*Four things are needed to have an NGC instance system in Google cloud where one can train your NN on NVIDIA GPUs:*

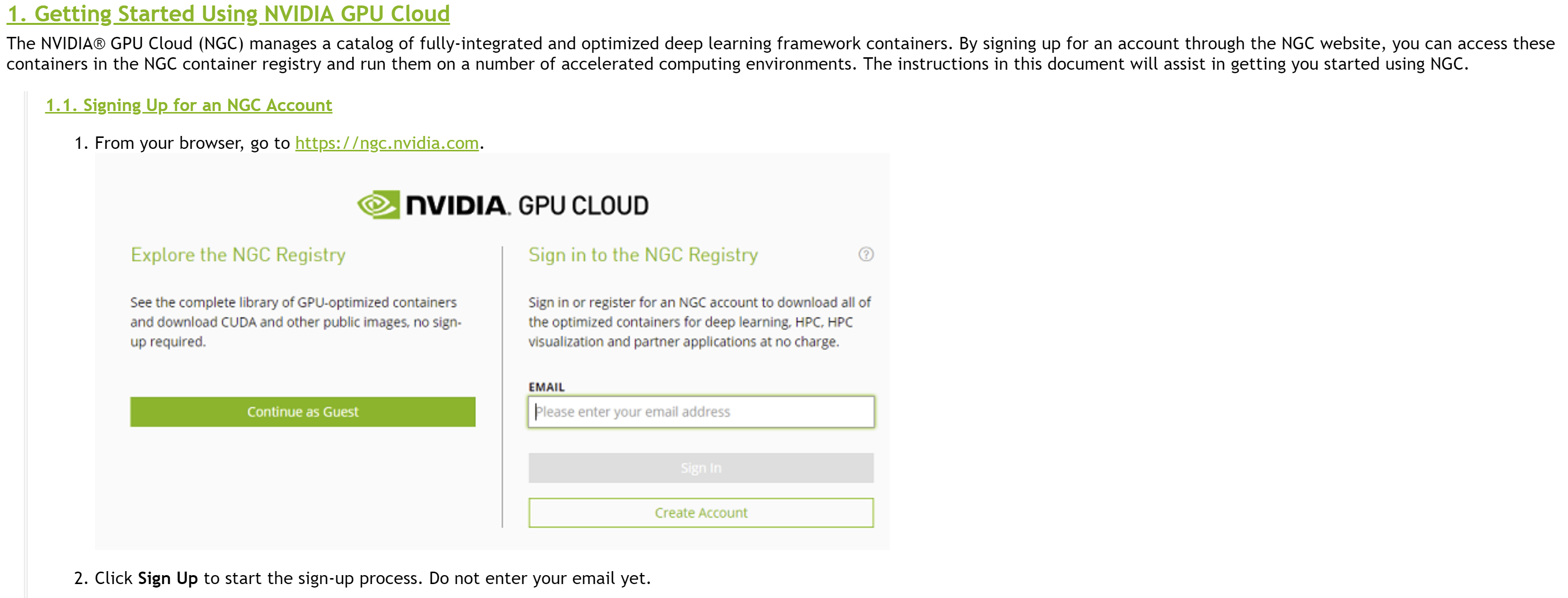
*1) NVDIA GPU Cloud account*

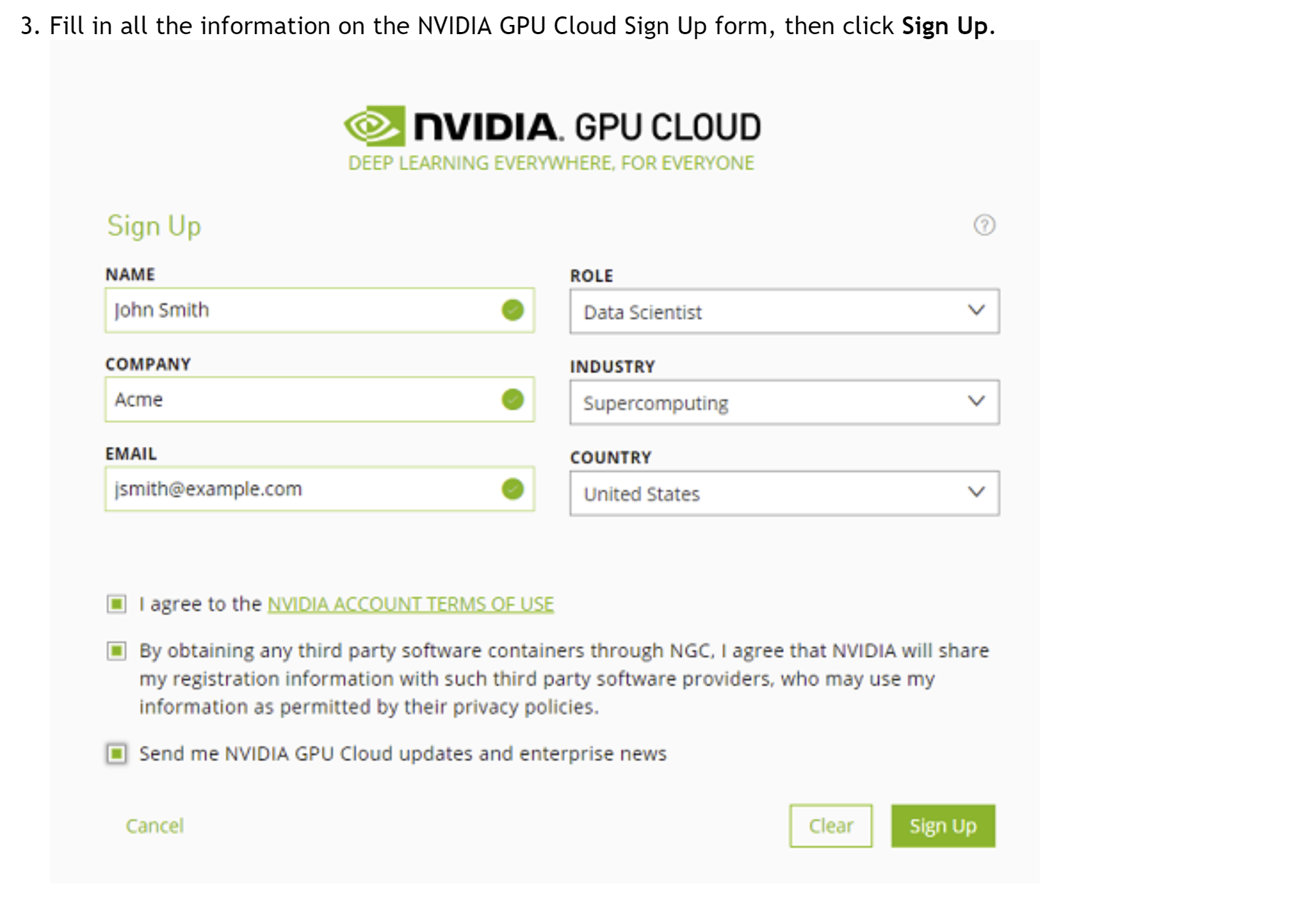
*2) Google Cloud account and starting a new project (https://console.cloud.google.com/)*

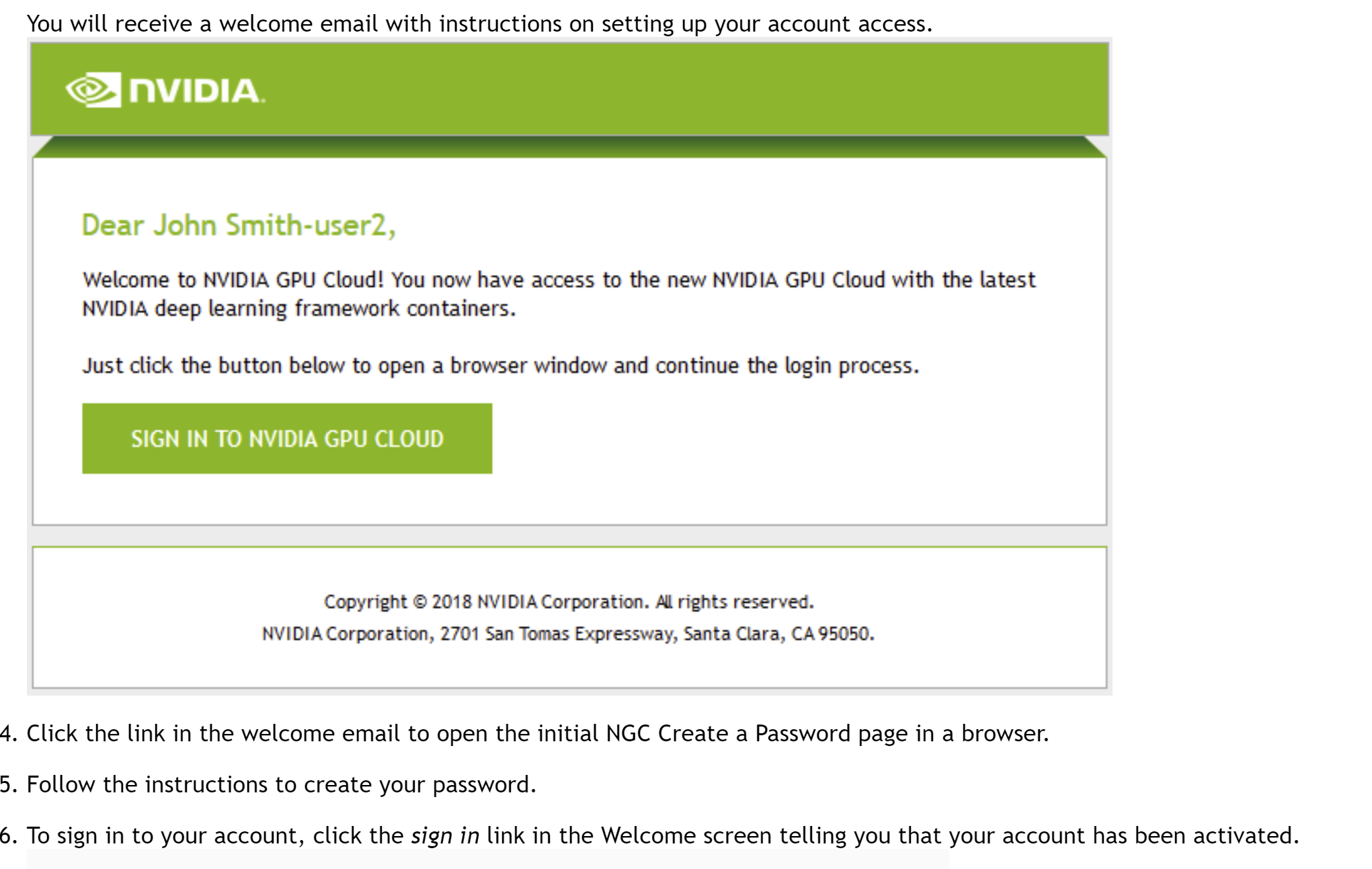
*3) Generate your own SSH keys and adding them to metadata of the new project in Google Cloud*

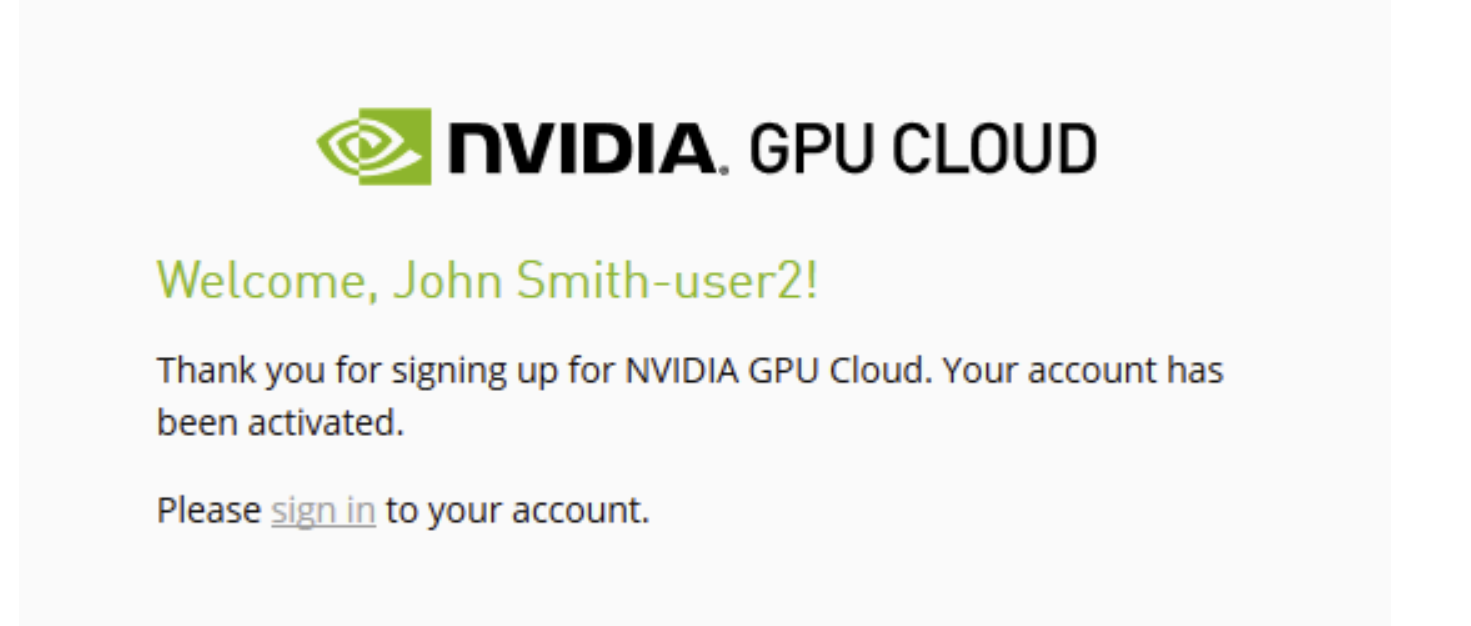
*4) Setting Firewall rules and Creating a GPU instance in Google cloud*

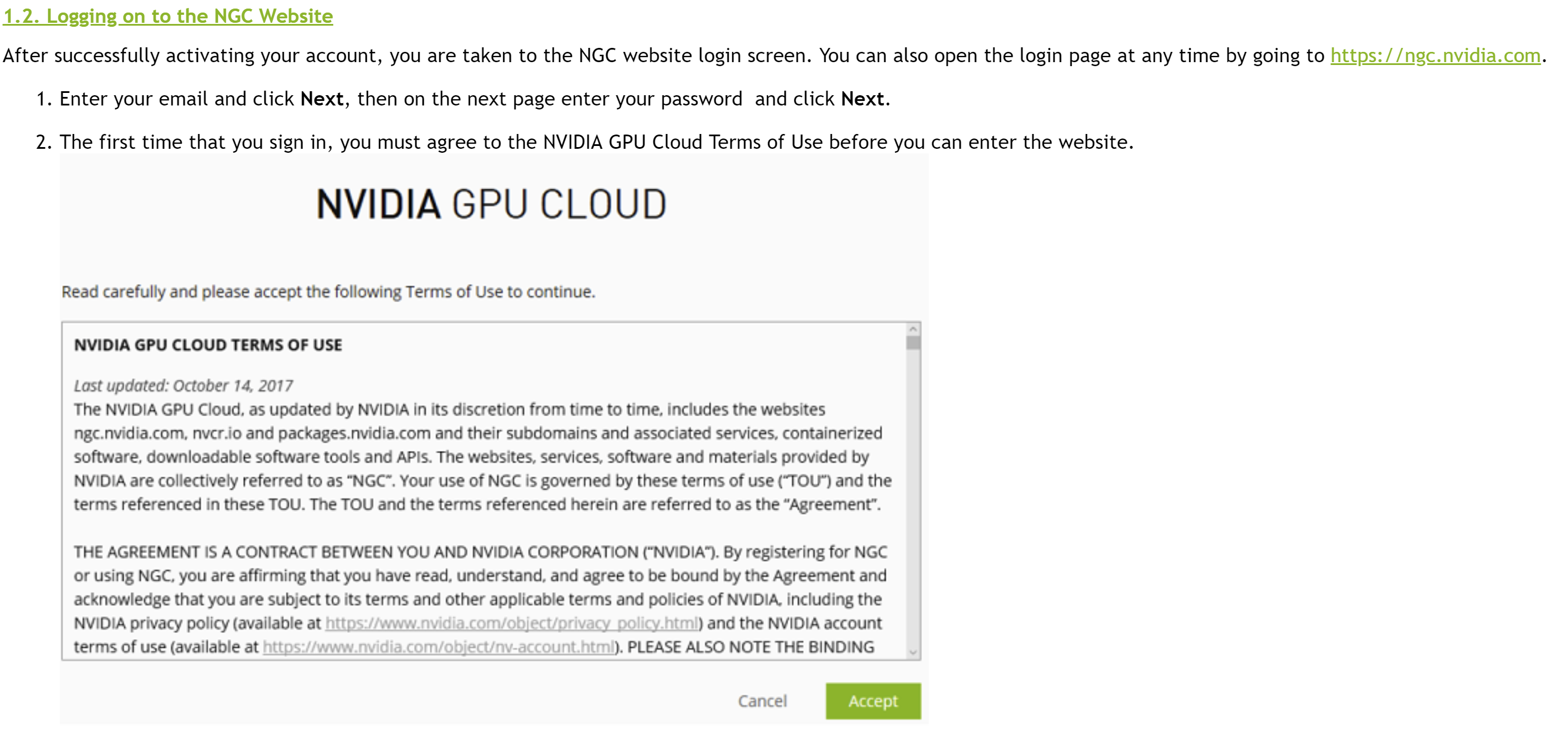
**Setting up the NVIDIA GPU Cloud account and generating the API Key:**

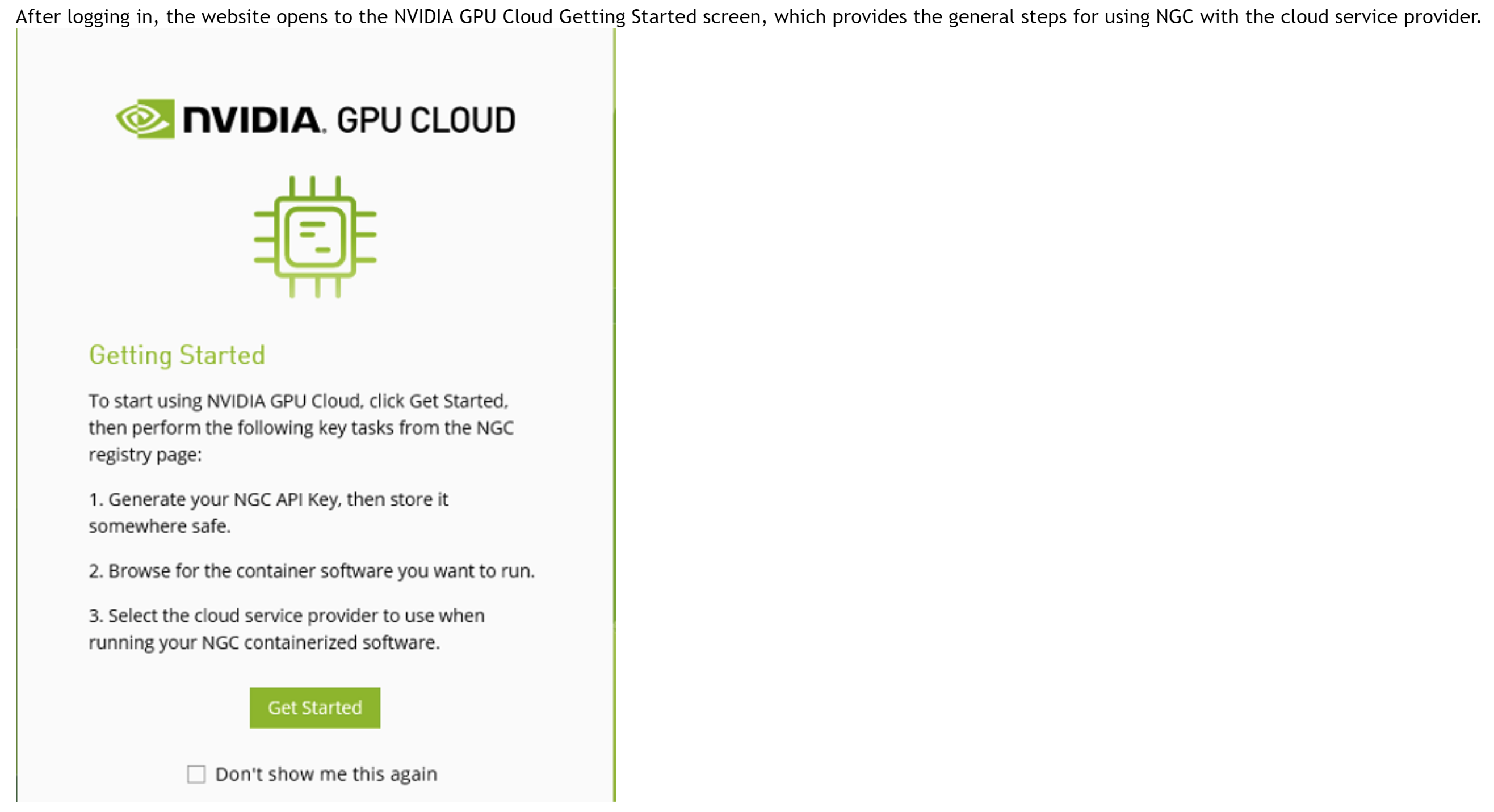


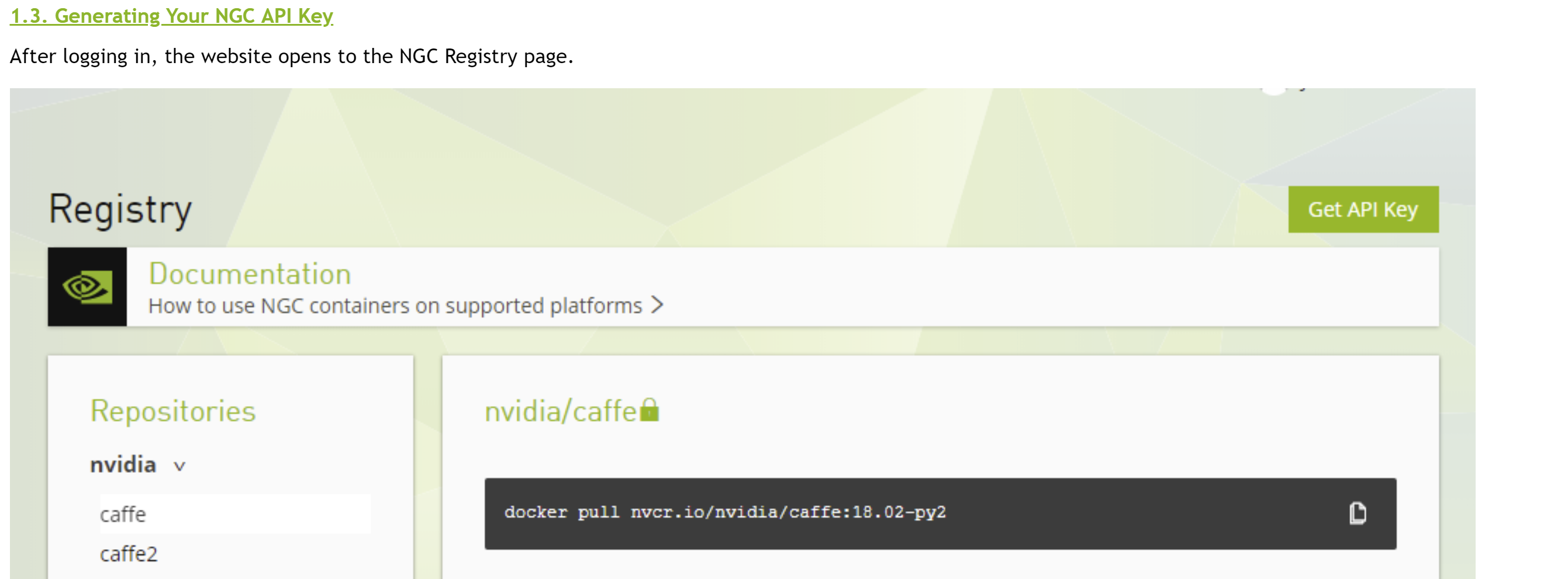


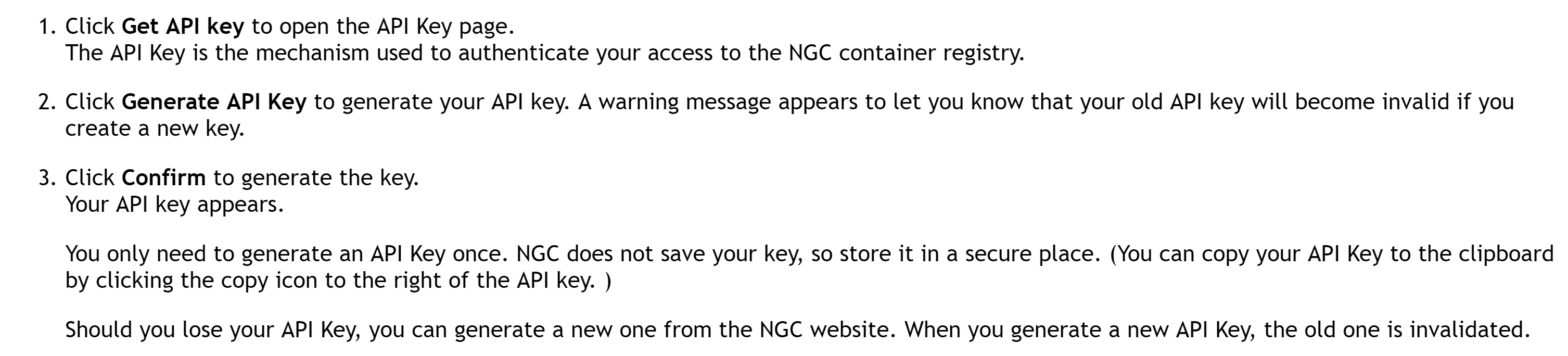












**Generate your own SSH keys and adding them to metadata of the new project in Google Cloud**

**[Setting Up SSH Keys](https://docs.nvidia.com/ngc/ngc-gcp-setup-guide/before-you-start.html" \l "setting-up-ssh-keys)**

The Google Compute Engine generates and manages an SSH key automatically for logging into your instance (see the Google Cloud documentation [Connecting to Instances](https://cloud.google.com/compute/docs/instances/connecting-to-instance).). However, to facilitate logging into the NGC container registry upon the initial connection to the VM instance, you need to

1. Generate your own SSH keys (see [Creating a new SSH key](https://cloud.google.com/compute/docs/instances/adding-removing-ssh-keys#createsshkeys) for instructions), and then
2. Add them to the metadata for your project (see [Adding or Removing Project-Wide Public SSH Keys](https://cloud.google.com/compute/docs/instances/adding-removing-ssh-keys#project-wide)for instructions).

If you do not prepare your SSH keys before launching and connecting to your VM instance, you will not be able to access the NGC container registry initially. In that case you will need to

1. Add yourself to the docker group after connecting to the instance.

sudo usermod -aG docker $USER

1. Restart the session.

**Setting Firewall rules and Creating a GPU instance in Google cloud**

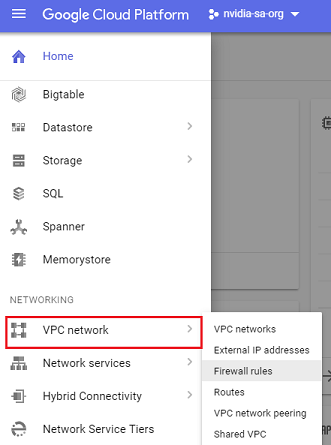
**[Setting Firewall Rules](https://docs.nvidia.com/ngc/ngc-gcp-setup-guide/before-you-start.html" \l "setting-firewall-rules)**

NVIDIA recommends setting firewall rules to allow external access to ports 443 (HTTPS), 8888 (DIGITS), and any other ports that may be needed. This should be done before launching an instance to avoid having to stop the instance when setting any firewall rules later.

**Note:**

You can specify that HTTPS traffic be allowed using the VM Instance Details page, but changing that setting also requires that the instance be stopped.

1. Log in to [https://console.cloud.google.com](https://console.cloud.google.com/)
2. Verify you are in the correct **Project**.
3. Click the Products and Services menu icon, then scroll down to the **Networking** section and click **VPC Network**->**Firewall Rules**.



1. Click **Create Firewall Rule**.
2. Enter the following information to specify the firewall rule you want to create.
   * **Name**: NVIDIA recommends the following naming format

For HTTPS: “default-allow-https”

For DIGITS: “default-allow-digits”

You can also create rules for other DIGITS versions, such as DIGITS4

* + **Direction of traffic**: "Ingress"
  + **Action on match**: "Allow"
  + **Targets**: "All instances in the network"
  + **Source filter**: "IP ranges"
  + **Source IP ranges**: "0.0.0.0/0"
  + **Protocols and ports**: "Specified protocols and ports", then enter

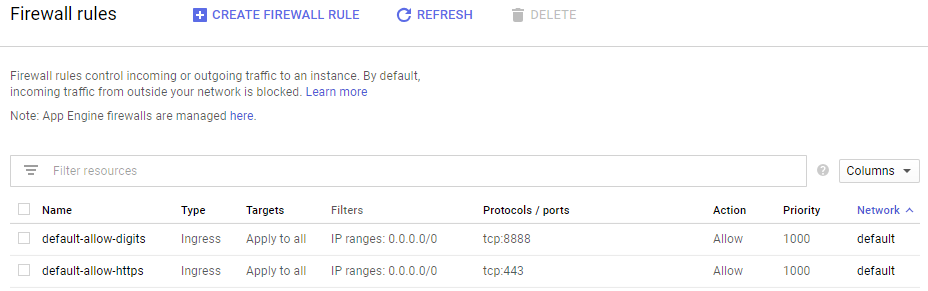
For HTTPS: “tcp:443”

For DIGITS: “tcp:8888”

You can enter ports for other DIGITS versions as well

1. Click **Create**.

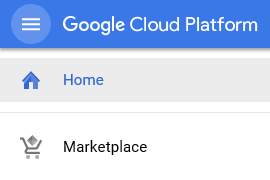
Your new firewall rules should appear on the Firewall Rules page.



**[Creating Your GPU Instance](https://docs.nvidia.com/ngc/ngc-gcp-setup-guide/deploying-ngc-image-from-gcp-console.html" \l "creating-gpu-instance-gcp-console)**

1. Log in to [https://console.cloud.google.com](https://console.cloud.google.com/).
2. Verify you are in the correct project.
3. Open the Google Cloud Platform Marketplace page.

Click the Products and Services menu icon and select **Marketplace**.

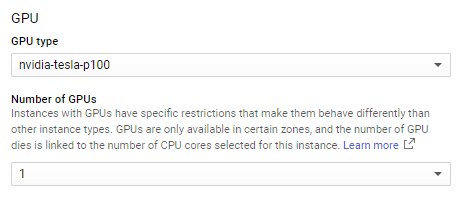


1. Search for NVIDIA GPU Cloud Image.
2. Click the NVIDIA GPU Cloud Image result.
3. From the NVIDIA GPU Cloud Image information page, click **Launch on Compute Engine**.

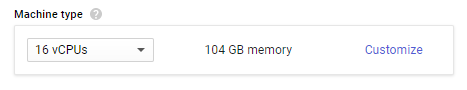
https://docs.nvidia.com/ngc/ngc-gcp-setup-guide/graphics/ui-launch-on-compute-engine.png

1. Configure the NVIDIA GPU Cloud Image deployment.
   1. In "**Name**", enter your new deployment name.
   2. In "**Zone**", select the zone to create the instance (select [one that features the appropriate GPU](https://cloud.google.com/compute/docs/regions-zones/#available)).
   3. Under the **GPU** section, select the GPU type and Number of GPUs.

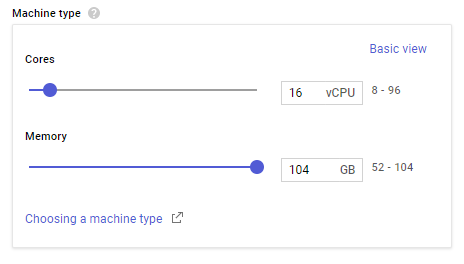
**Note:** If it appears, do not select the **k80** GPU as it is not supported.



* 1. In the "**Machine Type**" section, click **Customize** to open the customize view.



* 1. Assign the Cores (vCPUs) and Memory.



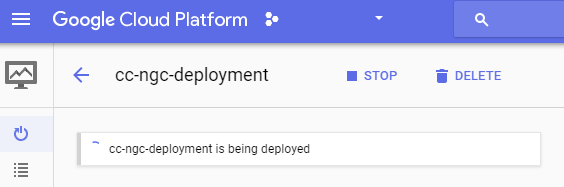
The following ratio is recommended: 1x GPU : 10x vCPU: 60 GB mem)

* 1. In the "**Boot disk**" section, select Standard Persistent Disk.

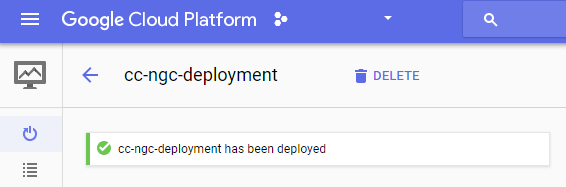


* 1. Make other changes as needed for Networking, Firewall and IP.

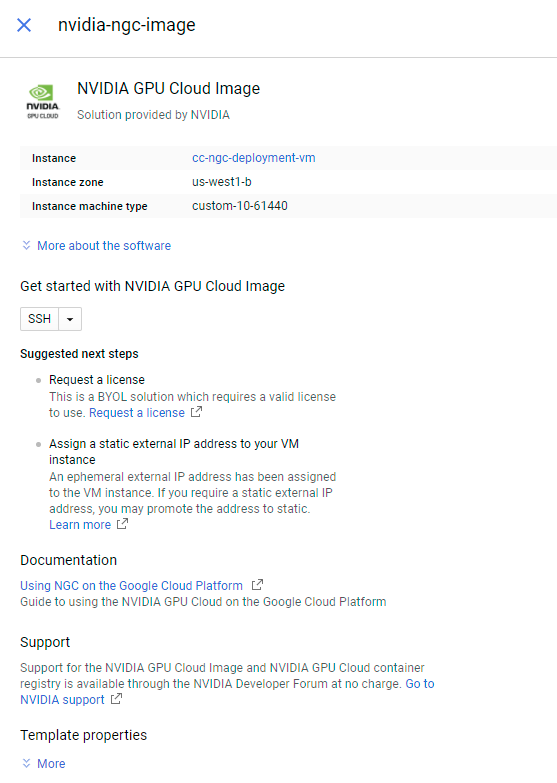
1. Click **Deploy** from the bottom of the page. It may take a few minutes for the deployment process to complete.



Wait for the message that your solution has been deployed, then you can connect to your running instance.



The right side of the deployment page provides information about your deployed image.

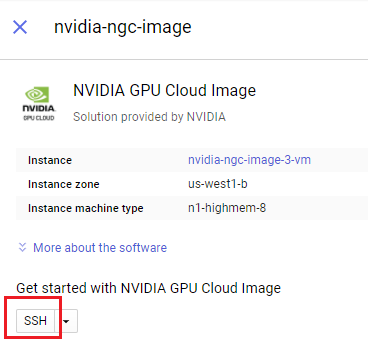


Make a note of the name for your instance. You can stop and restart this instance, as well as attach additional storage to it for datasets.

**[Connecting to Your GPU Instance](https://docs.nvidia.com/ngc/ngc-gcp-setup-guide/deploying-ngc-image-from-gcp-console.html" \l "connecting-to-your-gpu-instance-console)**

Connect to your instance from the Deployment page or from the VM Instance details page.

1. Connect to your instance.
   * If you are still on the Deployment page, you can click **SSH** to connect to your instance.



* + If you are no longer on the Deployment page, you can return to your instance and connect as follows.
    1. Click the Products and Services menu icon, then scroll down to the Compute Engine section and click **VM Instances**.
    2. Either click **SSH** by your listed deployed instance, or click your deployed instance and then click **SSH** from the VM instance details page.

The latest NVIDIA drivers must be installed on the NVIDIA GPU Cloud Image instance before running. If the drivers have not yet been installed on this instance, then upon connecting, the instance startup script asks if you want to download and install the latest NVIDIA drivers.

… NVIDIA GPU Cloud (NGC) is an optimized software environment that

requires the latest NVIDIA drivers to operate. If you do not

download the NVIDIA drivers at this time, your instance will shut down.

Would you like to download the latest NVIDIA drivers so NGC can finish installing? (Y/n)

1. Press **Y** to install the latest NVIDIA drivers and proceed with the connection.

If you press **N**, then the connection process will abort and the instance will be stopped.

The script also initiates the Docker login process automatically, at which point you must enter your NGC API Key.

1. Enter your NGC API Key to complete the login.

After you connect, you can use the terminal to run commands on your Linux instance. When you are done, use the exit command to disconnect from the instance.