stable release of the CUDA Toolkit. When a new version is available, use the following commands to upgrade the toolkit and driver:

```
sudo dnf install cuda-toolkit # Fedora, RHEL9, RHEL8, \ and KylinOS
```

```
sudo zypper install cuda-toolkit # OpenSUSE and SLES
```
Installation Guide for Linux, Release 12.5

```
sudo apt-get install cuda-toolkit       # Ubuntu and Debian
```

The `cuda-cross-<arch>` packages can also be upgraded in the same manner.

The `cuda-drivers` package points to the latest driver release available in the CUDA repository. When a new version is available, use the following commands to upgrade the driver:

```
sudo dnf module install nvidia-driver:latest-dkms  # Fedora, RHEL9, RHEL8, ...
```

```
sudo zypper install cuda-drivers nvidia-gfxG04-kmp-default  # OpenSUSE and SLES
```

```
sudo apt-get install cuda-drivers            # Ubuntu and Debian
```

Some desktop environments, such as GNOME or KDE, will display a notification alert when new packages are available.

To avoid any automatic upgrade, and lock down the toolkit installation to the XY release, install the `cuda-X-Y` or `cuda-cross-<arch>-X-Y` package.

Side-by-side installations are supported. For instance, to install both the XY CUDA Toolkit and the X.Y+1 CUDA Toolkit, install the `cuda-X.Y` and `cuda-X.Y+1` packages.
Chapter 7. Driver Installation

This section is for users who want to install a specific driver version.

**For Debian and Ubuntu:**

```bash
sudo apt-get install cuda-drivers-<driver_branch>
```

For example:

```bash
sudo apt-get install cuda-drivers-535
```

**For OpenSUSE and SLES:**

```bash
sudo zypper -v install cuda-drivers-<driver_branch>
```

For example:

```bash
sudo zypper -v install cuda-drivers-550
```

This allows you to get the highest version in the specified branch.

**For Fedora and RHEL8+:**

```bash
sudo dnf module install nvidia-driver:<stream>∕<profile>
```

where profile by default is "default" and does not need to be specified.

- Example dkms streams: `450-dkms` or `latest-dkms`
- Example precompiled streams: `450` or `latest`

**Note:** Precompiled streams are only supported on RHEL8 x86_64 and RHEL9 x86_64.

To uninstall or change streams on Fedora and RHEL8:

```bash
sudo dnf module remove --all nvidia-driver
sudo dnf module reset nvidia-driver
```
Chapter 8. NVIDIA Open GPU Kernel Modules

The NVIDIA Linux GPU Driver contains several kernel modules:

- nvidia.ko
- nvidia-modeset.ko
- nvidia-uvm.ko
- nvidia-drm.ko
- nvidia-peermem.ko

Starting in the 515 driver release series, two "flavors" of these kernel modules are provided:

- Proprietary - this is the flavor that NVIDIA has historically shipped.
- Open-source - published kernel modules that are dual licensed MIT/GPLv2. These are new starting in release 515. With every driver release, the source code to the open kernel modules will be published on [https://github.com/NVIDIA/open-gpu-kernel-modules](https://github.com/NVIDIA/open-gpu-kernel-modules) and a tarball will be provided on [https://download.nvidia.com/XFree86/](https://download.nvidia.com/XFree86/).

Verify that your NVIDIA GPU is at least Turing or newer generation.

```
lspci | grep VGA
```

Experimental support for GeForce and Quadro SKUs can be enabled with:

```
echo "options nvidia NVreg_OpenRmEnableUnsupportedGpus=1" | sudo tee /etc/modprobe.d/
   ...nvidia-gsp.conf
```

To install NVIDIA Open GPU Kernel Modules, follow the instructions below.

8.1. CUDA Runfile

Pass the CLI argument to the CUDA runfile to opt in to NVIDIA Open GPU Kernel Modules:

```
sh cuda_<release>_<version>_linux.run -m=kernel-open
```
8.2. Debian

1. Install the NVIDIA Open GPU Kernel Modules package:
   ```bash
   sudo apt-get install nvidia-kernel-open-dkms
   ```

2. Install the rest of the NVIDIA driver packages:
   ```bash
   sudo apt-get install cuda-drivers
   ```

**OR to install a specific driver version**

1. Install the NVIDIA Open GPU Kernel Modules package:
   ```bash
   sudo apt-get install -v nvidia-kernel-open-dkms=<version>-1
   ```

2. Install the rest of the NVIDIA driver packages:
   ```bash
   sudo apt-get install -v cuda-drivers=<driver_branch>
   ```

For example:
   ```bash
   sudo apt-get install -v nvidia-kernel-open-dkms=550.90.07-1
   sudo apt-get install -v cuda-drivers-550
   ```

8.3. Fedora

1. Install the NVIDIA Open GPU Kernel Modules package and the rest of the NVIDIA driver packages:
   ```bash
   sudo dnf module install nvidia-driver:open-dkms
   ```

**OR to install a specific driver version**

1. Install the NVIDIA Open GPU Kernel Modules package and the rest of the NVIDIA driver packages:
   ```bash
   sudo dnf module install nvidia-driver:<driver_branch>-open
   ```

8.4. KylinOS 10

1. Install the NVIDIA Open GPU Kernel Modules package and the rest of the NVIDIA driver packages:
   ```bash
   sudo dnf module install nvidia-driver:open-dkms
   ```

**OR to install a specific driver version**

1. Install the NVIDIA Open GPU Kernel Modules package and the rest of the NVIDIA driver packages:
   ```bash
   sudo dnf module install nvidia-driver:<driver_branch>-open
   ```
8.5. RHEL 9 and Rocky 9

1. Install the NVIDIA Open GPU Kernel Modules package and the rest of the NVIDIA driver packages:

   ```
sudo dnf module install nvidia-driver:open-dkms
   ```

**OR to install a specific driver version**

1. Install the NVIDIA Open GPU Kernel Modules package and the rest of the NVIDIA driver packages:

   ```
sudo dnf module install nvidia-driver:<driver_branch>-open
   ```

8.6. RHEL 8 and Rocky 8

1. Install the NVIDIA Open GPU Kernel Modules package and the rest of the NVIDIA driver packages:

   ```
sudo dnf module install nvidia-driver:open-dkms
   ```

**OR to install a specific driver version**

1. Install the NVIDIA Open GPU Kernel Modules package and the rest of the NVIDIA driver packages:

   ```
sudo dnf module install nvidia-driver:<driver_branch>-open
   ```

8.7. OpenSUSE and SLES

1. Install the NVIDIA Open GPU Kernel Modules package:

   ```
sudo zypper install nvidia-open-driver-G06-kmp-default
   ```

2. Install the rest of the NVIDIA driver packages:

   ```
sudo zypper install cuda-drivers
   ```

**OR to install a specific driver version**

1. Install the NVIDIA Open GPU Kernel Modules package:

   ```
sudo zypper -v install $(zypper search -s nvidia-open-driver-G06-kmp-<flavor> | sed 's| ||g' | awk -F '|' '{print $2"="$4}')
   ```

2. Install the rest of the NVIDIA driver packages:

   ```
sudo zypper -v install cuda-drivers-<driver_branch>
   ```

   For example:

   ```
sudo zypper -v install $(zypper search -s nvidia-open-driver-G06-kmp-default | sed 's| ||g' | awk -F '|' '{print $2"="$4}')
sudo zypper -v install cuda-drivers-550
   ```
8.8. Ubuntu

1. Install the NVIDIA Open GPU Kernel Modules package:
   `sudo apt-get install nvidia-driver-<driver_branch>-open`

2. Install the rest of the NVIDIA driver packages:
   `sudo apt-get install cuda-drivers-<driver_branch>`

**Note:** End-users on Ubuntu should upgrade their NVIDIA Open GPU kernel modules using the following:

   `sudo apt-get install --verbose-versions nvidia-kernel-source-550-open cuda-drivers-550`

**OR to install a specific driver version**

1. Install the NVIDIA Open GPU Kernel Modules package:
   `sudo apt-get install -v nvidia-driver-<driver_branch>-open`

2. Install the rest of the NVIDIA driver packages:
   `sudo apt-get install -v cuda-drivers-<driver_branch>`

For example:

   `sudo apt-get install -v nvidia-driver-550-open`
   `sudo apt-get install -v cuda-drivers-550`
Chapter 9. Precompiled Streams

Precompiled streams offer an optional method of streamlining the installation process.

The advantages of precompiled streams:

- Precompiled: faster boot up after driver and/or kernel updates
- Pre-tested: kernel and driver combination has been validated
- Removes gcc dependency: no compiler installation required
- Removes dkms dependency: enabling EPEL repository not required
- Removes kernel-devel and kernel-headers dependencies: no black screen if matching packages are missing

When using precompiled drivers, a plugin for the dnf package manager is enabled that cleans up stale .ko files. To prevent system breakages, the NVIDIA dnf plugin also prevents upgrading to a kernel for which no precompiled driver yet exists. This can delay the application of security fixes but ensures that a tested kernel and driver combination is always used. A warning is displayed by dnf during that upgrade situation:

```
NOTE: Skipping kernel installation since no NVIDIA driver kernel module package kmod-nvidia-{$driver}-${{kernel}} ... could be found
```

Packaging templates and instructions are provided on GitHub to allow you to maintain your own precompiled kernel module packages for custom kernels and derivative Linux distros: NVIDIA/yum-packaging-precompiled-kmod

To use the new driver packages on RHEL 8 or RHEL 9:

1. First, ensure that the Red Hat repositories are enabled:
   
   **RHEL 8:**
   ```
   subscription-manager repos --enable=rhel-8-for-x86_64-appstream-rpms
   subscription-manager repos --enable=rhel-8-for-x86_64-baseos-rpms
   ```
   or
   
   **RHEL 9:**
   ```
   subscription-manager repos --enable=rhel-9-for-x86_64-appstream-rpms
   subscription-manager repos --enable=rhel-9-for-x86_64-baseos-rpms
   ```

2. Choose one of the four options below depending on the desired driver:
   ```
   ▶ latest always updates to the highest versioned driver (precompiled):
   ```
Installation Guide for Linux, Release 12.5

```bash
sudo dnf module install nvidia-driver:latest
```

▶ `<id>` locks the driver updates to the specified driver branch (precompiled):
  ```bash
  sudo dnf module install nvidia-driver:<id>
  ```

**Note:** Replace `<id>` with the appropriate driver branch streams, for example 520, 515, 470, or 450.

▶ `latest-dkms` always updates to the highest versioned driver (non-precompiled):
  ```bash
  sudo dnf module install nvidia-driver:latest-dkms
  ```

**Note:** This is the default stream.

▶ `<id>-dkms` locks the driver updates to the specified driver branch (non-precompiled):
  ```bash
  sudo dnf module install nvidia-driver:<id>-dkms
  ```

**Note:** Valid streams include 520-dkms, 515-dkms, 470-dkms, and 450-dkms.

## 9.1. Precompiled Streams Support Matrix

This table shows the supported precompiled and legacy DKMS streams for each driver.

<table>
<thead>
<tr>
<th>NVIDIA Driver</th>
<th>Precompiled Stream</th>
<th>Legacy DKMS Stream</th>
<th>Open DKMS Stream</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highest version</td>
<td>latest</td>
<td>latest-dkms</td>
<td>open-dkms</td>
</tr>
<tr>
<td>Locked at 520.x</td>
<td>520</td>
<td>520-dkms</td>
<td>520-open</td>
</tr>
<tr>
<td>Locked at 515.x</td>
<td>515</td>
<td>515-dkms</td>
<td>515-open</td>
</tr>
</tbody>
</table>

Prior to switching between module streams, first reset:
```
sudo dnf module reset nvidia-driver
```

**Note:** This is also required for upgrading between branch locked streams.

Or alternatively:
```
sudo dnf module switch-to nvidia-driver:<stream>
```
9.2. Modularity Profiles

Modularity profiles work with any supported modularity stream and allow for additional use cases. These modularity profiles are available on RHEL8+ and Fedora.

Table 6: List of nvidia-driver Module Profiles

<table>
<thead>
<tr>
<th>Stream</th>
<th>Profile</th>
<th>Use Case</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default</td>
<td>default</td>
<td>Installs all the driver packages in a stream.</td>
</tr>
<tr>
<td>Kickstart</td>
<td>ks</td>
<td>Performs unattended Linux OS installation using a config file.</td>
</tr>
<tr>
<td>NVSwitch Fabric</td>
<td>fm</td>
<td>Installs all the driver packages plus components required for bootstrapping an NVSwitch system (including the Fabric Manager and NSCQ telemetry).</td>
</tr>
<tr>
<td>Source</td>
<td>src</td>
<td>Source headers for compilation (precompiled streams only).</td>
</tr>
</tbody>
</table>

For example:

```bash
sudo dnf module nvidia-driver:<stream>/default
sudo dnf module nvidia-driver:<stream>/ks
sudo dnf module nvidia-driver:<stream>/fm
sudo dnf module nvidia-driver:<stream>/src
```

You can install multiple modularity profiles using BASH curly brace expansion, for example:

```bash
sudo dnf module install nvidia-driver:latest/{default,src}
```

Chapter 10. Kickstart Installation

10.1. RHEL 8 / Rocky Linux 8

1. Enable the EPEL repository:
   
   ```
   repo --name=epel --baseurl=http://download.fedoraproject.org/pub/epel/8/
   →Everything/x86_64/
   ```

2. Enable the CUDA repository:
   
   ```
   repo --name=cuda-rhel8 --baseurl=https://developer.download.nvidia.com/compute/
   →cuda/repos/rhel8/x86_64/
   ```

3. In the packages section of the ks.cfg file, make sure you are using the /ks profile and :latest-dkms stream:
   
   ```
   @nvidia-driver:latest-dkms/ks
   ```

4. Perform the post-installation actions.

10.2. RHEL 9 / Rocky Linux 9

1. Enable the EPEL repository:
   
   ```
   repo --name=epel --baseurl=http://download.fedoraproject.org/pub/epel/9/
   →Everything/x86_64/
   ```

2. Enable the CUDA repository:
   
   ```
   repo --name=cuda-rhel9 --baseurl=https://developer.download.nvidia.com/compute/
   →cuda/repos/rhel9/x86_64/
   ```

3. In the packages section of the ks.cfg file, make sure you are using the /ks profile and :latest-dkms stream:
   
   ```
   @nvidia-driver:latest-dkms/ks
   ```

4. Perform the post-installation actions.
Chapter 11. Runfile Installation

Basic instructions can be found in the Quick Start Guide. Read on for more detailed instructions.
This section describes the installation and configuration of CUDA when using the standalone installer.
The standalone installer is a ".run" file and is completely self-contained.

11.1. Runfile Overview

The Runfile installation installs the NVIDIA Driver and CUDA Toolkit via an interactive ncurses-based interface.
The installation steps are listed below. Distribution-specific instructions on disabling the Nouveau drivers as well as steps for verifying device node creation are also provided.
Finally, advanced options for the installer and uninstallation steps are detailed below.
The Runfile installation does not include support for cross-platform development. For cross-platform development, see the CUDA Cross-Platform Environment section.

11.2. Installation

1. Perform the pre-installation actions.
2. Disable the Nouveau drivers.
3. Reboot into text mode (runlevel 3).
   This can usually be accomplished by adding the number “3” to the end of the system’s kernel boot parameters.
   Since the NVIDIA drivers are not yet installed, the text terminals may not display correctly. Temporarily adding “nomodeset” to the system’s kernel boot parameters may fix this issue.
   Consult your system’s bootloader documentation for information on how to make the above boot parameter changes.
   The reboot is required to completely unload the Nouveau drivers and prevent the graphical interface from loading. The CUDA driver cannot be installed while the Nouveau drivers are loaded or while the graphical interface is active.
4. Verify that the Nouveau drivers are not loaded. If the Nouveau drivers are still loaded, consult your distribution’s documentation to see if further steps are needed to disable Nouveau.
5. Run the installer and follow the on-screen prompts:

```
sudo sh cuda_<version>_linux.run
```

The installer will prompt for the following:

- EULA Acceptance
- CUDA Driver installation
- CUDA Toolkit installation, location, and /usr/local/cuda symbolic link

The default installation location for the toolkit is /usr/local/cuda-12.4:

The /usr/local/cuda symbolic link points to the location where the CUDA Toolkit was installed. This link allows projects to use the latest CUDA Toolkit without any configuration file update.

The installer must be executed with sufficient privileges to perform some actions. When the current privileges are insufficient to perform an action, the installer will ask for the user’s password to attempt to install with root privileges. Actions that cause the installer to attempt to install with root privileges are:

- installing the CUDA Driver
- installing the CUDA Toolkit to a location the user does not have permission to write to
- creating the /usr/local/cuda symbolic link

Running the installer with `sudo`, as shown above, will give permission to install to directories that require root permissions. Directories and files created while running the installer with `sudo` will have root ownership.

If installing the driver, the installer will also ask if the OpenGL libraries should be installed. If the GPU used for display is not an NVIDIA GPU, the NVIDIA openGL libraries should not be installed. Otherwise, the openGL libraries used by the graphics driver of the non-NVIDIA GPU will be overwritten and the GUI will not work. If performing a silent installation, the `--no-opengl-libs` option should be used to prevent the openGL libraries from being installed. See the Advanced Options section for more details.

If the GPU used for display is an NVIDIA GPU, the X server configuration file, /etc/X11/xorg.conf, may need to be modified. In some cases, nvidia-xconfig can be used to automatically generate an xorg.conf file that works for the system. For non-standard systems, such as those with more than one GPU, it is recommended to manually edit the xorg.conf file. Consult the xorg.conf documentation for more information.

**Note:** Installing Mesa may overwrite the /usr/lib/libGL.so that was previously installed by the NVIDIA driver, so a reinstallation of the NVIDIA driver might be required after installing these libraries.

6. Reboot the system to reload the graphical interface:

```
sudo reboot
```

7. Verify the device nodes are created properly.

8. Perform the post-installation actions.
11.3. Disabling Nouveau

To install the Display Driver, the Nouveau drivers must first be disabled. Each distribution of Linux has a different method for disabling Nouveau.

The Nouveau drivers are loaded if the following command prints anything:

```bash
dmesg | grep nouveau
```

11.3.1. Fedora

1. Create a file at `/usr/lib/modprobe.d/blacklist-nouveau.conf` with the following contents:

   ```
   blacklist nouveau
   options nouveau modeset=0
   ```

2. Regenerate the kernel initramfs:

   ```bash
   sudo dracut --force
   ```

3. Run the following command:

   ```bash
   sudo grub2-mkconfig -o /boot/grub2/grub.cfg
   ```

4. Reboot the system.

11.3.2. RHEL / Rocky and KylinOS

1. Create a file at `/etc/modprobe.d/blacklist-nouveau.conf` with the following contents:

   ```
   blacklist nouveau
   options nouveau modeset=0
   ```

2. Regenerate the kernel initramfs:

   ```bash
   sudo dracut --force
   ```

11.3.3. OpenSUSE

1. Create a file at `/etc/modprobe.d/blacklist-nouveau.conf` with the following contents:

   ```
   blacklist nouveau
   options nouveau modeset=0
   ```

2. Regenerate the kernel initrd:
11.3.4. SLES

No actions to disable Nouveau are required as Nouveau is not installed on SLES.

11.3.5. WSL

No actions to disable Nouveau are required as Nouveau is not installed on WSL.

11.3.6. Ubuntu

1. Create a file at `/etc/modprobe.d/blacklist-nouveau.conf` with the following contents:
   
   ```
   blacklist nouveau
   options nouveau modeset=0
   ```

2. Regenerate the kernel initramfs:
   
   ```
   sudo update-initramfs -u
   ```

11.3.7. Debian

1. Create a file at `/etc/modprobe.d/blacklist-nouveau.conf` with the following contents:
   
   ```
   blacklist nouveau
   options nouveau modeset=0
   ```

2. Regenerate the kernel initramfs:
   
   ```
   sudo update-initramfs -u
   ```

11.4. Device Node Verification

Check that the device files `/dev/nvidia*` exist and have the correct (0666) file permissions. These files are used by the CUDA Driver to communicate with the kernel-mode portion of the NVIDIA Driver. Applications that use the NVIDIA driver, such as a CUDA application or the X server (if any), will normally automatically create these files if they are missing using the `setuidnvidia-modprobe` tool that is bundled with the NVIDIA Driver. However, some systems disallow setuid binaries, so if these files do not exist, you can create them manually by using a startup script such as the one below:
#!/bin/bash
/sbin/modprobe nvidia
if [ "$?" -eq 0 ]; then
    # Count the number of NVIDIA controllers found.
    NVDEV=`lspci | grep -i NVIDIA`
    N3D=`echo "$NVDEV" | grep "3D controller" | wc -l`
    NVGA=`echo "$NVDEV" | grep "VGA compatible controller" | wc -l`
    N=`expr $N3D + $NVGA - 1`
    for i in `seq 0 $N`; do
        mknod -m 666 /dev/nvidia$i c 195 $i
    done
    mknod -m 666 /dev/nvidiactl c 195 255
else
    exit 1
fi
/sbin/modprobe nvidia-uvm
if [ "$?" -eq 0 ]; then
    # Find out the major device number used by the nvidia-uvm driver
    D=`grep nvidia-uvm /proc/devices | awk '{print $1}'`
    mknod -m 666 /dev/nvidia-uvm c $D 0
else
    exit 1
fi
## 11.5. Advanced Options

<table>
<thead>
<tr>
<th>Action</th>
<th>Options Used</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Silent Installation</td>
<td>--silent</td>
<td>Required for any silent installation. Performs an installation with no further user-input and minimal command-line output based on the options provided below. Silent installations are useful for scripting the installation of CUDA. Using this option implies acceptance of the EULA. The following flags can be used to customize the actions taken during installation. At least one of --driver, --uninstall, and --toolkit must be passed if running with non-root permissions.</td>
</tr>
<tr>
<td></td>
<td>--driver</td>
<td>Install the CUDA Driver.</td>
</tr>
<tr>
<td></td>
<td>--toolkit</td>
<td>Install the CUDA Toolkit.</td>
</tr>
<tr>
<td></td>
<td>--toolkitpath=&lt;path&gt;</td>
<td>Install the CUDA Toolkit to the &lt;path&gt; directory. If not provided, the default path of /usr/local/cuda-12.4 is used.</td>
</tr>
<tr>
<td></td>
<td>--defaultroot=&lt;path&gt;</td>
<td>Install libraries to the &lt;path&gt; directory. If the &lt;path&gt; is not provided, then the default path of your distribution is used. This only applies to the libraries installed outside of the CUDA Toolkit path.</td>
</tr>
<tr>
<td>Extraction</td>
<td>--extract=&lt;path&gt;</td>
<td>Extracts to the &lt;path&gt; the following: the driver runfile, the raw files of the toolkit to &lt;path&gt;. This is especially useful when one wants to install the driver using one or more of the command-line options provided by the driver installer which are not exposed in this installer.</td>
</tr>
<tr>
<td>Overriding Installation Checks</td>
<td>--override</td>
<td>Ignores compiler, third-party library, and toolkit detection checks which would prevent the CUDA Toolkit from installing.</td>
</tr>
<tr>
<td>No OpenGL Libraries</td>
<td>--no-opengl-libs</td>
<td>Prevents the driver installation from installing NVIDIA's GL libraries. Useful for systems where the display is driven by a non-NVIDIA GPU. In such systems, NVIDIA's GL libraries could prevent X from loading properly.</td>
</tr>
<tr>
<td>No man pages</td>
<td>--no-man-page</td>
<td>Do not install the man pages under /usr/share/man.</td>
</tr>
<tr>
<td>Overriding Kernel Source</td>
<td>--kernel-source-path=&lt;path&gt;</td>
<td>Tells the driver installation to use &lt;path&gt; as the kernel source directory when building the NVIDIA kernel module. Required for systems where the kernel source is installed to a non-standard location.</td>
</tr>
<tr>
<td>Running nvidia-xconfig</td>
<td>--run-nvidia-xconfig</td>
<td>Tells the driver installation to run nvidia-xconfig to update the system X configuration file so that the NVIDIA X driver is used. The pre-existing X configuration file will be backed up.</td>
</tr>
<tr>
<td>No nvidia-drm kernel module</td>
<td>--no-drm</td>
<td>Do not install the nvidia-drm kernel module. This option should only be used to work around failures to build or install the nvidia-drm kernel module on systems that do not need the provided features.</td>
</tr>
<tr>
<td>Custom Temporary Directory Selection</td>
<td>--tmpdir=&lt;path&gt;</td>
<td>Performs any temporary actions within a subdirectory of /tmp. Useful in cases where /tmp cannot be used (doesn't exist, is full, is mounted with 'noexec', etc.).</td>
</tr>
<tr>
<td>Show Installer Options</td>
<td>--help</td>
<td>Prints the list of command-line options to stdout.</td>
</tr>
</tbody>
</table>
11.6. Uninstallation

To uninstall the CUDA Toolkit, run the uninstallation script provided in the bin directory of the toolkit. By default, it is located in `/usr/local/cuda-12.4/bin`:

```
sudo /usr/local/cuda-12.4/bin/cuda-uninstaller
```

To uninstall the NVIDIA Driver, run `nvidia-uninstall`:

```
sudo /usr/bin/nvidia-uninstall
```

To enable the Nouveau drivers, remove the blacklist file created in the Disabling Nouveau section, and regenerate the kernel initramfs/initrd again as described in that section.
Chapter 12. Conda Installation

This section describes the installation and configuration of CUDA when using the Conda installer. The Conda packages are available at https://anaconda.org/nvidia.

12.1. Conda Overview

The Conda installation installs the CUDA Toolkit. The installation steps are listed below.

12.2. Installing CUDA Using Conda

To perform a basic install of all CUDA Toolkit components using Conda, run the following command:

```
conda install cuda -c nvidia
```

12.3. Uninstalling CUDA Using Conda

To uninstall the CUDA Toolkit using Conda, run the following command:

```
conda remove cuda
```

12.4. Installing Previous CUDA Releases

All Conda packages released under a specific CUDA version are labeled with that release version. To install a previous version, include that label in the `install` command such as:

```
conda install cuda -c nvidia/label/cuda-11.3.0
```
12.5. Upgrading from cudatoolkit Package

If you had previously installed CUDA using the cudatoolkit package and want to maintain a similar install footprint, you can limit your installation to the following packages:

- cuda-libraries-dev
- cuda-nvcc
- cuda-nvtx
- cuda-cupti

**Note:** Some extra files, such as headers, will be included in this installation which were not included in the cudatoolkit package. If you need to reduce your installation further, replace `cuda-libraries-dev` with the specific libraries you need.
Chapter 13. Pip Wheels

NVIDIA provides Python Wheels for installing CUDA through pip, primarily for using CUDA with Python. These packages are intended for runtime use and do not currently include developer tools (these can be installed separately).

Please note that with this installation method, CUDA installation environment is managed via pip and additional care must be taken to set up your host environment to use CUDA outside the pip environment.

Prerequisites

To install Wheels, you must first install the nvidia-pyindex package, which is required in order to set up your pip installation to fetch additional Python modules from the NVIDIA NGC PyPI repo. If your pip and setuptools Python modules are not up-to-date, then use the following command to upgrade these Python modules. If these Python modules are out-of-date then the commands which follow later in this section may fail.

```bash
python3 -m pip install --upgrade setuptools pip wheel
```

You should now be able to install the nvidia-pyindex module.

```bash
python3 -m pip install nvidia-pyindex
```

If your project is using a requirements.txt file, then you can add the following line to your requirements.txt file as an alternative to installing the nvidia-pyindex package:

```bash
--extra-index-url https://pypi.org/simple
```

Procedure

Install the CUDA runtime package:

```bash
python3 -m pip install nvidia-cuda-runtime-cu12
```

Optionally, install additional packages as listed below using the following command:

```bash
python3 -m pip install nvidia-<library>
```

Metapackages

The following metapackages will install the latest version of the named component on Linux for the indicated CUDA version. "cu12" should be read as "cuda12".

- nvidia-cuda-runtime-cu12
- nvidia-cuda-cccl-cu12
- nvidia-cuda-cupti-cu12
Installation Guide for Linux, Release 12.5

- nvidia-cuda-nvcc-cu12
- nvidia-cuda-opencl-cu12
- nvidia-cuda-nvrtc-cu12
- nvidia-cublas-cu12
- nvidia-cuda-sanitizer-api-cu12
- nvidia-cufft-cu12
- nvidia-cusolver-cu12
- nvidia-cusparse-cu12
- nvidia-npp-cu12
- nvidia-nvfatbin-cu12
- nvidia-nvjitlink-cu12
- nvidia-nvjpeg-cu12
- nvidia-nvml-dev-cu12
- nvidia-nvtx-cu12

These metapackages install the following packages:

- nvidia-cuda-runtime-cu125
- nvidia-cuda-cccl-cu125
- nvidia-cuda-cupti-cu125
- nvidia-cuda-nvcc-cu125
- nvidia-cuda-opencl-cu125
- nvidia-cublas-cu125
- nvidia-cuda-sanitizer-api-cu125
- nvidia-cuda-nvrtc-cu125
- nvidia-cufft-cu125
- nvidia-cusolver-cu125
- nvidia-cusparse-cu125
- nvidia-npp-cu125
- nvidia-nvfatbin-cu125
- nvidia-nvjitlink-cu125
- nvidia-nvjpeg-cu125
- nvidia-nvml-dev-cu125
- nvidia-nvtx-cu125
Chapter 14. Tarball and Zip Archive Deliverables

In an effort to meet the needs of a growing customer base requiring alternative installer packaging formats, as well as a means of input into community CI/CD systems, tarball and zip archives are available for each component.

These tarball and zip archives, known as binary archives, are provided at https://developer.download.nvidia.com/compute/cuda/redist/.

These component .tar.xz and .zip binary archives do not replace existing packages such as .deb, .rpm, runfile, conda, etc. and are not meant for general consumption, as they are not installers. However this standardized approach will replace existing .txz archives.

For each release, a JSON manifest is provided such as redistrib_11.4.2.json, which corresponds to the CUDA 11.4.2 release label (CUDA 11.4 update 2) which includes the release date, the name of each component, license name, relative URL for each platform and checksums.

Package maintainers are advised to check the provided LICENSE for each component prior to redistribution. Instructions for developers using CMake and Bazel build systems are provided in the next sections.
14.1. Parsing Redistrib JSON

The following example of a JSON manifest contains keys for each component: name, license, version, and a platform array which includes relative_path, sha256, md5, and size (bytes) for each archive.

```json
{
  "release_date": "2021-09-07",
  "cuda_cudart": {
    "name": "CUDA Runtime (cudart)",
    "license": "CUDA Toolkit",
    "version": "11.4.108",
    "linux-x86_64": {
      "relative_path": "cuda_cudart/linux-x86_64/cuda_cudart-linux-x86_64-11.4.108-archive.tar.xz",
      "sha256": "d08a1b731e5175aa3ae06a6d1c6b3059dd9ea13836d947018ea5e3ec2ca3d62b",
      "md5": "da1985656b27a3559004c3b7f20e5d074",
      "size": "828300"
    },
    "linux-ppc64le": {
      "relative_path": "cuda_cudart/linux-ppc64le/cuda_cudart-linux-ppc64le-11.4.108-archive.tar.xz",
      "sha256": "831dffe062ae3ebda3d3c4010d0ee4e40a01fd5e6358098a87bb318ea7c79e0c",
      "md5": "ca73328e3f8e2bb5b1f2184c98c3a510",
      "size": "776840"
    },
    "linux-sbsa": {
      "relative_path": "cuda_cudart/linux-sbsa/cuda_cudart-linux-sbsa-11.4.108-archive.tar.xz",
      "sha256": "2ab9599baebebdf59add73d1f1a352ae619f8cb5cc2c254093c98efd4c14553c",
      "md5": "aeb5c19661f066b6398741015ba368102",
      "size": "782372"
    },
    "windows-x86_64": {
      "relative_path": "cuda_cudart/windows-x86_64/cuda_cudart-windows-x86_64-11.4.108-archive.zip",
      "sha256": "b59756c27658d1ea87a17c06d064d1336576431cd64da5d1790d909e455d06d3",
      "md5": "7f6837a46b78198402429a3760ab28fc",
      "size": "2897751"
    }
  }
}
```

A JSON schema is provided at https://developer.download.nvidia.com/compute/redist/redistrib-v2.schema.json.

A sample script that parses these JSON manifests is available on GitHub:

- Downloads each archive
- Validates SHA256 checksums
- Extracts archives
- Flattens into a collapsed directory structure
### Table 7: Available Tarball and Zip Archives

<table>
<thead>
<tr>
<th>Product</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>CUDA Toolkit</td>
<td>./parse_redist.py --product cuda --label 12.3.2</td>
</tr>
<tr>
<td>cuBLASMp</td>
<td>./parse_redist.py --product cublasmp --label 0.1.0</td>
</tr>
<tr>
<td>cuDNN</td>
<td>./parse_redist.py --product cudnn --label 8.9.6.50</td>
</tr>
<tr>
<td>cuDSS</td>
<td>./parse_redist.py --product cudss --label 0.1.0</td>
</tr>
<tr>
<td>cuQuantum</td>
<td>./parse_redist.py --product cuquantum --label 23.10.0</td>
</tr>
<tr>
<td>cuSPARSELt</td>
<td>./parse_redist.py --product cusparselt --label 0.5.2</td>
</tr>
<tr>
<td>cuTENSOR</td>
<td>./parse_redist.py --product cutensor --label 1.7.0</td>
</tr>
<tr>
<td>NVIDIA driver</td>
<td>./parse_redist.py --product nvidia-driver --label 535.129.03</td>
</tr>
<tr>
<td>nvJPEG2000</td>
<td>./parse_redist.py --product nvjpeg2000 --label 0.7.5</td>
</tr>
<tr>
<td>NVPL</td>
<td>./parse_redist.py --product nvpl --label 23.11</td>
</tr>
<tr>
<td>nvTIFF</td>
<td>./parse_redist.py --product nvtiff --label 0.3.0</td>
</tr>
</tbody>
</table>

### 14.2. Importing Tarballs into CMake

The recommended module for importing these tarballs into the CMake build system is via FindCUDA-Toolkit (3.17 and newer).

**Note:** The FindCUDA module is deprecated.

The path to the extraction location can be specified with the CUDAToolkit_ROOT environmental variable. For example CMakeLists.txt and commands, see cmake/1_FindCUDAToolkit/.

For older versions of CMake, the ExternalProject_Add module is an alternative method. For example CMakeLists.txt file and commands, see cmake/2_ExternalProject/.

### 14.3. Importing Tarballs into Bazel

The recommended method of importing these tarballs into the Bazel build system is using http_archive and pkg_tar.

For an example, see bazel/1_pkg_tar/.
Chapter 15. CUDA Cross-Platform Environment

Cross development for arm64-sbsa is supported on Ubuntu 20.04, Ubuntu 22.04, RHEL 8, RHEL 9, and SLES 15.

Cross development for arm64-Jetson is only supported on Ubuntu 20.04

We recommend selecting a host development environment that matches the supported cross-target environment. This selection helps prevent possible host/target incompatibilities, such as GCC or GLIBC version mismatches.

15.1. CUDA Cross-Platform Installation

Some of the following steps may have already been performed as part of the native Ubuntu installation. Such steps can safely be skipped.

These steps should be performed on the x86_64 host system, rather than the target system. To install the native CUDA Toolkit on the target system, refer to the native Ubuntu installation section.

1. Perform the pre-installation actions.
2. Install repository meta-data package with:

   sudo dpkg -i cuda-repo-cross-<identifier>_all.deb

   where <identifier> indicates the operating system, architecture, and/or the version of the package.
3. Update the Apt repository cache:

   sudo apt-get update

4. Install the appropriate cross-platform CUDA Toolkit:
   a. For aarch64:

      sudo apt-get install cuda-cross-aarch64

   b. For QNX:

      sudo apt-get install cuda-cross-qnx

5. Perform the post-installation actions.
15.2. CUDA Cross-Platform Samples

CUDA Samples are now located in https://github.com/nvidia/cuda-samples, which includes instructions for obtaining, building, and running the samples.
Chapter 16. Post-installation Actions

The post-installation actions must be manually performed. These actions are split into mandatory, recommended, and optional sections.

16.1. Mandatory Actions

Some actions must be taken after the installation before the CUDA Toolkit and Driver can be used.

16.1.1. Environment Setup

The PATH variable needs to include

```
export PATH=/usr/local/cuda-12.4/bin:${PATH:+:${PATH}}. Nsight Compute has moved to /opt/nvidia/nsight-compute/ only in rpm/deb installation method. When using .run installer it is still located under /usr/local/cuda-12.4/.
```

To add this path to the PATH variable:

```
export PATH=/usr/local/cuda-12.4/bin:${PATH:+:${PATH}}
```

In addition, when using the runfile installation method, the LD_LIBRARY_PATH variable needs to contain /usr/local/cuda-12.4/lib64 on a 64-bit system, or /usr/local/cuda-12.4/lib on a 32-bit system

- To change the environment variables for 64-bit operating systems:
  ```
  export LD_LIBRARY_PATH=/usr/local/cuda-12.4/lib64
  ${LD_LIBRARY_PATH:+:${LD_LIBRARY_PATH}}
  ```

- To change the environment variables for 32-bit operating systems:
  ```
  export LD_LIBRARY_PATH=/usr/local/cuda-12.4/lib
  ${LD_LIBRARY_PATH:+:${LD_LIBRARY_PATH}}
  ```

Note that the above paths change when using a custom install path with the runfile installation method.
16.2. Recommended Actions

Other actions are recommended to verify the integrity of the installation.

16.2.1. Install Persistence Daemon

NVIDIA is providing a user-space daemon on Linux to support persistence of driver state across CUDA job runs. The daemon approach provides a more elegant and robust solution to this problem than persistence mode. For more details on the NVIDIA Persistence Daemon, see the documentation here.

The NVIDIA Persistence Daemon can be started as the root user by running:

```
∕usr∕bin∕nvidia-persistenced --verbose
```

This command should be run on boot. Consult your Linux distribution’s init documentation for details on how to automate this.

16.2.2. Install Writable Samples

CUDA Samples are now located in https://github.com/nvidia/cuda-samples, which includes instructions for obtaining, building, and running the samples.

16.2.3. Verify the Installation

Before continuing, it is important to verify that the CUDA toolkit can find and communicate correctly with the CUDA-capable hardware. To do this, you need to compile and run some of the sample programs, located in https://github.com/nvidia/cuda-samples.

Note: Ensure the PATH and, if using the runfile installation method, LD_LIBRARY_PATH variables are set correctly.

16.2.3.1 Verify the Driver Version

If you installed the driver, verify that the correct version of it is loaded. If you did not install the driver, or are using an operating system where the driver is not loaded via a kernel module, such as L4T, skip this step.

When the driver is loaded, the driver version can be found by executing the command

```
cat /proc/driver/nvidia/version
```

Note that this command will not work on an iGPU/dGPU system.
16.2.3.2 Running the Binaries

After compilation, find and run `deviceQuery` from https://github.com/nvidia/cuda-samples. If the CUDA software is installed and configured correctly, the output for `deviceQuery` should look similar to that shown in Figure 1.

![deviceQuery output](image)

**Figure 1:** Valid Results from deviceQuery CUDA Sample

The exact appearance and the output lines might be different on your system. The important outcomes are that a device was found (the first highlighted line), that the device matches the one on your system (the second highlighted line), and that the test passed (the final highlighted line).

If a CUDA-capable device and the CUDA Driver are installed but `deviceQuery` reports that no CUDA-capable devices are present, this likely means that the `/dev/nvidia*` files are missing or have the wrong permissions.

On systems where SELinux is enabled, you might need to temporarily disable this security feature to run `deviceQuery`. To do this, type:

```
setenforce 0
```

from the command line as the superuser.

Running the `bandwidthTest` program ensures that the system and the CUDA-capable device are able to communicate correctly. Its output is shown in Figure 2.

![bandwidthTest output](image)

**Figure 2:** Valid Results from bandwidthTest CUDA Sample

Note that the measurements for your CUDA-capable device description will vary from system to system. The important point is that you obtain measurements, and that the second-to-last line (in Figure 2) confirms that all necessary tests passed.

Should the tests not pass, make sure you have a CUDA-capable NVIDIA GPU on your system and make sure it is properly installed.
If you run into difficulties with the link step (such as libraries not being found), consult the Linux Release Notes found in https://github.com/nvidia/cuda-samples.

16.2.4. Install Nsight Eclipse Plugins

To install Nsight Eclipse plugins, an installation script is provided:

```
/usr/local/cuda-12.4/bin/nsight_ee_plugins_manage.sh install <eclipse-dir>
```

Refer to Nsight Eclipse Plugins Installation Guide for more details.

16.2.5. Local Repo Removal

Removal of the local repo installer is recommended after installation of CUDA SDK.

**Ubuntu and Debian**

```
sudo apt-get remove --purge "cuda-repo-<distro>-X-Y-local*"
```

**Fedora**

```
sudo dnf remove "cuda-repo-<distro>-X-Y-local*"
```

**RHEL 9 / Rocky Linux 9 and RHEL 8 / Rocky Linux 8**

```
sudo dnf remove "cuda-repo-<distro>-X-Y-local*"
```

**openSUSE 15 and SLES 15**
sudo zypper remove "cuda-repo-<distro>-X-Y-local*"

Removal of the local repo installer is recommended after installation of NVIDIA driver.

Ubuntu and Debian

sudo apt-get remove --purge "nvidia-driver-local-repo-<distro>**"

Fedora

sudo dnf remove "nvidia-driver-local-repo-<distro>**"

RHEL 9 / Rocky Linux 9 and RHEL 8 / Rocky Linux 8

sudo dnf remove "nvidia-driver-local-repo-<distro>**"

openSUSE 15 and SLES 15

sudo zypper remove "nvidia-driver-local-repo-<distro>**"

16.3. Optional Actions

Other options are not necessary to use the CUDA Toolkit, but are available to provide additional features.

16.3.1. Install Third-party Libraries

Some CUDA samples use third-party libraries which may not be installed by default on your system. These samples attempt to detect any required libraries when building.

If a library is not detected, it waives itself and warns you which library is missing. To build and run these samples, you must install the missing libraries. In cases where these dependencies are not installed, follow the instructions below.

RHEL 8 / Rocky Linux 8

sudo dnf install freeglut-devel libXI-devel libXmu-devel make mesa-libGLU-devel freeimage-devel libglfw3-devel

RHEL 9 / Rocky Linux 9

sudo dnf install freeglut-devel libXI-devel libXmu-devel make mesa-libGLU-devel freeimage-devel libglfw3-devel

KylinOS 10

sudo dnf install freeglut-devel libXI-devel libXmu-devel make mesa-libGLU-devel freeimage-devel libglfw3-devel

Fedora
16.3.2. Install the Source Code for cuda-gdb

The cuda-gdb source must be explicitly selected for installation with the runfile installation method. During the installation, in the component selection page, expand the component “CUDA Tools 12.4” and select cuda-gdb-src for installation. It is unchecked by default.

To obtain a copy of the source code for cuda-gdb using the RPM and Debian installation methods, the cuda-gdb-src package must be installed.

The source code is installed as a tarball in the /usr/local/cuda-12.4/extras directory.

16.3.3. Select the Active Version of CUDA

For applications that rely on the symlinks /usr/local/cuda and /usr/local/cuda-MAJOR, you may wish to change to a different installed version of CUDA using the provided alternatives.

To show the active version of CUDA and all available versions:

```
update-alternatives --display cuda
```

To show the active minor version of a given major CUDA release:

```
update-alternatives --display cuda-12
```

To update the active version of CUDA:

```
sudo update-alternatives --config cuda
```
Chapter 17. Advanced Setup

Below is information on some advanced setup scenarios which are not covered in the basic instructions above.
Table 8: Advanced Setup Scenarios when Installing CUDA

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Instructions</th>
</tr>
</thead>
</table>
| Install CUDA using the Package Manager installation method without installing the NVIDIA GL libraries. | **Fedora**  
Install CUDA using the following command:  
sudo dnf install cuda-toolkit-12-4 \  
nvidia-driver-cuda akmmod-nvidia  
Follow the instructions [here](#) to ensure that Nouveau is disabled.  
If performing an upgrade over a previous installation, the NVIDIA kernel module may need to be rebuilt by following the instructions [here](#).  
**OpenSUSE/SLES**  
On some system configurations the NVIDIA GL libraries may need to be locked before installation using:  
sudo zypper addlock nvidia-g104  
Install CUDA using the following command:  
sudo zypper install --no-recommends cuda-toolkit-12-4 \  
nvidia-computeG04 \  
nvidia-gfxG04-kmp-default  
Follow the instructions [here](#) to ensure that Nouveau is disabled.  
**Ubuntu**  
This functionality isn't supported on Ubuntu. Instead, the driver packages integrate with the Bumblebee framework to provide a solution for users who wish to control what applications the NVIDIA drivers are used for. See Ubuntu’s [Bumblebee wiki](#) for more information. |
| Upgrade from a RPM/Deb driver installation which includes the diagnostic driver packages to a driver installation which does not include the diagnostic driver packages. | **RHEL/CentOS**  
Remove diagnostic packages using the following command:  
sudo yum remove cuda-drivers-diagnostic \  
xorg-x11-drv-nvidia-diagnostic  
Follow the instructions [here](#) to continue installation as normal.  
**Fedora**  
Remove diagnostic packages using the following command:  
sudo dnf remove cuda-drivers-diagnostic \  
xorg-x11-drnvidia-diagnostic  
Follow the instructions [here](#) to continue installation as normal.  
**OpenSUSE/SLES**  
Remove diagnostic packages using the following command:  
sudo zypper remove cuda-drivers-diagnostic \  
nvidia-diagnosticG04  
Follow the instructions [here](#) to continue installation as normal.  
**Ubuntu**  
Remove diagnostic packages using the following command: |
Chapter 18. Frequently Asked Questions

18.1. How do I install the Toolkit in a different location?

The Runfile installation asks where you wish to install the Toolkit during an interactive install. If installing using a non-interactive install, you can use the --toolkitpath parameter to change the install location:

```
./runfile.run --silent \n   --toolkit --toolkitpath=/my/new/toolkit
```

The RPM and Deb packages cannot be installed to a custom install location directly using the package managers. See the “Install CUDA to a specific directory using the Package Manager installation method” scenario in the Advanced Setup section for more information.

18.2. Why do I see “nvcc: No such file or directory” when I try to build a CUDA application?

Your PATH environment variable is not set up correctly. Ensure that your PATH includes the bin directory where you installed the Toolkit, usually /usr/local/cuda-12.4/bin.

```
export PATH=/usr/local/cuda-12.4/bin$PATH:+$PATH
```
18.3. Why do I see “error while loading shared libraries: <lib name>: cannot open shared object file: No such file or directory” when I try to run a CUDA application that uses a CUDA library?

Your LD_LIBRARY_PATH environment variable is not set up correctly. Ensure that your LD_LIBRARY_PATH includes the lib and/or lib64 directory where you installed the Toolkit, usually /usr/local/cuda-12.4/lib{,64}:

```bash
export LD_LIBRARY_PATH=/usr/local/cuda-12.4/lib\${LD_LIBRARY_PATH:+:${LD_LIBRARY_PATH}}
```

18.4. Why do I see multiple “404 Not Found” errors when updating my repository meta-data on Ubuntu?

These errors occur after adding a foreign architecture because apt is attempting to query for each architecture within each repository listed in the system’s sources.list file. Repositories that do not host packages for the newly added architecture will present this error. While noisy, the error itself does no harm. Please see the Advanced Setup section for details on how to modify your sources.list file to prevent these errors.

18.5. How can I tell X to ignore a GPU for compute-only use?

To make sure X doesn’t use a certain GPU for display, you need to specify which other GPU to use for display. For more information, please refer to the “Use a specific GPU for rendering the display” scenario in the Advanced Setup section.
18.6. Why doesn’t the cuda-repo package install the CUDA Toolkit and Drivers?

When using RPM or Deb, the downloaded package is a repository package. Such a package only informs the package manager where to find the actual installation packages, but will not install them. See the Package Manager Installation section for more details.

18.7. How do I get CUDA to work on a laptop with an iGPU and a dGPU running Ubuntu14.04?

After installing CUDA, set the driver value for the intel device in /etc/X11/xorg.conf to 'modesetting' as shown below:

```plaintext
Section "Device"
  Identifier "intel"
  Driver "modesetting"
... 
EndSection
```

To prevent Ubuntu from reverting the change in xorg.conf, edit /etc/default/grub to add "nogpumanager" to GRUB_CMDLINE_LINUX_DEFAULT.

Run the following command to update grub before rebooting:

```
sudo update-grub
```

18.8. What do I do if the display does not load, or CUDA does not work, after performing a system update?

System updates may include an updated Linux kernel. In many cases, a new Linux kernel will be installed without properly updating the required Linux kernel headers and development packages. To ensure the CUDA driver continues to work when performing a system update, rerun the commands in the Kernel Headers and Development Packages section.

Additionally, on Fedora, the Akmods framework will sometimes fail to correctly rebuild the NVIDIA kernel module packages when a new Linux kernel is installed. When this happens, it is usually sufficient to invoke Akmods manually and regenerate the module mapping files by running the following commands in a virtual console, and then rebooting:

```
sudo akmods --force
sudo depmod
```
You can reach a virtual console by hitting `ctrl+alt+f2` at the same time.

18.9. How do I install a CUDA driver with a version less than 367 using a network repo?

To install a CUDA driver at a version earlier than 367 using a network repo, the required packages will need to be explicitly installed at the desired version. For example, to install 352.99, instead of installing the cuda-drivers metapackage at version 352.99, you will need to install all required packages of cuda-drivers at version 352.99.

18.10. How do I install an older CUDA version using a network repo?

Depending on your system configuration, you may not be able to install old versions of CUDA using the cuda metapackage. In order to install a specific version of CUDA, you may need to specify all of the packages that would normally be installed by the cuda metapackage at the version you want to install.

If you are using yum to install certain packages at an older version, the dependencies may not resolve as expected. In this case you may need to pass `--setopt=obsoletes=0` to yum to allow an install of packages which are obsoleted at a later version than you are trying to install.

18.11. Why does the installation on SUSE install the Mesa-dri-nouveau dependency?

This dependency comes from the SUSE repositories and shouldn't affect the use of the NVIDIA driver or the CUDA Toolkit. To disable this dependency, you can lock that package with the following command:

```
sudo zypper al Mesa-dri-nouveau
```
18.12. How do I handle “Errors were encountered while processing: glx-diversions”?

This sometimes occurs when trying to uninstall CUDA after a clean .deb installation. Run the following commands:

```bash
sudo apt-get install glx-diversions --reinstall
sudo apt-get remove nvidia-alternative
```

Then re-run the commands from Removing CUDA Toolkit and Driver.
Chapter 19. Additional Considerations

Now that you have CUDA-capable hardware and the NVIDIA CUDA Toolkit installed, you can examine and enjoy the numerous included programs. To begin using CUDA to accelerate the performance of your own applications, consult the CUDA C++ Programming Guide, located in /usr/local/cuda-12.4/doc.

A number of helpful development tools are included in the CUDA Toolkit to assist you as you develop your CUDA programs, such as NVIDIA® Nsight™ Eclipse Edition, NVIDIA Visual Profiler, CUDA-GDB, and CUDA-MEMCHECK.

Chapter 20. Switching between Driver Module Flavors

Use the following steps to switch between the NVIDIA driver legacy and open module flavors on your system.

**Note:** If switching to open module, experimental support for GeForce and Quadro SKUs can be enabled with:

```bash
echo "options nvidia NVreg_OpenRmEnableUnsupportedGpus=1" | sudo tee /etc/modprobe.d/nvidia-gsp.conf
```

**Note:** Replace XXX with the NVIDIA driver branch number such as 550.

**Fedora, RHEL 9 / Rocky Linux 9, RHEL 8 / Rocky Linux 8**
To switch between legacy and open: uninstall, then reinstall.

**Kylin OS**
To switch between legacy and open: uninstall, then reinstall.

**Ubuntu**
To switch from legacy to open:

```bash
sudo apt-get --purge remove nvidia-kernel-source-XXX
sudo apt-get install --verbose-versions nvidia-kernel-open-XXX
sudo apt-get install --verbose-versions cuda-drivers-XXX
```

To switch from open to legacy:

```bash
sudo apt-get remove --purge nvidia-kernel-open-XXX
sudo apt-get install --verbose-versions cuda-drivers-XXX
```

**Debian**
To switch from legacy to open:

```bash
sudo apt-get --purge remove nvidia-kernel-dkms
sudo apt-get install --verbose-versions nvidia-kernel-open-dkms
sudo apt-get install --verbose-versions cuda-drivers-XXX
```

To switch from open to legacy:
sudo apt-get remove --purge nvidia-kernel-open-dkms
sudo apt-get install --verbose-versions cuda-drivers-XXX

**OpenSUSE**

To switch from legacy to open:

```bash
sudo zypper remove nvidia-driver-G06-kmp-default
sudo zypper install --details nvidia-open-driver-G06-kmp-default
sudo zypper install --details cuda-drivers-XXX
```

To switch from open to legacy:

```bash
sudo zypper remove nvidia-open-driver-G06-kmp-default
sudo zypper install --details cuda-drivers-XXX
```

**SLES**

To switch from legacy to open:

```bash
sudo zypper remove nvidia-driver-G06-kmp-default nvidia-driver-G06-kmp-azure
sudo zypper install --details nvidia-open-driver-G06-kmp-default nvidia-open-driver-G06-kmp-azure
sudo zypper install --details cuda-drivers-XXX
```

To switch from open to legacy:

```bash
sudo zypper remove nvidia-open-driver-G06-kmp-default nvidia-driver-G06-open-kmp-azure
sudo zypper install --details cuda-drivers-XXX
```

**Note:** The Azure package is only available for SLES (x86_64).
Chapter 21. Removing CUDA Toolkit and Driver

Follow the below steps to properly uninstall the CUDA Toolkit and NVIDIA Drivers from your system. These steps will ensure that the uninstallation will be clean.

**KylinOS 10**

To remove CUDA Toolkit:

```
sudo dnf remove "cuda* "*cublas*" "*cufft*" "*cufile* "*curand*" "
"*cusolver*" "*cusparse*" "*gds-tools*" "*npp*" "*nvjpeg*"

```

To remove NVIDIA Drivers:

```
sudo dnf module remove --all nvidia-driver
```

To reset the module stream:

```
sudo dnf module reset nvidia-driver
```

**RHEL 9 / Rocky Linux 9**

To remove CUDA Toolkit:

```
sudo dnf remove "cuda* "*cublas*" "*cufft*" "*cufile* "*curand*" "
"*cusolver*" "*cusparse*" "*gds-tools*" "*npp*" "*nvjpeg*" "nsight*" "*nvvm*"
```

To remove NVIDIA Drivers:

```
sudo dnf module remove --all nvidia-driver
```

To reset the module stream:

```
sudo dnf module reset nvidia-driver
```

**RHEL 8 / Rocky Linux 8**

To remove CUDA Toolkit:

```
sudo dnf remove "cuda* "*cublas*" "*cufft*" "*cufile* "*curand*" "
"*cusolver*" "*cusparse*" "*gds-tools*" "*npp*" "*nvjpeg*" "nsight*" "*nvvm*"
```

To remove NVIDIA Drivers:
sudo dnf module remove --all nvidia-driver

To reset the module stream:

sudo dnf module reset nvidia-driver

**Fedora**

To remove CUDA Toolkit:

```bash
sudo dnf remove "cuda*" "*cublas*" "*cufft*" "*cufile*" "*curand*" \ "*cusolver*" "*cusparse*" "*gds-tools*" "*npp*" "*nvjpeg*" "*nsight*" "*nvvm*"
```

To remove NVIDIA Drivers:

sudo dnf module remove --all nvidia-driver

To reset the module stream:

sudo dnf module reset nvidia-driver

To remove 3rd party NVIDIA Drivers:

sudo dnf remove "*nvidia*"

**OpenSUSE / SLES**

To remove CUDA Toolkit:

```bash
sudo zypper remove "cuda*" "*cublas*" "*cufft*" "*cufile*" "*curand*" \ "*cusolver*" "*cusparse*" "*gds-tools*" "*npp*" "*nvjpeg*" "*nsight*" "*nvvm*"
```

To remove NVIDIA Drivers:

sudo zypper remove "*nvidia*"

**Ubuntu and Debian**

To remove CUDA Toolkit:

```bash
sudo apt-get --purge remove "cuda*" "*cublas*" "*cufft*" "*cufile*" "*curand*" \ "*cusolver*" "*cusparse*" "*gds-tools*" "*npp*" "*nvjpeg*" "*nsight*" "*nvvm*"
```

To remove NVIDIA Drivers:

sudo apt-get --purge remove "*nvidia*" "libxnvctrl*"

To clean up the uninstall:

sudo apt-get autoremove
Chapter 22. Notices

22.1. 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.

22.2. OpenCL

OpenCL is a trademark of Apple Inc. used under license to the Khronos Group Inc.

22.3. Trademarks

NVIDIA and the NVIDIA logo are trademarks or registered trademarks of NVIDIA Corporation in the U.S. and other countries. Other company and product names may be trademarks of the respective companies with which they are associated.
Chapter 23. Copyright

© 2009-2024 NVIDIA Corporation & affiliates. All rights reserved.
This product includes software developed by the Syncro Soft SRL (http://www.sync.ro/).

Copyright

©2009-2024, NVIDIA Corporation & affiliates. All rights reserved