

# Cloud GPU Service

## Getting Started

### Product Documentation



## Copyright Notice

©2013–2026 Tencent Cloud. All rights reserved.

Copyright in this document is exclusively owned by Tencent Cloud. You must not reproduce, modify, copy or distribute in any way, in whole or in part, the contents of this document without Tencent Cloud's the prior written consent.

## Trademark Notice



All trademarks associated with Tencent Cloud and its services are owned by the Tencent corporate group, including its parent, subsidiaries and affiliated companies, as the case may be. Trademarks of third parties referred to in this document are owned by their respective proprietors.

## Service Statement

This document is intended to provide users with general information about Tencent Cloud's products and services only and does not form part of Tencent Cloud's terms and conditions. Tencent Cloud's products or services are subject to change. Specific products and services and the standards applicable to them are exclusively provided for in Tencent Cloud's applicable terms and conditions.

# Getting Started

Last updated: 2026-01-16 16:58:48

## Purchase NVIDIA GPU Instance

### Purchase Note

- Before purchasing, please ensure you are familiar with the [Tencent Cloud GPU Service](#) and understand [configuration and pricing](#), and purchase based on actual needs.
- Ensure you understand the AZ of the selected GPU instance. For AZ information, see [GPU Instance Types](#).
- For more information on configuration options during the purchase process, please refer to [Purchasing a Customized Linux Instance](#).

### Purchase Procedure

This guide uses the **GN10Xp instance type** as an example to guide how to quickly purchase a GPU instance:

#### Step 1: Log In to the Purchase Page

#### Step 2: Select Billing Mode, Network, Region, and Instance Type

After going to the purchase page, select the billing mode, region (AZ), and instance. As shown in the following figure:

Basic configurations

Billing mode

- Monthly subscription: Applicable to businesses with long-term and stable demands
- Pay-as-you-go: 22% off in regions outside the Chinese mainland. Applicable to scenarios where the demands are fluctuated significantly
- Spot instances: Up to 97% off in all regions. Up to 97% fees can be saved compared with pay-as-you-go, but the instances may be automatically released.
- Underwriting: Applicable to scenarios where the resources are bought out and one-time payment is not required

Region

China: Shenzhen (Sold Out), Qingyuan (New), Shenzhen Shenyu Tenpay (Sold Out), Shanghai, Shanghai Finance, Hangzhou ec, Nanjing (Promo), Shanghai Self-driving Cloud, Guangzhou

Asia Pacific and Middle East: Shenzhen Finance, Beijing, Tianjin (Sold Out), Beijing Finance (New), Chengdu, Chongqing, Hong Kong (China), Taiwan (China) (New)

Europe and America: (Empty)

Tencent Cloud products in different regions cannot communicate via a private network. The region cannot be changed after the creation. Please select the region closest to your customers to reduce access latency.

Availability zone

Nanjing Zone 3 (Selected) | Recommended Availability Zone | Nanjing Zone 1 (Recommend) | Nanjing Zone 2 | Nanjing Zone 3 (Recommend)

Cloud products in different availability zones within the same region can communicate over the private network.

Instance configurations

Instance

Filter: All CPU cores | All MEMs | Fuzzy search specification name, such as S5.Slv | Reset

Architecture: All Architectures | X86 computing | Heterogeneous Computing (Selected) | HPC Cluster

Instance family: All Instance Families | GPU-based (Selected)

Model: All models | GPU Compute GN7 | GPU Compute GN10X | GPU Compute GN10Xp (Selected) | GPU Graphic G1 | GPU Graphic GNV4 | Show all

Selected model: Selected model GN10Xp.2XLARGE40 (GPU Compute GN10Xp, 10 cores 40 GiB) | The current selected AZ is Nanjing Zone 3. If you need more quotas, apply in the console.

Start Model Comparison |  Hide sold-out products

Instance	Specifications	vCPU	MEM	GPU	GPU memory	CPU Clock Spee...	Processo	Reference fee...
<input checked="" type="radio"/> GPU Compute GN10Xp <small>Low Inventory</small> <a href="#">View more</a>	GN10Xp.2XLARGE40	10Core	40GiB	1 * NVIDIA V100	1 * 32GB	2.5GHz/-	Intel Xeon	
<input type="radio"/> GPU Compute GN10Xp <small>Low Inventory</small>	GN10Xp.5XLARGE80	20Core	80GiB	2 * NVIDIA V100	2 * 32GB	2.5GHz/-	Intel Xeon	

Selected GN10Xp.2XLARGE40 (GPU Compute GN10Xp, 10 cores 40 GiB) | Period: | Quantity: 1 | Configuration fee: | [Next: Configure network and host](#)

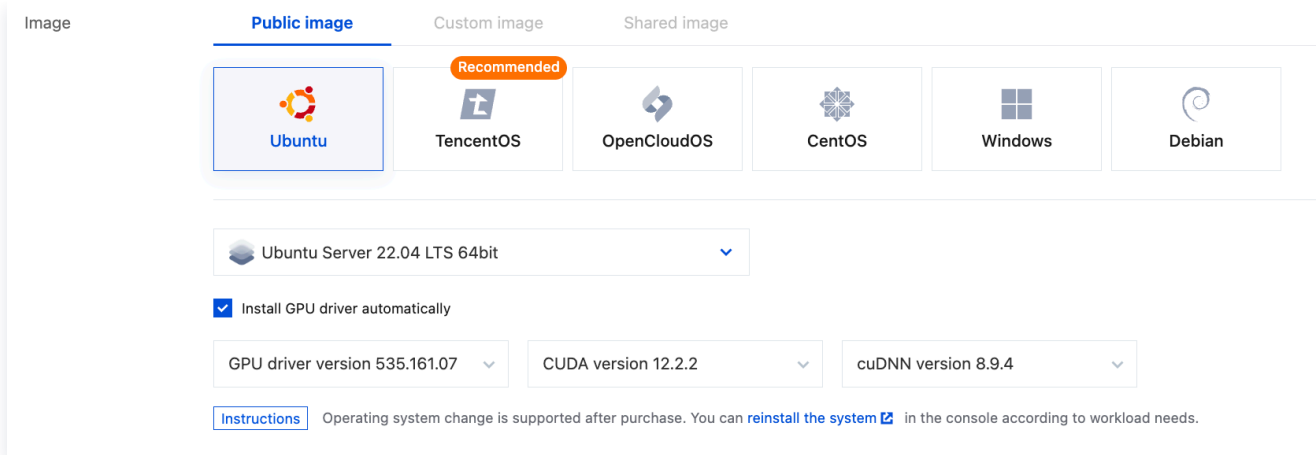
- **Billing Mode:** yearly/monthly subscription, spot instances, pay-as-you-go or underwriting. For details, see [Billing Overview](#).
- **Region and AZ:** You can select the AZ based on the actual purchase page. For details, see [Instance Types](#).
- **Architecture:** Select **Heterogeneous Computing**.
- **Type:** This guide uses **GPU computing instances GN10Xp** as an example. You can refer to [Instance Types](#) to select the most suitable instance configuration.

Step 3: Select Images and Storage

1. Select the image for the GPU.

GPU currently supports three image types: Public Images, Custom Images, and Shared Images. For details, see [Image Overview](#).

New Tencent Cloud users can select **Public Images** and choose the version as needed. As shown in the following figure:

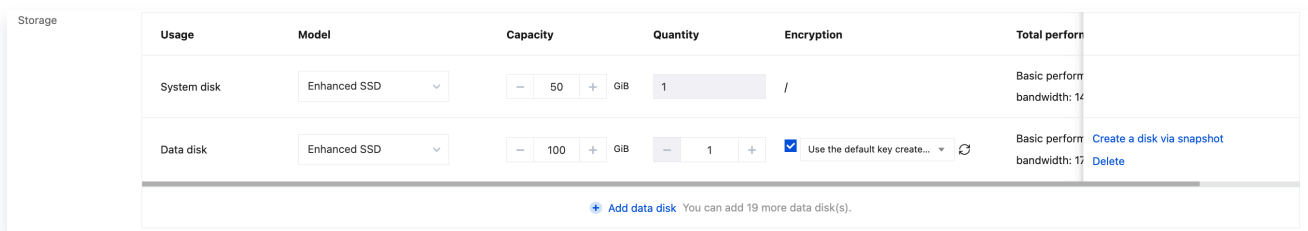


**Note:**

GPU must have the corresponding GPU driver installed to function properly. You can install the driver using the following methods:

- If you select **Public Images**, select **Install GPU drive automatically** to pre-install the corresponding driver version. We recommend using this method. It is only supported for some Linux public images. For details, see [NVIDIA Driver Installation Guide](#).
- If you select **Public Images**, you can manually install the corresponding driver by referring to [NVIDIA Driver Installation Guide](#) after the GPU instance is successfully created.
- If you purchase vGPU instances, you could select public images with pre-installed GRID drivers, with no need to install drivers separately.

2. Select the storage for the Cloud GPU Service. Both the system disk and data disk types and sizes can be flexibly selected. As shown in the following figure:



3. Main **Type Information** is as follows:

Balanced SSD, Enhanced SSD, CloudSSD, Premium Cloud Disk. For CBS types and corresponding billing rules, see [CBS Price Overview](#).

4. After the setting is completed, click **Next: Configure Network and Host**.

### Step 4: Configure Network, Security Group, and Host

1. Select the network and bandwidth for the Cloud GPU Service. As shown in the following figure:

**Custom configuration**

1 Select basic configurations | 2 **Configure network and host** | 3 Confirm configuration

**Network and bandwidth**

Network: [Default-VPC (default)] [Default-Subnet (default)] Available IP addresses within this subnet: 4093.

You can only select a subnet in (Nanjing Zone 3), where your instance is located.  
**The current network is the default VPC/subnet. Please adjust it based on your own needs.**  
 If the existing VPCs/subnets do not meet your requirements, [create a VPC](#) or [a subnet](#) in the console. You can also change the VPC and subnet later.

Manually assign an IP address

Public IP:  Assign Independent Public IP

Line type: BGP

Bandwidth billing mode: Monthly-subscribed bandwidth | **Bill by traffic** | Bandwidth package

Note: The traffic fee is settled on an hourly basis. When the account balance is not enough, the data transfer service will be suspended two hours later. Tencent Cloud provides a 2-Gbps capacity for DDoS mitigation by default.

Bandwidth cap: 1Mbps | 20Mbps | 40Mbps | 200Mbps | - 5 + Mbps

- **Network:** Select an existing virtual private cloud (VPC) or create a new one.
  - **Public IP:** If your instance requires public network access capabilities, you need to select this option, and a public IP address will be assigned to the instance after creation.
  - **Line type:** All public IPs are BGP IPs, which can fully utilize the advantages of Tencent Cloud's DDoS network: low latency and high protection.
  - **Bandwidth billing mode:** For public network bandwidth, the following three billing modes are available:
    - **Monthly-subscribed bandwidth:** Select a fixed bandwidth. When the bandwidth is exceeded, packets will be dropped. Suitable for scenarios with low network fluctuations.
    - **Bill by traffic:** Billed based on actual traffic usage. You can limit the peak bandwidth. Packets will be dropped when the instantaneous bandwidth exceeds the limit (suitable for scenarios with significant network fluctuations).
    - **Bandwidth package:** The bandwidth upper limit of the IP address is affected by the Bandwidth package. The actual limit is the lower of the two. For details, see [Bandwidth package Billing Overview](#).
  - **Bandwidth cap:** The upper limit of the instance's public network bandwidth, set it as needed.
2. Create a new security group or select an existing one to control the access scope of ports. As shown in the following figure:

**Security group**

Security group: **New security group** | Existing security group

Allow common IPs/ports: To open other ports, you can [create a security group](#).

ICMP (Ping the CVM from public network) |  TCP:22 (SSH remote login for Linux) |  TCP:3389 (RDP remote login for Windows)

TCP:80 (HTTP Web server) |  TCP:443 (HTTPS Web server) |  Open for private network (Private network access from other cloud resources (IPv4))

**Note** When "0.0.0.0/0" is entered for source/destination, it means all IP addresses are allowed. Please enter your frequently used IP addresses.

Please make sure port 22 (Linux SSH login) and port 3389 (Windows remote login) are open in the selected security group. You can go to the CVM console to [Modify security group](#).

[View security group rules](#)

### 3. Set the login password for the Cloud GPU Service.

- **Set password:** Password login is set. You can log in to the CVM console to reset the password.
- **SSH key pair:** Tencent Cloud recommends that you use an SSH key pair to log in to Linux instances. View the [SSH Keys Guide](#).
- **Reset password after creation:** After the instance is created, log in to the CVM console and "reset the password" to complete the setup.

### 4. Click **Next: Confirm Configuration**.

## Step 5: Confirm Configuration

### 1. Please verify the following content in the **Confirm Configuration** step. As shown in the following figure:

**Custom configuration**

Select basic configurations    
  Configure network and host    
 **3** Confirm configuration

**Basic and instance configurations** [Edit](#)

CVM billing mode	Monthly subscription	Region	Nanjing	Availability Zone	Nanjing Zone 3
Instance	GN10Xp.2XLARGE40 (GPU Compute GN10Xp, 10 cores 40 GiB)	Image	Public image   Ubuntu   img-487zeit5   Ubuntu Server 22.04 LTS 64bit   20GiB	System disk	Enhanced SSD   50 GiB
Data disk	Enhanced SSD x 1, 100GiB				

**Network and security group** [Edit](#)

Network	vpc-k1ka73x   Default-VPC (default)   10.206.0.0/16	Subnet	subnet-aotbtvkw   Default-Subnet (default)   10.206.16.0/20	Private IP	Not assigned
Public IP	Purchase	Bandwidth billing mode	Bill by traffic 5Mbps	Line type	BGP
Security group	Custom template				

**Other settings** [Reset password after creation](#)

[Save as template](#)   
 [Generate API Explorer best practice scripts](#)

auto-renew  When the device expires but there is sufficient balance in the account, **monthly** auto-renew [Configure automatic renewal](#)  
After purchase, you can modify the automatic renewal cycle in the console. Automatic renewal takes precedence over voucher deduction. [Voucher overview](#)

Protocol  I have read and agree to "[Tencent Cloud Service Terms](#)", [Purchase Notes](#)  
(Monthly subscription instances of Tencent Cloud do not support early returns. Please confirm your actual needs before you pay the order.)

Selected GN10Xp.2XLARGE40 (GPU Compute GN10Xp, 10 cores 40 GiB)    
 Configuration fee [View](#)    
 Public traffic fees [View](#)    
 Back    
 Buy Now

Period [View](#)    
 Quantity

- Verify that the configurations, including instance type, image, storage, bandwidth, and security group, meet your requirements.
- Please select and verify the purchase quantity and duration.

### 2. Read and select **Agree to the "Tencent Cloud Service Terms" and Purchase Notes**, then click **Buy Now**.

## Step 6: Verify the Order and Payment

Check the order information and select a payment method to pay.

After the payment is successful, go to the [Console](#). You can log in to the instance once it has been created and started.

## Getting Started with a Windows GPU Instance

The steps for creating and configuring a Windows GPU instance are the same as those for a Windows CVM instance. For details, see [Purchasing a Customized Windows Instance](#).

## Getting Started with a Linux GPU Instance

The steps for creating and configuring a Linux GPU instance are the same as those for a Linux CVM instance. For details, see [Purchasing a Customized Linux Instance](#).

## Appendix

### GPU Driver Pre-installation Instructions

Cloud GPU Service support automatic installation of GPU drivers for specific Linux public images. Upon initial startup, the GPU driver, CUDA, and cuDNN libraries will be automatically installed based on the selected version. This process takes approximately 15 to 20 minutes. Alternatively, you can manually install the required driver version. For details, please refer to [Driver Version Selection](#).

### Using API to Create a GPU Instance

If you need to create a GPU instance via API, the server provides the interface (RunInstances). Request domain: `cvm.intl.tencentcloudapi.com`. For details, see [RunInstances](#).