



NVIDIA CONTAINER RUNTIME FOR DOCKER

DA-09013-001 _v03 | July 2019

Upgrade Instructions for DGX Systems



TABLE OF CONTENTS

Chapter 1. About the NVIDIA Container Runtime for Docker.....	1
Chapter 2. What Happens During the Upgrade Process?.....	3
Chapter 3. Performing the Upgrade.....	4
Chapter 4. Using the NVIDIA Container Runtime for Docker.....	6

Chapter 1.

ABOUT THE NVIDIA CONTAINER RUNTIME FOR DOCKER

The NVIDIA Container Runtime for Docker is an improved mechanism for allowing the Docker Engine to support NVIDIA GPUs used by GPU-accelerated containers. This new runtime replaces the Docker Engine Utility for NVIDIA GPUs.

You now have the option to upgrade your DGX system environment to use the NVIDIA Container Runtime for Docker.



Important Some applications, such as Kubernetes, as well as CUDA 10 NGC containers, require the new runtime.

The Docker Engine Utility for NVIDIA GPUs is implemented with the installation of the **nvidia-docker** package. To upgrade your DGX system environment to use the NVIDIA Container Runtime for Docker, you must install the **nvidia-docker2** package.¹

This document provides instructions for performing the upgrade.

Prerequisites

The instructions in this document require that your DGX system is installed with DGX OS Software Release 3.1, version 3.1.6 or higher. To determine the DGX OS Software version installed on your system, run the following command.

```
$ grep VERSION /etc/dgx-release
```

This will show the version installed on the system from the ISO image (**DGX_SWBUILD_VERSION**) as well as any subsequent updates performed on the system (**DGX_OTA_VERSION**). The latest version listed is the currently installed version.

Example for DGX-1:

```
DGX_NAME="DGX Server"
```

¹ For this reason, the NVIDIA Container Runtime for Docker is also referred to by the updated package "**nvidia-docker2**".

```
DGX_PRETTY_NAME="NVIDIA DGX Server"
DGX_SWBUILD_DATE="2017-09-02"
DGX_SWBUILD_VERSION="3.1.2"
...
DGX_OTA_VERSION="3.1.6"
DGX_OTA_DATE="Wed May 9 13:47:33 PDT 2018"
```

In this example, the installed version is 3.1.6.



DGX systems installed with DGX OS release 4 automatically include the NVIDIA Container Runtime for Docker.

If you need to update your DGX OS Software, refer to the [Release Note and Upgrade Instructions](#) for DGX-1, or the [DGX Station User Guide](#) for DGX Station.

Chapter 2.

WHAT HAPPENS DURING THE UPGRADE PROCESS?

Only GPU-accelerated containers are affected during the upgrade process. Other containers remain untouched. Specifically, the upgrade process does the following.

- ▶ Stops all running GPU-accelerated containers that were started using `nvidia-docker run`.
- ▶ Commits all GPU-accelerated containers on the system.

Images are saved in a file with the following naming format:

image_name-container_id

For example, a container started with image `nvcr.io/nvidia/caffe:18.05-py2` and with container ID `a4093a4040de` would be saved as `nvcr.io/nvidia/caffe:18.05-py2-a4093a4040de`.

- ▶ Removes the containers after committing them.



Important If you have edited the `/etc/docker/daemon.json` file, your edits may be overridden by the upgrade. See the section [Performing the Upgrade](#) for details on managing the changes.

Chapter 3.

PERFORMING THE UPGRADE

Upgrading to the new runtime involves updating the **nvidia-docker** package and then installing the **nvidia-docker2** package.

The instructions apply to DGX systems installed with the Docker Engine Utility for NVIDIA GPUs. To determine your installation, run the following command.

```
$ nvidia-docker version
```

If the command returns 2.0.x, then your system already contains the upgrade to the NVIDIA Container Runtime for Docker and no further action is needed.

1. Stop, commit, and then remove all GPU-accelerated containers that you want to keep.

While the upgrade process commits all your GPU-accelerated containers, you should commit the containers yourself as a fail-safe. Removing the containers afterwards will avoid duplicate copies.

2. Update the list of available packages and their versions.

```
$ sudo apt update
```

3. Verify the **nvidia-docker** package is updated to at least version 1.0.1-3.

```
$ dpkg -l nvidia-docker
```

If the version is lower than 1.0.1-3, then update the **nvidia-docker** package as follows.

```
$ sudo apt install nvidia-docker
```

The updated package ensures the upgrade to the NVIDIA Container Runtime for Docker is performed cleanly and reliably.

4. Install the **nvidia-docker2** package.

```
$ sudo apt install nvidia-docker2
```

You may be presented with a configuration choice such as the following:

```
Configuration file '/etc/docker/daemon.json'
```

```

==> File on system created by you or by a script.
==> File also in package provided by package maintainer.
What would you like to do about it ? Your options are:
Y or I : install the package maintainer's version
N or O : keep your currently-installed version
D : show the differences between the versions
Z : start a shell to examine the situation
The default action is to keep your current version.
*** daemon.json (Y/I/N/O/D/Z) [default=N] ?

```

Press **Y** at the prompt to install the **nvidia-docker2** configuration file. Your original file is renamed **/etc/docker/daemon.json.dpkg-old**. You can merge the contents afterwards to continue using your configuration changes.

5. Restart Docker.

```
$ sudo systemctl restart docker
```

6. (Optional) If the system has network access to the public Docker repository, verify the upgrade was successful.

```
$ sudo docker run --runtime=nvidia --rm nvcr.io/nvidia/cuda:9.0-devel-ubuntu16.04 nvidia-smi
```

Chapter 4.

USING THE NVIDIA CONTAINER RUNTIME FOR DOCKER



These instructions assume the user is part of the `docker` group. For details, see the relevant sections in the [DGX-1 User Guide](#) or in the [DGX-Station User Guide](#).

After updating to the NVIDIA Container Runtime for Docker, you can run GPU-accelerated containers in one of the following ways.

- ▶ Use **docker run** and specify **runtime=nvidia**.

```
$ docker run --runtime=nvidia ...
```

- ▶ Use **nvidia-docker run**.

```
$ nvidia-docker run ...
```

The new package provides backward compatibility, so you can still run GPU-accelerated containers by using this command, and the new runtime will be used.

- ▶ Use **docker run** with **nvidia** as the default runtime.

You can set **nvidia** as the default runtime, for example, by adding the following line to the `/etc/docker/daemon.json` configuration file as the first entry.

```
"default-runtime": "nvidia",
```

The following is an example of how the added line appears in the JSON file. Do not remove any pre-existing content when making this change.

```
{
  "default-runtime": "nvidia",
  "runtimes": {
    "nvidia": {
      "path": "/usr/bin/nvidia-container-runtime",
      "runtimeArgs": []
    }
  },
}
```

You can then use **docker run** to run GPU-accelerated containers.


```
$ docker run ...
```



Caution If you build Docker images while `nvidia` is set as the default runtime, make sure the build scripts executed by the Dockerfile specify the GPU architectures that the container will need. Failure to do so may result in the container being optimized only for the GPU architecture on which it was built. Instructions for specifying the GPU architecture depend on the application and are beyond the scope of this document. Consult the specific application build process for guidance.

Notice

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

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

NVIDIA makes no representation or warranty that the product described in this guide will be suitable for any specified use without further testing or modification. Testing of all parameters of each product is not necessarily performed by NVIDIA. It is customer's sole responsibility to ensure the product is suitable and fit for the application planned by customer and to do the necessary testing for the application in order to avoid a default of the application or the product. Weaknesses in customer's product designs may affect the quality and reliability of the NVIDIA product and may result in additional or different conditions and/or requirements beyond those contained in this guide. NVIDIA does not accept any liability related to any default, damage, costs or problem which may be based on or attributable to: (i) the use of the NVIDIA product in any manner that is contrary to this guide, or (ii) customer product designs.

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

Trademarks

NVIDIA, the NVIDIA logo, DGX, DGX-1, DGX-2, and DGX Station are trademarks and/or registered trademarks of NVIDIA Corporation in the United States and other countries. Other company and product names may be trademarks of the respective companies with which they are associated.

Copyright

© 2019 NVIDIA Corporation. All rights reserved.