NGC Overview

Application Note
Table of Contents

Chapter 1. What is NVIDIA NGC ................................................................. 1
  1.1. NGC Catalog .................................................................................... 1
  1.2. NGC-Ready Program ....................................................................... 2
Chapter 2. Why NGC Software ................................................................. 3
Chapter 3. Accessing NGC Software ........................................................ 4
Chapter 4. Registering and Activating Your NGC Account ..................... 5
  4.1. Signing Up for an NGC Registered Account .................................... 5
  4.2. Generating Your NGC API Key ....................................................... 8
Chapter 5. Obtaining a Private Registry .................................................. 10
Chapter 6. NGC Support Services ............................................................ 13
Chapter 7. Introduction to the NGC Catalog and Registry CLIs ............... 15
  7.1. Installing NGC Catalog CLI ........................................................... 15
  7.2. Installing NGC Registry CLI ......................................................... 16
  7.3. Getting Help Using NGC CLI ......................................................... 16
  7.4. Running the Diagnostics ............................................................... 16
Chapter 1. What is NVIDIA NGC

An overview of what constitutes NVIDIA NGC.

1.1. NGC Catalog

The NGC Catalog is a curated set of GPU-optimized software for AI, HPC and Visualization.

The content provided by NVIDIA and third party ISVs simplify building, customizing and the integration of GPU-optimized software into workflows, accelerating the time to solutions for users.

The NGC Catalog consists of containers, pre-trained models, Helm charts for Kubernetes deployments and industry specific AI toolkits with software development kits (SDKs).

Containers

Containers package software applications, libraries, dependencies, and run time compilers in a self-contained environment so they can be easily deployed across various compute environments. They enable software portability and through a single command, users can pull, run and scale applications across the cloud, the datacenter and the edge. The NGC Catalog hosts a broad range of containers including deep learning frameworks, machine learning, HPC and visualization applications that maximize the utilization of GPU environments.

Models and Resources

The NGC Catalog offers pre-trained models for a wide range of common AI tasks that are optimized for NVIDIA Tensor Core GPUs. The pre-trained models can be used for inference or fine tuned with transfer learning, saving data scientists and developers valuable time. Resources provide reference neural network architectures across all domains and popular frameworks with the state-of-the-art accuracy to enable reproducibility as well as documentation and code samples which make it easy to get started with deep learning.

Helm Charts

Kubernetes is a container orchestrator that facilitates the deployment and management of containerized applications and microservices. A Helm chart is a package manager that allows DevOps to more easily configure, deploy and update applications across Kubernetes
environments. The NGC Catalog provides Helm charts for the deployment of GPU-optimized applications and SDKs.

Software Development Kits

SDKs deliver all the tooling users need to build and deploy AI applications across domains such as medical imaging, conversational AI or video analytics. They include annotation tools for data labeling, pre-trained models for customization with transfer learning and SDKs that enable deployment across the cloud, the data center, or the edge for low-latency inference.

Users can learn more about the NGC Catalog content by visiting NGC directly.

1.2. NGC-Ready Program

An overview of the NGC-Ready program.

Another feature of NGC is the NGC-Ready program which validates the performance of AI, ML and DL workloads using NVIDIA GPUs on leading servers and public clouds. It allows server manufacturers and public clouds to qualify their NVIDIA GPU equipped systems on a wide variety of AI workloads ranging from training to inference on on-premise servers, cloud infrastructure and edge devices.
Chapter 2. Why NGC Software

NGC provides software to meet the needs of data scientists, developers, and researchers with various levels of AI expertise.

Software hosted on NGC undergoes scans against an aggregated set of common vulnerabilities and exposures (CVEs), crypto and private keys.

In addition to security scanning, NGC software is tested against a broad range of GPU-enabled platforms including public cloud instances, workstations and OEM servers targeted for the datacenter or edge deployments. Supported GPUs include V100, A100, T4, Jetson and the RTX Quadro.

NGC software is tested and assured to scale to multiple GPUs and in some cases, to scale to multi-node ensuring users maximize the use of their gpu-powered servers out of the box.

For a select set of containers, NVIDIA offers NGC Support Services for running software on DGX platforms or certified OEM servers. The service gives enterprise IT direct access to NVIDIA subject matter experts to quickly address software issues and minimize system downtime.
Chapter 3. Accessing NGC Software

How to access NGC software.
There are three ways to access software on NGC.

Guest Access: Content under guest access does not require users to register or sign-in to pull or download the software.

The majority of software on the NGC Catalog falls under guest access. However, it is important to note that many third party applications require a license key which can be sourced directly from ISVs. So while pulling the container from the NGC Catalog does not require sign-in, a user might have to reach out to an ISV to obtain a license key for further use.

Authenticated Access: Software under authenticated access requires a user to create an NVIDIA NGC account or sign-in to the NGC Catalog. Pulling or downloading the software requires the use of their API key.

Approved Access: To access the software under this category, the publisher must grant user approval. Instructions to request access are provided in the overview section of the respective software. Once access is granted, the user will receive a notification with further instructions for accessing the software.
Chapter 4. Registering and Activating Your NGC Account

Instructions for registering and activating your NGC account.

4.1. Signing Up for an NGC Registered Account

This section describes the process of signing up for an NGC account.

1. From your browser, go to the NGC Sign In page and then click Continue by the NVIDIA Account choice to create a new NVIDIA account.

2. At the Log In screen, click Create Account.
3. At the Create an Account screen, enter your display name, email address, and new password, then agree to the NVIDIA Account Terms of Use and click Create Account.

A verification email is sent to your email address.
4. Open the email and click **Verify Email Address**.
5. At the **Almost done!** dialog, select the desired information and then click **Submit**.

6. Agree to the NVIDIA Account Terms of Use and select desired options and then click **Continue**.

7. The first time that you sign in, you must agree to the NVIDIA GPU Cloud Terms of Use before you can enter the website.
Accept the Terms of Use to enter the NGC web UI under your new account.

4.2. Generating Your NGC API Key

This section describes how to obtain an API key to access locked container images from the NGC Registry.

1. Sign in to the NGC website.
   From a browser, go to https://ngc.nvidia.com/signin/email and then enter your email and password.
2. In the top right corner, click your user account icon and select Setup.
3. Click Get API key to open the Setup > API Key page.
The API Key is the mechanism used to authenticate your access to the NGC container registry.

4. Click **Generate API Key** to generate your API key.
   
   A warning message appears to let you know that your old API key will become invalid if you create a new key.

5. Click **Confirm** to generate the key.

   Your API key appears.

You only need to generate an API Key once. NGC does not save your key, so store it in a secure place.

**Tip:** You can copy your API Key to the clipboard by clicking the copy icon to the right of the API key.

Should you lose your API Key, you can generate a new one from the NGC website. When you generate a new API Key, the old one is invalidated.
Chapter 5. Obtaining a Private Registry

This chapter provides instructions for DGX customers on obtaining a private registry.

After purchasing a support entitlement with NVIDIA, the end-customer will receive an NVIDIA Entitlement Certificate via email. The email will include all the pertinent instructions needed to register for technical support.

The following is an example of the NVIDIA Entitlement Certificate email.

The Entitlement Certificate itself is provided as a PDF attachment. The following is an example of an NVIDIA Entitlement Certificate.
The PDF also includes instructions for using the certificate.
If you already have an account, you can log into the NVIDIA Enterprise Support portal right away.

If you are a new user without an NGC Support account, click the NVIDIA Enterprise Support Registration Form link. This link will have embedded information regarding your account and so, it is very IMPORTANT NOT to share this entitlement link outside of your organization.

Registration will provide an NGC private registry account in addition to the NVIDIA Enterprise Support account. You’ll receive a welcome email at which time you can activate (create) your NGC private registry account.
Chapter 6. NGC Support Services

Overview of NGC support services.

NGC Support Services provides enterprise-grade support to ensure optimal performance and maximized system utility for NGC-Ready systems. The support service is purchased through NVIDIA’s OEM partners. The service gives enterprise customers direct phone access to NVIDIA’s subject matter experts to quickly address L1 through L3 software issues during local business hours or via web and email. This support for open source AI software combined with the OEM’s hardware support gives enterprises to build optimized AI solutions at scale, minimize system downtime and accelerate their time to solution.

NGC support services cover machine learning, deep learning and inference containers available from NGC, running on both bare metal or virtual machines. Support is also provided for CUDA and associated drivers. Below is the breakdown of the various softwares that are covered under the contract. The OEM partner is typically responsible for providing support at the hardware and OS level with additional support for container orchestration tools.
Figure 1. Components of the Stack Covered Under NGC Support Services
Chapter 7. Introduction to the NGC Catalog and Registry CLIs

Introduction to NGC CLIs

The NGC CLIs are command-line interfaces for managing content within the NGC Registry. The CLI operates within a shell and lets you use scripts to automate commands.

**NGC Catalog CLI**

The NGC Catalog CLI is available to you if you have guest access to the NGC Registry, and with it you can

- View a list of GPU-accelerated Docker container images, pre-trained deep-learning models, and scripts for creating deep-learning models.
- Download container images, models, and resources.

**NGC Registry CLI**

The NGC Registry CLI is available to you if you are logged in with your own NGC account or with an NGC Private Registry account, and with it you can

- View a list of GPU-accelerated Docker containers available to you as well as detailed information about each container image.
- See a list of deep-learning models and resources as well as detailed information about them.
- Download container images, models, and resources.
- Upload container images, models, and resources.
- Create and manage users and teams (available to NGC Private Registry administrators).

### 7.1. Installing NGC Catalog CLI

To install NGC Catalog CLI,

1. Enter the NGC website [https://ngc.nvidia.com](https://ngc.nvidia.com) as a guest user.
2. In the top right corner, click **Welcome Guest** and then select **Setup** from the menu.
3. Click **Downloads** under **Install NGC CLI** from the Setup page.
4. From the CLI Install page, click the Windows, Linux, or MacOS tab, according to the platform from which you will be running NGC Catalog CLI.
5. Follow the instructions to install the CLI.
6. Verify the installation by entering `ngc --version`.
   The output should be “NGC CLI x.y.z” where x.y.z indicates the version.

7.2. Installing NGC Registry CLI

To install NGC Registry CLI,

1. Log in to your enterprise account on the NGC website (https://ngc.nvidia.com).
2. In the top right corner, click your user account icon and select Setup, then click Downloads under Install NGC CLI from the Setup page.
3. From the CLI Install page, click the Windows, Linux, or MacOS tab, according to the platform from which you will be running NGC Registry CLI.
4. Follow the instructions to install the CLI.
5. Verify the installation by entering `ngc --version`.
   The output should be “NGC CLI x.y.z” where x.y.z indicates the version.

7.3. Getting Help Using NGC CLI

To run an NGC CLI command, enter “ngc” followed by the appropriate options.

To see a description of available options and command descriptions, use the option `-h` after any command or option. Because the help content is targeted to your user role, be sure to obtain an API key and then run `ngc config set` before using the `-h` option.

**Example 1:** To view a list of all the available options for ngc, enter

C:\>ngc -h

**Example 2:** To view a description of the registry image command and options, enter

C:\>ngc registry image -h

**Example 2:** To view a description of the registry image info command and options, enter

C:\>ngc registry image info -h

7.4. Running the Diagnostics

You can view the following diagnostic information for the NGC CLI user:

- Current time
- Operating system
- Disk usage
- Current directory size
Introduction to the NGC Catalog and Registry CLIs

- Memory usage
- NGC CLI environment variables set
- API gateway connectivity
- API connectivity to the container registry
- API connectivity to the model registry
- External IP
- NGC CLI configuration values
- User information (ID, name, and email)
- User org roles
- User team roles

**Syntax**

```
ngc diag [all,client,install,server,user]
```

where

- **all**
  Produces the maximum amount of diagnostic output
- **client**
  Produces diagnostic output only for the client machine.
- **install**
  Produces diagnostic output only for the local installation.
- **server**
  Produces diagnostic output only for the remote server.
- **user**
  Produces diagnostic output only for the user configuration.

**Note:**

nvcr.io/nvidia, nvcr.io/partners and nvcr.io/hpc are reserved namespaces and do not grant user permissions to push or delete container images, models, helm and all artifacts.
Trademarks

NVIDIA, the NVIDIA logo, and Volta are trademarks and/or registered trademarks of NVIDIA Corporation in the United 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

© 2020-2021 NVIDIA Corporation. All rights reserved.