NGC Overview

Application Note
Table of Contents

Chapter 1. What is NVIDIA NGC? ........................................................................................................ 1
  1.1. NGC Catalog .............................................................................................................................. 1
  1.2. NGC Private Registry ................................................................................................................ 2
  1.3. NVIDIA-Certified Systems ....................................................................................................... 2

Chapter 2. Why NGC Software ........................................................................................................ 4

Chapter 3. NGC Organizations and Teams .................................................................................. 5
  3.1. NGC Org Owner and Other Org Users .................................................................................... 6
  3.1.1. Example of Adding a User to an Org ................................................................................ 7
  3.2. New User Accepting an NGC Org Invitation ........................................................................... 8
  3.3. NGC Org Authentication Using an External SSO Provider .................................................... 9

Chapter 4. Accessing NGC Software .......................................................................................... 11

Chapter 5. Registering and Activating a New NGC Org to Obtain Authenticated Access .......... 12
  5.1. Signing Up for an NVIDIA Account and Activating a New NGC Org ......................... 12
  5.2. Signing In to your NGC Org Owner Account ...................................................................... 16
  5.3. Activating Your Subscription [Offer Dependent] ................................................................. 18
  5.4. Generating Your NGC API Key ............................................................................................ 21

Chapter 6. Obtaining a Private Registry ...................................................................................... 23

Chapter 7. NGC Support Services ............................................................................................. 26

Chapter 8. Introduction to the NGC Catalog and Registry CLIs ............................................... 28
  8.1. Installing NGC Catalog CLI ................................................................................................. 28
  8.2. Installing NGC Registry CLI ............................................................................................... 29
Chapter 1. What is NVIDIA NGC?

NVIDIA NGC™ is the cloud platform offering fully managed services, including NeMo LLM, BioNemo, and Riva Studio for NLU and speech AI solutions. AI practitioners can take advantage of NVIDIA Base Command for model training, NVIDIA Fleet Command for model management, and the NGC Private Registry for securely sharing proprietary AI software. NGC also hosts a catalog of GPU-optimized AI software, SDKs, and Jupyter Notebooks that help accelerate AI workflows and offers support through NVIDIA AI Enterprise.

Enterprises gain access to their AI cloud services through a dedicated virtual NGC organization where their services are enabled.

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 simplifies building, customizing, and integrating 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 data center, 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 everyday AI tasks 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 state-of-the-art accuracy to enable reproducibility as well as documentation and code samples making 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 configure, deploy and update applications across Kubernetes environments more easily. The NGC Catalog provides Helm charts for deploying 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.

**Quick Deploy**

NGC provides a “Quick Deploy” feature that allows you to deploy content directly to Google Cloud Vertex AI Workbench service from the NGC Catalog, making it easier to start building with NVIDIA AI. With just a click of a button, the NGC quick deploy feature takes care of end-to-end setup requirements such as fetching the Jupyter notebooks, configuring the GPU instance, installing dependencies, and running a JupyterLab interface to get started with the development quickly.

For more information, visit the [NGC Catalog](#) website.

### 1.2. **NGC Private Registry**

NVIDIA has developed the NGC private registry to provide a secure space to store and share custom containers, models, Jupyter notebooks, and Helm charts within your enterprise. The NGC private registry is available to DGX and NVIDIA AI Enterprise customers. For more information, refer to [Obtaining a Registry](#).

### 1.3. **NVIDIA-Certified Systems**

Another feature of NGC is the NVIDIA-Certified Systems program which assembles the industry’s most complete set of accelerated workload performance tests to help its partners deliver the highest-performing systems. NVIDIA-Certified Systems are tested with the most powerful enterprise NVIDIA GPUs and networking and are evaluated by NVIDIA engineers for performance, functionality, scalability, and security. NVIDIA-Certified Systems have been proven to deliver predictable performance and enable enterprises to deploy optimized
platforms for AI quickly, Data Analytics, HPC, high-density VDI, and other accelerated workloads in the data center, at the edge, and on the desktop.
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 data center or edge deployments. Supported GPUs include H100, 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. NGC Organizations and Teams

An NGC org (organization) is a dedicated virtual instance used to enable and manage NVIDIA cloud services.

When users sign up to NGC using the self-sign-up steps, they are assigned an individual org with the NGC Catalog service enabled which can be accessed by a single user, the org owner.

An NVIDIA premium cloud service subscription, such as NVAIE or NeMoLLM, will be granted through purchase, an early access program, or through the NGC self-service portal. To enable the subscription on NGC, a dedicated org is created as an enterprise org, or an individual’s org is converted to an enterprise {multi-user} org.
3.1. NGC Org Owner and Other Org Users

The Org Owner is the initial role assigned to a user when a new org is created.

There are two ways NVIDIA collects the user contact information to assign the Org Owner role:

1. A user can sign up for a new NVIDIA account through the NGC sign-in page and register for a new NGC individual org.

2. A user is onboarded by an NVIDIA representative directly. The user provides the org owner information to an NVIDIA representative, and an NVIDIA NGC org invite email is sent to the user’s email address. The user is then requested to sign into NGC or is prompted to create an NVIDIA account that is used to sign in to NGC as Org Owner.

Org Owner possesses the highest admin privileges in an org. An org owner of an enterprise org can add/edit/delete teams and users. The org owner can define access control rules against each added user by managing the assignment of teams and roles. When a new user is added, the org owner can assign access to the entire org or to any number of teams within the org. Then, the org owner can control user access by assigning a role based on each user’s requirement to perform their functions.

To verify the type of org you are an owner of, select **Organization > Organization Profile** from the left navigation menu.
3.1.1. Example of Adding a User to an Org

The following is an example of adding a user to an enterprise org that is enabled with Base Command cloud service.

To add a new user to an org or team, select Organization > Users from the left navigation menu.

On the Users page, click Invite New User to open the Add User page.

In this example, the org owner adds a new user to team “team1” and assigns admin roles across all available cloud services, plus grants the privilege to add new users by assigning the org “User Admin” role. To limit the user’s access, the org owner can assign a lesser-privileged role under each cloud service.

To learn more about NGC cloud services user roles, refer to the links below.

▶ Base Command roles
NGC Organizations and Teams

- Fleet Command roles
- Private Registry roles
- NVAIE viewer role

Omniverse roles are explained in the table below:

<table>
<thead>
<tr>
<th>Role</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>OMNIVERSE_ADMIN</td>
<td>Can create new Nucleus Cloud instances and access all features of Nucleus Cloud instances owned by the Organization</td>
</tr>
<tr>
<td>OMNIVERSE_USER</td>
<td>Can access Nucleus Cloud instances owned by the Organization and use all features</td>
</tr>
<tr>
<td>OMNIVERSE_READ</td>
<td>Can access Nucleus Cloud instances owned by the Organization and use all features</td>
</tr>
</tbody>
</table>

Note that by assigning the user at the team level, the new user can only perform commands within the assigned team. If the new user must possess access privilege to the entire org, the user must be assigned at the “Organization” level.

After the org owner assigns the roles to the user being created, the owner must click Add Role to confirm the assignment of roles. After the roles are assigned, the Create User button becomes enabled for the org owner to select to complete the creation of the user.

The org owner or user admin can check the new user was created successfully by reviewing the list of Pending Invitations in the Users panel.

A new user may not receive the NGC invite email, possibly due to temporary internet issues, the email ending up in the spam folder, or a corporate email firewall blocking the email. The org owner or user admin can resend the email from the Pending Invitations list by clicking on the ... (ellipses) icon next to the desired user and selecting Resend Invitation.

3.2. New User Accepting an NGC Org Invitation

A user newly added to an org will receive an NGC invite email message requesting the user to join the NGC org.
After clicking the acceptance link in the email message, the user is redirected to the NGC sign-in page, where they enter their email address to sign in. If this is the user’s first time signing into NVIDIA NGC, they will be automatically prompted to create a new NVIDIA account.

**Note:** NGC cloud service users will not require an NVIDIA account if the NGC org has been federated with an external SSO/IdP identity service. Users can be created using their company email address, and authentication to NGC is performed by the external SSO/IdP identity service. To learn more about how to bring your own SSO/IdP provider, see [NGC Org Authentication Using an External SSO Provider](#).

### 3.3. NGC Org Authentication Using an External SSO Provider

In special cases, an enterprise org can be federated to an external SSO/IdP identity service to centrally manage a company’s rules for user authentication to cloud services.
After an NGC org is federated against an enterprise SSO/IdP identity service, the users signing into NGC will automatically be prompted to authenticate against their enterprise SSO/IdP service and redirected back to NGC after a successful sign in. To add new users to an org federated to an external SSO/IdP provider, the org owner follows the same steps described in Section 3.5.2: Example of Adding a User to an Org.

Note the org owner role is created as the break-glass/recovery account and must always authenticate through an NVIDIA account. This provides access to the org in case the external SSO/IdP service is unavailable. See section Signing In to your NGC Org Owner Account to learn how to sign in to your org owner account.

NVIDIA recommends org owners add themselves to their enterprise org as new users and assign access rules set at the org level with admin roles across all org services. Using the same org owner email address to add themselves as new admin users is supported. The users’ new service admin accounts will be authenticated using the external SSO/IdP authentication service and inherit the same role privileges as their org owner account.
Chapter 4. Accessing NGC Software

There are four 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 that 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 sign into their NGC org using an NVIDIA account or sign in to NGC using SSO if the user’s org is federated to their external SSO/IdP service. Pulling or downloading the software requires the user to provide their API key.

**Approved Access:** The publisher must grant user approval to access the software under this category. Once access is granted, the user will receive a notification with further instructions for accessing the software. Instructions to request access are provided in the overview section of the respective software.

**Subscription:** To access subscription-based software, users must provide Business Address details and a token (serial number, activation code, and so on). Note that this category also requires Authenticated Access.
Chapter 5. Registering and Activating a New NGC Org to Obtain Authenticated Access

Instructions for registering and activating a new NGC org.

5.1. Signing Up for an NVIDIA Account and Activating a New NGC Org

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

1. Go to the NGC sign-in page from your browser, enter your email address, and then click Continue.
2. At the **Create your Account** screen, verify your email address and create a password. Review the NVIDIA Account Terms of Use and Privacy Policy and click **Create Account** to accept and proceed with account creation.
A verification email is sent to your email address.

3. Open the NVIDIA account creation email and click **Verify Email Address**.
You are automatically directed to nvidia.com and see an Email verified successfully page.

4. At the Almost done! dialog, set your communications preferences, and then click Submit.

5. Complete your user profile at the Set Your Profile screen, agree to the NVIDIA GPU Cloud Terms of Use, and then click Submit.
Your NVIDIA account is created, and you will be automatically logged in to your NGC org.

5.2. Signing In to your NGC Org Owner Account

NVIDIA accounts always handle an NGC org owner authentication. If your org is federated to an external SSO/IdP service as explained in NGC Org Authentication Using an External SSO Provider, the company email address used to create the org owner account will be redirected to the company’s external SSO provider when the email address is entered at the NGC sign-in page. This prevents access to the org owner account. To gain access to your org owner account, you must go through NGC’s alternate sign-in method.

1. At the NGC sign-in page, select Use alternate method to access the NVIDIA account login page.
2. At the **Your NVIDIA Account** sign-in page, enter your email address and click **Continue**. You will be prompted to enter your password. Click **Log in** when done.
5.3. Activating Your Subscription (Offer Dependent)

This section describes activating a subscription and linking it to your NGC Account.

Note: These steps are only required for customers who have been given an activation code as part of the purchase of a GPU or DPU.

1. Access the activation page directly via [Activate Subscription](#).

2. Sign in to NGC with your email address and password if prompted. If you have not created an NGC account, [create one now](#).

3. On the [Activate Subscription](#) page, enter your Business Information using your company’s headquarters address and the serial number or activation code described by the specific offer.

   If entering multiple serial numbers or activation codes, use a comma to separate each.
4. Click **Activate Subscription**.

5. Once the system validates the serial numbers, review the information displayed and click **Request Activation**.
6. The **Subscriptions** page will display for your organization with the active NVIDIA AI Enterprise subscription.

7. Use the left navigation and click **Enterprise Catalog** to access the NVIDIA AI Enterprise software suite.
5.4. Generating Your NGC API Key

This section describes obtaining 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://nvc.nvidia.com/signin/email and then enter your email and password.
2. Click your user account icon in the top right corner and select Setup.
3. Click **Get API key** to open the Setup > API Key page.

   The API Key is the mechanism to authenticate your access to the NGC container registry.

4. Click **Generate API Key** to generate your API key.

   A warning message shows that your old API key will become invalid if you create a new one.

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 securely.

   **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 6. 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 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 immediately log into the NVIDIA Enterprise Support portal.

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. It is essential **not** to share this entitlement link outside of your organization.

Registration will provide an NGC private registry and NVIDIA Enterprise Support accounts. You’ll receive a welcome email, at which time you can activate your NGC private registry account.
Chapter 7. 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 and virtual machines. Support is also provided for CUDA and associated drivers. The OEM partner is typically responsible for providing support at the hardware and OS level with additional support for container orchestration tools. The following is the breakdown of the various software covered under the contract.
Figure 1. Components of the Stack Covered Under NGC Support Services
Chapter 8. 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 and detailed information about each image.
- See a list of deep-learning models and resources and 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).

For more details and best practices, visit the NGC CLI documentation page.

8.1. 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 CLI from the Setup page.
4. From the CLI Install page, click the Windows, Linux, or macOS tab, according to the platform 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.

8.2. 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 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.
Trademarks

NVIDIA and the NVIDIA logo 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.