TABLE OF CONTENTS

Chapter 1. Introduction to NVIDIA GPU Cloud............................................................. 1
  1.1. NGC Containers........................................................................................... 1
  1.2. NGC Container Registry..................................................................................2
  1.3. NGC Website...............................................................................................2
  1.4. Optimized Accelerated Computing Environments..............................................2
  1.5. Running Deep Learning Frameworks Using NGC.............................................2
Chapter 1.  
INTRODUCTION TO NVIDIA GPU CLOUD

NVIDIA GPU Cloud (NGC) is a GPU-accelerated cloud platform optimized for deep learning and scientific computing. In this release, NGC includes NGC containers, the NGC container registry, the NGC website, and platform software for running the deep learning containers. This document provides an overview of the NVIDIA GPU Cloud and how to use it.

1.1. NGC Containers

NGC Containers are designed to enable a software platform centered around minimal OS requirements, Docker and driver installation on the server or workstation, and provisioning of all application and SDK software in the NGC containers through the NGC container registry.

NGC manages a catalog of fully integrated and optimized deep learning framework containers that take full advantage of NVIDIA GPUs in both single GPU and multi-GPU configurations. They include CUDA Toolkit, DIGITS workflow, and the following deep learning frameworks: NVCaffe, Caffe2, Microsoft Cognitive Toolkit (CNTK), MXNet, PyTorch, TensorFlow, Theano, and Torch. These framework containers are delivered ready-to-run, including all necessary dependencies such as CUDA runtime, NVIDIA libraries, and an operating system.

Each framework container image also includes the framework source code to enable custom modifications and enhancements, along with the complete software development stack.

NVIDIA updates these deep learning containers monthly to ensure they continue to provide peak performance.

In addition NGC provides a catalog of HPC visualization containers, now available in beta, featuring the industry’s leading visualization tools, including ParaView with NVIDIA IndeX volume renderer, NVIDIA OptIX ray-tracing library and NVIDIA Holodeck for interactive real-time visualization and high-quality visuals.
NGC also hosts popular third-party GPU ready HPC application containers which conform to NGC container standards and best practices, making it easy to get the latest GPU optimized HPC software up and running quickly.

1.2. NGC Container Registry

The NGC container registry stores the container images for distribution at nvcr.io. With an NGC API Key, you can pull and run NGC containers from the registry.

1.3. NGC Website

The NGC website (https://ngc.nvidia.com) is the portal for managing NGC. Easily view the contents of the NGC container registry, create your API Key that authorizes you to use the containers, and see which cloud service providers offer virtual machine instances that are optimized for NGC containers.

1.4. Optimized Accelerated Computing Environments

All NGC containers are qualified to take full advantage of NVIDIA GPUs and are ready-to-run on supported platforms such as NVIDIA DGX systems and supported cloud service providers.

1.5. Running Deep Learning Frameworks Using NGC

The process for running deep learning framework containers can be summarized as follows:

**Preparing the Platform**

Prepare the platform for running the NGC containers. See the following documentation for instructions:

- NVIDIA GPU Cloud (NGC) (supported cloud service providers documentation)
- Preparing to Use Containers (for DGX systems)
- NGC Container User Guide (for other systems)
Obtaining NGC Credentials and Selecting a Container

Sign up for an NGC account at https://ngc.nvidia.com, log in, then create your NGC API Key which is needed to use the containers in the NGC container registry.

Browse the Registry section of the NGC website to identify a container and tag to use.

See Getting Started with NGC for details.

Running Containers

Connect to the platform log into nvcr.io, then enter commands to run the containers that you identified from the Registry.

See the NVIDIA Docker Containers for Deep Learning Frameworks User Guide for more information on running deep learning containers.

See the NGC Container User Guide for more information on running other NGC containers.
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, and Volta are trademarks and/or registered trademarks of NVIDIA Corporation in the Unites States and other countries.

Docker and the Docker logo are trademarks or registered trademarks of Docker, Inc. in the United States and/or other countries.

Other company and product names may be trademarks of the respective companies with which they are associated.

Copyright

© 2018 NVIDIA Corporation. All rights reserved.

www.nvidia.com