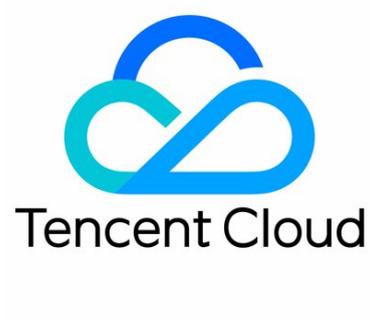


# GPU 云服务器 操作指南



## Copyright Notice

©2013–2025 Tencent Cloud. All rights reserved.

The complete copyright of this document, including all text, data, images, and other content, is solely and exclusively owned by Tencent Cloud Computing (Beijing) Co., Ltd. ("Tencent Cloud"); Without prior explicit written permission from Tencent Cloud, no entity shall reproduce, modify, use, plagiarize, or disseminate the entire or partial content of this document in any form. Such actions constitute an infringement of Tencent Cloud's copyright, and Tencent Cloud will take legal measures to pursue liability under the applicable laws.

## Trademark Notice

 Tencent Cloud

This trademark and its related service trademarks are owned by Tencent Cloud Computing (Beijing) Co., Ltd. and its affiliated companies ("Tencent Cloud"). The trademarks of third parties mentioned in this document are the property of their respective owners under the applicable laws. Without the written permission of Tencent Cloud and the relevant trademark rights owners, no entity shall use, reproduce, modify, disseminate, or copy the trademarks as mentioned above in any way. Any such actions will constitute an infringement of Tencent Cloud's and the relevant owners' trademark rights, and Tencent Cloud will take legal measures to pursue liability under the applicable laws.

## Service Notice

This document provides an overview of the as-is details of Tencent Cloud's products and services in their entirety or part. The descriptions of certain products and services may be subject to adjustments from time to time.

The commercial contract concluded by you and Tencent Cloud will provide the specific types of Tencent Cloud products and services you purchase and the service standards. Unless otherwise agreed upon by both parties, Tencent Cloud does not make any explicit or implied commitments or warranties regarding the content of this document.

## Contact Us

We are committed to providing personalized pre-sales consultation and technical after-sale support. Don't hesitate to contact us at 4009100100 or 95716 for any inquiries or concerns.

# Contents

## 操作指南

登录实例

重启实例

### 安装 NVIDIA 驱动

#### NVIDIA 驱动安装指引

##### 安装 Tesla 驱动

创建实例时自动安装 Tesla 驱动-Linux (推荐)

创建实例后快速安装 Tesla 驱动-Linux (推荐)

手动安装 Tesla 驱动-Linux

手动安装 Tesla 驱动-Windows

使用预装 Tesla 驱动的云市场镜像

##### 安装 GRID 驱动

使用预装 GRID 驱动的镜像 (推荐)

申请 License 并安装 GRID 驱动

##### 安装 CUDA 驱动

### 卸载 NVIDIA 驱动

卸载 Tesla 驱动

卸载 GRID 驱动

### 升级 NVIDIA 驱动

升级 Tesla 驱动

升级 GRID 16 驱动

使用 GPU 监控和告警

# 操作指南

## 登录实例

Last updated: 2024-12-24 15:16:32

在购买并启动了 GPU 实例后，您可以连接并登录它。根据您的本地的操作系统、GPU 实例操作系统和 GPU 实例是否可被 Internet 访问的情况，您可以使用不同的登录方式。

### 先决条件

- 使用密码登录到 GPU 云服务器时，需要使用管理员账号和对应的密码。
- 使用密钥登录到 GPU 云服务器时，需要创建并下载私钥。

### 登录指引

- 若 GPU 实例为 Linux 实例，具体登录指引可参见 [登录 Linux 实例](#)。
- 若 GPU 实例为 Windows 实例，具体登录指引可参见 [登录 Windows 实例](#)。

# 重启实例

Last updated: 2024-09-19 09:49:23

重启操作是维护 GPU 云服务器的一种常用方式，重启实例相当于本地计算机的重启操作系统操作。

## 概述

- **重启准备：**重启期间实例将无法提供服务，因此在重启之前，请确保 GPU 云服务器已暂停业务请求。
- **重启操作方式：**建议使用腾讯云提供的重启操作进行实例重启，而非在实例中运行重启命令（如 Windows 下的重新启动命令及 Linux 下的 Reboot 命令）。
- **重启时间：**一般来说重启操作后只需要几分钟时间。
- **实例物理特性：**重启实例不改变实例的物理特性。实例的公网 IP、内网 IP、存储的任何数据都不会改变。
- **计费相关：**重启实例不启动新的实例计费时间。

## 使用控制台重启实例

登录 [云服务器控制台](#)：

- 重启单个实例：勾选需要重启的实例，在列表顶部，单击**重启**。或在右侧操作栏中，选择**更多 > 实例状态 > 重启**。
- 重启多个实例：勾选所有需要重启的实例，在列表顶部，单击**重启**。即可批量重启实例。不能重启的实例会显示原因。

## 使用 API 重启实例

请参见 [RebootInstances 接口](#)。

# 安装 NVIDIA 驱动

## NVIDIA 驱动安装指引

Last updated: 2024-11-05 14:53:42

NVIDIA GPU 实例需要安装驱动后才可以正常使用，根据业务属于计算或渲染场景，需选择适用的驱动类型和安装方式。腾讯云提供多样化的安装方法，使您能够在创建 GPU 实例时或创建后根据具体需求选择最适合的驱动部署方式，以确保 GPU 云服务器的正常运作并实现更高效的计算和渲染性能。

- 部署免费的 Tesla Driver 和 CUDA SDK，可用作深度学习、科学计算等通用计算场景。
- 部署 GRID Driver 并且使用付费的 License，可开启 GPU 的 OpenGL 或 DirectX 图形加速能力。

### GPU 驱动类型选型指引

实例类型	虚拟化类型	场景	驱动类型	支持安装方式
计算型实例	直通卡	通用计算	Tesla 驱动	<ul style="list-style-type: none"> <li>创建实例时自动安装 Tesla 驱动-Linux (推荐)</li> <li>创建实例后快速安装 Tesla 驱动-Linux (推荐)</li> <li>手动安装 Tesla 驱动-Windows</li> <li>手动安装 Tesla 驱动-Linux</li> </ul>
		图形渲染	GRID 驱动	<ul style="list-style-type: none"> <li>申请 License 并安装 GRID 驱动</li> </ul>
渲染型实例	直通卡	通用计算	Tesla 驱动	<ul style="list-style-type: none"> <li>登录实例使用脚本安装 Tesla 驱动-Linux (推荐)</li> <li>手动安装 Tesla 驱动-Linux</li> <li>手动安装 Tesla 驱动-Windows</li> </ul>
		图形渲染	GRID 驱动	<ul style="list-style-type: none"> <li>使用预装 GRID 驱动的镜像</li> </ul>
	vGPU	图形渲染	GRID 驱动	<ul style="list-style-type: none"> <li>使用预装 GRID 驱动的镜像</li> </ul>

#### 说明:

- NVIDIA GPU 实例的类型包括物理直通卡实例（整卡 GPU）和 vGPU 实例（非单颗 GPU，例如1/4GPU）。
  - 物理直通 GPU 可以使用 Tesla 驱动来支持计算场景，也可以使用 GRID 驱动支持渲染场景。
  - vGPU 只能使用特定版本的 GRID 驱动支持不同场景。
- NVIDIA [渲染型实例](#) 腾讯云提供 GRID 驱动 License，如果您选择使用 [计算型实例](#) 用于渲染场景，需要前往 [NVIDIA 官网购买 License](#)。

### 实例驱动版本选型

实例	GPU 类型	支持驱动类型	自动安装推荐驱动版本	自动安装推荐 CUDA 版本	其他安装方式驱动版本支持
计算型 PNV4	NVIDIA A10	Tesla 驱动 GRID 驱动	Tesla 驱动 535.161.07	CUDA 12.4.0	<b>Tesla 驱动</b> 无特殊要求，官方支持版本即可。 <b>GRID 驱动</b> <ul style="list-style-type: none"> <li>GRID17 驱动</li> <li>GRID16 驱动</li> </ul> 说明：手动安装 GRID 驱动需要您前往 NVIDIA 官网下载并购买 License。版本选择请参见 <a href="#">GRID 驱动版本说明</a> 。
计算型 GT4	NVIDIA A100				
计算型 GN10XP	NVIDIA V100	Tesla 驱动 535.161.07	CUDA 12.4.0 CUDA 11.4.3		
计算型 GN7	NVIDIA T4	Tesla 驱动 470.182.03			
推理型 GI3X	NVIDIA T4				
计算型 GN10X	NVIDIA V100				

计算型 GN8	NVIDIA P40				
计算型 GN6/GN6S	NVIDIA P4				
渲染型 GNV4	NVIDIA A10		-	-	<p><b>Tesla 驱动</b> 无特殊要求，官方支持版本即可。</p> <p><b>GRID 驱动</b></p> <ul style="list-style-type: none"> <li>GRID17 驱动</li> </ul> <p>说明：安装 GRID 17 驱动需要您前往 NVIDIA 官网下载。请参见 <a href="#">GRID 驱动版本说明</a>。获取 License 请联系 <a href="#">售前在线咨询</a>。</p> <ul style="list-style-type: none"> <li>GRID16 驱动 Windows-GRID16.2 驱动 537.70</li> </ul>
渲染型 GNV4v	NVIDIA A10	GRID 驱动	不支持自动安装，可使用预装 GRID驱动的公共镜像	-	Windows-GRID16.2 驱动 537.70
渲染型 GN7vw	NVIDIA T4				<ul style="list-style-type: none"> <li>Linux-GRID16.2 驱动 535.129.03</li> <li>Windows-GRID16.2 驱动 537.70</li> </ul>

# 安装 Tesla 驱动

## 创建实例时自动安装 Tesla 驱动-Linux（推荐）

Last updated: 2024-11-05 14:53:42

### 操作场景

GPU 云服务器正常工作需提前安装正确的基础设施软件，对 NVIDIA 系列 GPU 而言，有以下两个层次的软件包需要安装：

- 驱动 GPU 工作的硬件驱动程序。
- 上层应用程序所需要的库。

为方便用户使用，购买页提供了多种安装 GPU 驱动及关联 CUDA、cuDNN 的方法。在创建 GPU 实例时，可以根据业务需求选择不同方法完成驱动部署。

### 安装方式

安装方式	描述	链接
方式一：方式选择公共镜像后自动安装驱动	在购买页选择镜像步骤，选择公共镜像并勾选后台自动安装 GPU 驱动选项。	<a href="#">选择公共镜像后自动安装驱动（推荐）</a>
方式二：使用脚本安装 GPU 驱动	在购买页自定义数据文本框中，输入驱动自动安装脚本。	<a href="#">使用脚本安装 GPU 驱动</a>
方式三：使用预装 Tesla 驱动的云市场镜像	在购买页选择镜像步骤，选择预装了 Tesla 驱动的云市场镜像。	<a href="#">使用预装 Tesla 驱动的镜像</a>

### 方式一：选择公共镜像后自动安装驱动

1. 在云服务器 [购买页](#) 创建实例过程中，选择镜像步骤请选择 CentOS 或 Ubuntu 镜像。

选择后即出现后台自动安装 GPU 驱动选项，勾选后即可按需选择 CUDA 和 cuDNN 版本。如下图所示：



#### 说明：

仅计算型实例部分镜像版本支持 Tesla 驱动自动安装，以购买页面展示为准。

2. 购买页其他配置选择请参见 [购买 GPU 实例](#)，创建完成后请前往控制台，找到实例并等待10分钟左右驱动安装完成。
3. 参见 [使用标准登录方式登录 Linux 实例](#)，登录实例。
4. 执行以下命令，验证驱动是否安装成功。

```
nvidia-smi
```

如返回类似下图中的 GPU 信息，则说明驱动安装成功。

```
ubuntu@VM-2-124-ubuntu:~$ nvidia-smi
Thu Oct 24 22:11:12 2024

+-----+
| NVIDIA-SMI 535.161.07           Driver Version: 535.161.07   CUDA Version: 12.2   |
+-----+-----+-----+-----+-----+
| GPU  Name                   Persistence-M | Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp   Perf           Pwr:Usage/Cap |      Memory-Usage | GPU-Util  Compute M. |
|                                           |              |                  |
+-----+-----+-----+-----+-----+
|   0   Tesla T4              On          | 00000000:00:08:0 | Off      |          0         |
| N/A   30C    P8              9W / 70W   |      2MiB / 15360MiB |      0%    Default  |
|                                           |              |                  |
+-----+-----+-----+-----+-----+

Processes:
+-----+-----+-----+-----+-----+
| GPU  GI   CI          PID  Type   Process name          GPU Memory |
|   ID  ID  ID              |                   |           Usage      |
+-----+-----+-----+-----+-----+
| No running processes found |
+-----+-----+-----+-----+-----+
```

## 方式二：使用脚本安装 GPU 驱动

### 操作步骤

1. 在云服务器 [购买页](#) 选择 GPU 实例，选择镜像步骤 **不勾选** 后台自动安装 GPU 驱动选项。



2. 在设置网络和主机步骤中的其他设置中单击高级设置，如下图所示：

3. 在高级设置的自定义数据文本框中，输入驱动自动安装脚本。如下图所示：

完整脚本内容如下，请参考 [参数说明](#) 更新参数后将以下脚本完整复制到自定义数据中：

```
#!/bin/bash
sudo rm -f /tmp/user_define_install_info.ini
sudo rm -f /tmp/auto_install.sh
sudo rm -f /tmp/auto_install.log
sudo echo "
DRIVER_VERSION=535.161.07
CUDA_VERSION=12.4.0
CUDNN_VERSION=8.9.7
DRIVER_URL=
CUDA_URL=
CUDNN_URL=
" > /tmp/user_define_install_info.ini
sudo wget https://mirrors.tencentyun.com/install/GPU/auto_install.sh -O /tmp/auto_install.sh && sudo
chmod +x /tmp/auto_install.sh && sudo /tmp/auto_install.sh > /tmp/auto_install.log 2>&1 &
```

- 按照界面信息逐步操作，完成创建云服务器。
- 登录实例**，可输入 `ps aux | grep auto_install` 查看后台运行自动安装脚本。
- 等待10-20分钟后，验证驱动安装完成。使用 `nvidia-smi` 检查驱动安装成功。

```
ubuntu@VM-0-91-ubuntu:~$ nvidia-smi
Mon Aug 12 21:34:07 2024

+-----+
| NVIDIA-SMI 535.161.07           Driver Version: 535.161.07   CUDA Version: 12.2   |
+-----+-----+-----+-----+-----+
| GPU  Name                   Persistence-M | Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp   Perf           Pwr:Usage/Cap |      Memory-Usage | GPU-Util  Compute M. |
|                                           |                  |           MIG M. |
+-----+-----+-----+-----+-----+
|   0   Tesla T4              On          | 00000000:00:08:0 Off |             0         |
| N/A   39C    P8             22W / 70W   |  2MiB / 15360MiB |      0%    Default   |
|                                           |                  |           N/A   |
+-----+-----+-----+-----+-----+

+-----+
| Processes:                      |
| GPU   GI   CI        PID   Type   Process name                      GPU Memory |
|      ID   ID                 |                 |           Usage |
+-----+-----+-----+-----+-----+
| No running processes found      |
+-----+
```

输入 `grep -i "finished" /tmp/auto_install.log` 查看 CUDA、cuDNN 安装记录。

```
ubuntu@VM-0-91-ubuntu:~$ grep -i "finished" /tmp/auto_install.log
install nvidia driver finished
install nvidia cuda finished
install nvidia cudnn finished
driver/cuda/cudnn install finished.....
```

## 参数说明

使用驱动自动安装脚本时，您可以通过指定驱动版本号来安装驱动，也可以传入对应的驱动下载地址安装。

- 根据创建的实例规格和镜像，参考支持组合范围修改对应的 Tesla 驱动、CUDA 和 cuDNN 库版本参数：

```
DRIVER_VERSION=535.161.07
CUDA_VERSION=12.4.0
CUDNN_VERSION=8.9.7
DRIVER_URL=
CUDA_URL=
CUDNN_URL=
```

### 说明：

- 仅 NVIDIA 计算型实例部分 Linux 镜像支持 Tesla 驱动安装脚本。
- 推荐选择最新版本的 Tesla 驱动、CUDA、cuDNN 库。
- 创建实例后，执行脚本安装耗时约10~20分钟。

支持的机型、镜像、Tesla 驱动、CUDA、cuDNN 组合范围如下：

实例类型	公共镜像	Tesla 驱动版本	CUDA 驱动版本	cuDNN 版本
GT4、PNV4、GN10Xp、GN10X、GN8、GN7、BMG5t、BMG5v、	TencentOS Server 3.1 (TK4) Ubuntu Server 22.04 LTS 64位	550.90.07	12.4.0	8.9.7

HCCPNV4h、HCCG5v、 HCCG5vm、HCCPNV4sn、 HCCPNV4sne、HCCPNV5v	Ubuntu Server 20.04 LTS 64位			
GT4、PNV4、GN10Xp、 GN10X、GN8、GN7、 BMG5t、BMG5v、 HCCPNV4h、HCCG5v、 HCCG5vm、HCCPNV4sn、 HCCPNV4sne、HCCPNV5v	TencentOS Server 3.1 (TK4) TencentOS Server 2.4 (TK4)	535.183.06 535.161.07	12.4.0	8.9.7
	Ubuntu Server 22.04 LTS 64位 Ubuntu Server 20.04 LTS 64位 Ubuntu Server 18.04 LTS 64位	535.183.06 535.161.07	12.2.2	8.9.4
GT4、PNV4、GN10Xp、 GN10X、GN8、GN7、 BMG5t、BMG5v、 HCCPNV4h、HCCG5v、 HCCG5vm、HCCPNV4sn、 HCCPNV4sne、HCCPNV5v	TencentOS Server 3.1 (TK4) TencentOS Server 2.4 (TK4)	525.105.17	12.0.1	8.8.0
	Ubuntu Server 20.04 LTS 64位 Ubuntu Server 18.04 LTS 64位			
GT4、PNV4、GN10Xp、 GN10X、GN8、GN7、 BMG5t、BMG5v、 HCCPNV4h、HCCG5v、 HCCG5vm、HCCPNV4sn、 HCCPNV4sne	CentOS 7.x 64位 CentOS 8.x 64位	470.182.03	11.4.3	8.2.4

- 根据创建的实例规格和镜像，参见 [英伟达驱动](#)、[CUDA](#)、[cuDNN 官网文档](#) 指定对应的 Tesla 驱动、CUDA 和 cuDNN 库版本组合，下载后另存为实例可访问的 URL 地址填入参数：

```
DRIVER_VERSION=
CUDA_VERSION=
CUDNN_VERSION=
#确保实例可正常下载URL对应安装包
DRIVER_URL=http://mirrors.tencentyun.com/install/GPU/NVIDIA-Linux-x86_64-535.161.07.run
CUDA_URL=http://mirrors.tencentyun.com/install/GPU/cuda_12.4.0_550.54.14_linux.run
#cuDNN建议使用tar.xz、tgz格式安装包
CUDNN_URL=http://mirrors.tencentyun.com/install/GPU/cudnn-linux-x86_64-8.9.7.29_cuda12-archive.tar.xz
```

#### 说明：

- 如果填写 DRIVER\_URL、CUDA\_URL、CUDNN\_URL 的任意一个参数，DRIVER\_VERSION、CUDA\_VERSION、CUDNN\_VERSION 参数将失效。
- 仅 NVIDIA 计算型实例 Linux 镜像支持 Tesla 驱动安装脚本，可能存在卡型、镜像、GPU 驱动、CUDA 和 cuDNN 库安装包不兼容的风险，推荐您使用指定驱动版本号的方式安装驱动。
- 如果填写非腾讯云内网下载地址，将产生 [网络流量费用](#)，且下载时间会更长。

### 方式三：使用预装 Tesla 驱动的云市场镜像

操作步骤可参见 [使用预装 Tesla 驱动的云市场镜像](#)。

# 创建实例后快速安装 Tesla 驱动-Linux (推荐)

Last updated: 2024-11-21 15:27:22

## 操作场景

GPU 云服务器正常工作需提前安装正确的基础设施软件，对 NVIDIA 系列 GPU 而言，有以下两个层次的软件包需要安装：

- 驱动 GPU 工作的硬件驱动程序。
- 上层应用程序所需要的库。

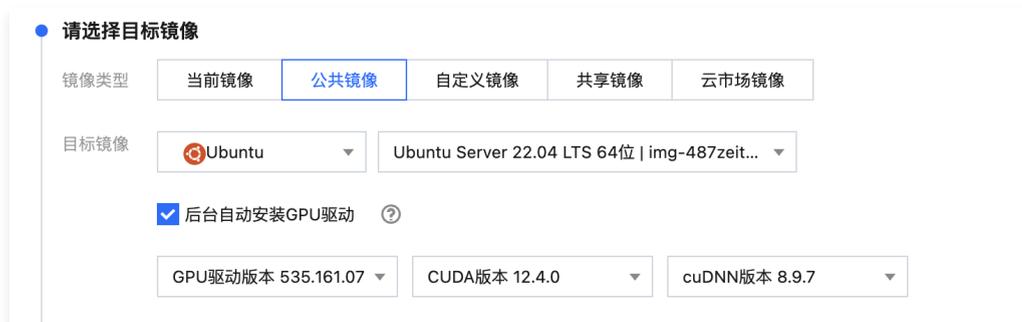
为方便用户使用，购买页提供了多种安装 GPU 驱动及关联 CUDA、cuDNN 的方法。在创建 GPU 实例时，可以根据业务需求选择不同方法完成驱动部署。

## 安装方式

安装方式	描述	链接
方式一：重装公共镜像时自动安装驱动	在重装系统页面选择镜像步骤，选择公共镜像并勾选后台自动安装 GPU 驱动选项。	<a href="#">重装公共镜像时自动安装驱动 (推荐)</a>
方式二：登录实例使用脚本安装 GPU 驱动	登录实例，使用驱动自动安装脚本安装驱动。	<a href="#">登录实例使用脚本安装 GPU 驱动</a>
方式三：使用 TAT 自动化助手安装驱动	进入控制台，使用 TAT 执行公共命令执行驱动安装脚本	<a href="#">使用 TAT 自动化助手安装驱动</a>
方式四：使用云市场镜像重装系统	在重装系统页面选择镜像步骤，选择预装了 Tesla 驱动的云市场镜像。	<a href="#">使用预装 Tesla 驱动的镜像</a>

### 方式一：重装公共镜像时自动安装驱动

1. 在云服务器 [重装系统](#) 过程中，选择镜像步骤请选择 CentOS、Ubuntu 或 TencentOS 镜像。选择后即出现后台自动安装 GPU 驱动选项，可按需选择 CUDA 和 cuDNN 版本。如下图所示：



#### 说明：

仅计算型实例部分镜像版本支持 Tesla 驱动自动安装，以重装系统页面展示为准。

2. 其他配置选择请参见 [重装系统](#)，配置完成后请前往控制台，找到实例并等待10分钟左右驱动安装完成。
3. 参见 [使用标准登录方式登录 Linux 实例](#)，登录实例。
4. 执行以下命令，验证驱动是否安装成功。

```
nvidia-smi
```

如返回类似下图中的 GPU 信息，则说明驱动安装成功。

```
ubuntu@VM-2-124-ubuntu:~$ nvidia-smi
Thu Oct 24 22:11:12 2024

+-----+
| NVIDIA-SMI 535.161.07          Driver Version: 535.161.07   CUDA Version: 12.2   |
+-----+-----+
| GPU   Name                   Persistence-M | Bus-Id        Disp.A | Volatile Uncorr. ECC | | | | |
| Fan  Temp  Perf              Pwr:Usage/Cap |      Memory-Usage | GPU-Util  Compute M. |
|      |      |      |              |                   |              | GPU-Util  Compute M. |
|      |      |      |              |                   |              | MIG M. |
+-----+-----+-----+
|    0   Tesla T4               On          | 00000000:00:08:0 Off |                    0 | | | | |
| N/A   30C    P8              9W / 70W   |  2MiB / 15360MiB |      0%    Default |
|      |      |      |              |                   |              |      N/A |
+-----+-----+-----+

+-----+
| Processes:                      |
| GPU   GI   CI        PID   Type   Process name                      GPU Memory |
|      |   |   |          |      |           |                      Usage |
+-----+-----+-----+
| No running processes found      |
+-----+
```

## 方式二：登录实例使用脚本安装 GPU 驱动

### 操作步骤

1. 登录 [云服务器控制台](#)，选择需要登录的 GPU 云服务器，单击右侧的登录，根据实际需求选择连接方式 [登录实例](#)。



2. 复制下列命令，参考参数说明更新参数后将驱动自动安装脚本保存为 `driver_install.sh`。

```
cat > driver_install.sh << EOF
#!/bin/bash
sudo rm -f /tmp/user_define_install_info.ini
sudo rm -f /tmp/auto_install.sh
sudo rm -f /tmp/auto_install.log
sudo echo "
DRIVER_VERSION=535.161.07
CUDA_VERSION=12.4.0
CUDNN_VERSION=8.9.7
DRIVER_URL=
CUDA_URL=
CUDNN_URL=
" > /tmp/user_define_install_info.ini
sudo wget https://mirrors.tencentyun.com/install/GPU/auto_install.sh -O /tmp/auto_install.sh && sudo
chmod +x /tmp/auto_install.sh && sudo /tmp/auto_install.sh > /tmp/auto_install.log 2>&1 &
EOF
```

如下图所示：



```
ubuntu@VM-0-91-ubuntu:~$ grep -i "finished" /tmp/auto_install.log
install nvidia driver finished
install nvidia cuda finished
install nvidia cudnn finished
driver/cuda/cudnn install finished.....
```

## 参数说明

使用驱动自动安装脚本时，支持两种参数指定版本方式：

1. 指定驱动版本号来安装驱动。
2. 指定安装包下载地址安装驱动。

### 指定驱动版本号来安装驱动

根据创建的实例规格和镜像，参考支持组合范围修改对应的 Tesla 驱动、CUDA 和 cuDNN 库版本参数：

```
DRIVER_VERSION=535.161.07
CUDA_VERSION=12.4.0
CUDNN_VERSION=8.9.7
DRIVER_URL=
CUDA_URL=
CUDNN_URL=
```

#### 说明：

- 仅 NVIDIA 计算型实例部分 Linux 镜像支持 Tesla 驱动安装脚本。
- 推荐选择最新版本的 Tesla 驱动、CUDA、cuDNN 库。
- 创建实例后，执行脚本安装耗时约10–20分钟。

支持的机型、镜像、Tesla 驱动、CUDA、cuDNN 组合范围如下：

实例类型	公共镜像	Tesla 驱动版本	CUDA 驱动版本	cuDNN 版本
GT4、PNV4、GN10Xp、GN10X、GN8、GN7、BMG5t、BMG5v、HCCPNV4h、HCCG5v、HCCG5vm、HCCPNV4sn、HCCPNV4sne、HCCPNV5v	TencentOS Server 3.1 (TK4)	550.90.07	12.4.0	8.9.7
	Ubuntu Server 22.04 LTS 64位 Ubuntu Server 20.04 LTS 64位			
GT4、PNV4、GN10Xp、GN10X、GN8、GN7、BMG5t、BMG5v、HCCPNV4h、HCCG5v、HCCG5vm、HCCPNV4sn、HCCPNV4sne、HCCPNV5v	TencentOS Server 3.1 (TK4)	535.183.06 535.161.07	12.4.0	8.9.7
	TencentOS Server 2.4 (TK4)			
	Ubuntu Server 22.04 LTS 64位 Ubuntu Server 20.04 LTS 64位 Ubuntu Server 18.04 LTS 64位	535.183.06 535.161.07	12.2.2	8.9.4
GT4、PNV4、GN10Xp、GN10X、GN8、GN7、BMG5t、BMG5v、HCCPNV4h、HCCG5v、HCCG5vm、HCCPNV4sn、HCCPNV4sne、HCCPNV5v	CentOS 7.x 64位 CentOS 8.x 64位	525.105.17	12.0.1	8.8.0
	TencentOS Server 3.1 (TK4) TencentOS Server 2.4 (TK4) Ubuntu Server 20.04 LTS 64位 Ubuntu Server 18.04 LTS 64位 CentOS 7.x 64位			

GT4、PNV4、GN10Xp、GN10X、GN8、GN7、BMG5t、BMG5v、HCCPNV4h、HCCG5v、HCCG5vm、HCCPNV4sn、HCCPNV4sne	CentOS 8.x 64位	470.182.03	11.4.3	8.2.4
--	----------------	------------	--------	-------

### 指定安装包下载地址安装驱动

根据创建的实例规格和镜像，参见 [英伟达驱动](#)、[CUDA](#)、[cuDNN 官网文档](#) 指定对应的 Tesla 驱动、CUDA 和 cuDNN 库版本组合，下载后另存为实例可访问的 URL 地址填入参数：

```
DRIVER_VERSION=
CUDA_VERSION=
CUDNN_VERSION=
#确保实例可正常下载URL对应安装包
DRIVER_URL=http://mirrors.tencentyun.com/install/GPU/NVIDIA-Linux-x86_64-535.161.07.run
CUDA_URL=http://mirrors.tencentyun.com/install/GPU/cuda_12.4.0_550.54.14_linux.run
#cuDNN建议使用tar.xz、tgz格式安装包
CUDNN_URL=http://mirrors.tencentyun.com/install/GPU/cudnn-linux-x86_64-8.9.7.29_cuda12-archive.tar.xz
```

#### 说明：

- 如果填写 DRIVER\_URL、CUDA\_URL、CUDNN\_URL 的任意一个参数，DRIVER\_VERSION、CUDA\_VERSION、CUDNN\_VERSION 参数将失效。
- 仅 NVIDIA 计算型实例 Linux 镜像支持 Tesla 驱动安装脚本，可能存在卡型、镜像、GPU 驱动、CUDA 和 cuDNN 库安装包不兼容的风险，**推荐您使用指定驱动版本号的方式安装驱动。**
- 如果填写非腾讯云内网下载地址，将产生 [网络流量费用](#)，且下载时间会更长。

### 方式三：使用 TAT 自动化助手安装驱动

1. 登录云服务器控制台，选择左侧导航栏中的 自动化助手 > [公共命令](#)。
2. 在公共命令页面上方选择实例所在地域，并单击 [安装 GPU 驱动-Linux 系统](#) 模块左下角的 [执行命令](#)。如下图所示：



3. 在执行命令页面，您可修改命令配置，参数配置参见 [参数说明](#)：

参数名	参数值	描述
DRIVER_VERSION	535.161.07	选择安装的 Tesla 驱动版本
CUDA_VERSION	12.4.0	选择安装的 CUDA 版本，和 GPU 驱动版本有依赖关系
CUDNN_VERSION	8.9.7	选择安装的 cuDNN 版本，和 CUDA 版本有依赖关系
DRIVER_URL	选填，填入后VERSION参数均失效	填入实例可访问的 GPU 驱动安装包地址，填入即前三个版本参数均失效。示例值： http://mirrors.tencentyun.com/install/GPU/NVIDIA-Linux-x86_64-535.161.07.run
CUDA_URL	选填，填入后VERSION参数均失效	填入实例可访问的 CUDA 安装包地址，填入即前三个版本参数均失效。示例值： http://mirrors.tencentyun.com/install/GPU/cuda_12.4.0_550.54.14_linux.run
CUDNN_URL	选填，填入后VERSION参数均失效	填入实例可访问的 cuDNN 安装包地址，建议使用 tar.xz、tgz 格式安装包，填入即前三个版本参数均失效。示例值： http://mirrors.tencentyun.com/install/GPU/cudnn-linux-x86_64-8.9.7.29_cuda12-archive.tar.xz

4. 勾选需执行命令的 GPU 实例，您可以使用实例类型筛选出 GPU 实例类型：

选择执行实例 已选择1个实例

实例类型 GPU计算型GN7 多个关键字用竖线“|”分隔，多个过滤标签用回车键分隔 Q

<input checked="" type="checkbox"/>	ID/名称	状态	可用区	实例类型	主IPv4地址	实例计费模式	所属项目	标签 (key:value)	客户端状态
<input checked="" type="checkbox"/>	未命名	运行中	北京六区	GPU计算型GN7	(公) (内)	包年包月	默认项目		在线
<input type="checkbox"/>		已关机	北京五区	GPU计算型GN7	(公) (内)	包年包月			未安装

5. 单击执行命令即可。

#### 方式四：使用云市场镜像重装系统

操作步骤可参见 [使用预装 Tesla 驱动的云市场镜像](#)。

# 手动安装 Tesla 驱动-Linux

Last updated: 2024-11-11 17:52:43

## 操作场景

GPU 云服务器正常工作需提前安装正确的基础设施软件，对 NVIDIA 系列 GPU 而言，有以下两个层次的软件包需要安装：

- 驱动 GPU 工作的硬件驱动程序。
- 上层应用程序所需要的库。

若把 NVIDIA GPU 用作通用计算，则需安装 Tesla Driver + CUDA。本文介绍如何安装 Tesla Driver，如何安装 CUDA 请参见 [安装 CUDA 指引](#)。

### 说明：

NVIDIA GPU 实例的类型包括物理直通卡实例（整卡 GPU）和 vGPU 实例（非单颗 GPU，例如1/4GPU）。

- 物理直通的 GPU 可以使用 Tesla 驱动来支持计算场景，也可以使用 GRID 驱动支持渲染场景。
- vGPU 只能使用特定版本的 GRID 驱动支持不同场景。

## 操作步骤

### Linux 驱动安装

Linux 驱动安装采用 Shell 脚本安装方式，适用于任何 Linux 发行版，包括 TencentOS，Ubuntu 等。

NVIDIA Tesla GPU 的 Linux 驱动在安装过程中需要编译 kernel module，系统需提前安装 gcc 和编译 Linux Kernel Module 所依赖的包，例如 `kernel-devel-$(uname -r)` 等。本文提供 TencentOS 及 Ubuntu 系统操作步骤，请您按需选择：

#### TencentOS

1. Tencent 执行以下命令，检查当前系统中是否已安装 dkms。

```
rpm -qa | grep -i dkms
```

返回结果如下图，则表示已安装 dkms。

```
[root@VM-0-11-tencentos ~]# rpm -qa | grep -i dkms
dkms-3.2.0-1.t12.noarch
```

如未安装 dkms，则执行以下命令进行安装。

```
sudo yum install -y dkms
```

2. 登录 [NVIDIA 驱动下载](#) 或访问 <http://www.nvidia.com/Download/Find.aspx>。
3. 根据实例操作系统及 GPU 规格，选择操作系统和安装包。GPU 规格信息请参见 [实例类型](#)。
4. 单击 **SEARCH** 搜寻驱动，选择要下载的驱动版本。本文以 V100 为例，如下图所示：

### 注意：

操作系统（Operating System）选择 Linux 64-bit 即表示下载 shell 安装文件。如果选择具体的发行版，则下载的文件是对应的包安装文件。

## NVIDIA Driver Downloads

Official Advanced Driver Search | NVIDIA

Product Type:

Data Center / Tesla

Operating System:

Linux 64-bit

Product Series:

V-Series

CUDA Toolkit:

12.2

Product:

Tesla V100

Language:

English (US)

Recommended/Beta:

All ?

Search

Name	Version	Release Date	CUDA Toolkit
<a href="#">Data Center Driver for Linux x64</a>	535.161.08	March 18, 2024	12.2
<a href="#">Data Center Driver for Linux x64</a>	535.161.07	February 22, 2024	12.2
<a href="#">Data Center Driver for Linux x64</a>	535.154.05	January 16, 2024	12.2
<a href="#">Data Center Driver for Linux x64</a>	535.129.03	October 31, 2023	12.2
<a href="#">Data Center Driver for Linux x64</a>	535.104.12	September 25, 2023	12.2
<a href="#">Data Center Driver for Linux x64</a>	535.104.05	August 29, 2023	12.2
<a href="#">Data Center Driver for Linux x64</a>	535.86.10	July 31, 2023	12.2
<a href="#">Data Center Driver for Linux x64</a>	535.54.03	June 26, 2023	12.2

5. 选择特定的版本进入下载页面，单击 **DOWNLOAD**。如下图所示：

### Data Center Driver For Linux X64

Version: 535.161.07  
 Release Date: 2024.2.22  
 Operating System: CBL Mariner, Linux 64-bit  
 CUDA Toolkit: 12.2  
 Language: English (US)  
 File Size: 325.94 MB

Download

6. 如有填写个人信息的页面可选择直接跳过，当出现以下页面时，右键单击 **AGREE&DOWNLOAD** 并选择菜单中的复制链接地址。如下图所示：

### Download

By clicking the "Agree & Download" button below, you are confirming that you have read and agree to be bound by the [License For Customer Use of NVIDIA Software](#) for use of the driver. The driver will begin downloading immediately after clicking on the "Agree & Download" button below. NVIDIA recommends users update to the latest driver version. Please review [NVIDIA Product Security](#) for more information.

Agree & Download

Decline

7. 参见 [使用标准方式登录 Linux 实例（推荐）](#)，登录 GPU 实例。您也可以根据实际操作习惯，选择其他不同的登录方式：

- [使用远程登录软件登录 Linux 实例](#)
- [使用 SSH 登录 Linux 实例](#)

8. 使用 `wget` 命令，粘贴 [步骤6](#) 中已获取的链接地址，下载安装包。如下图所示：

```
[root@VM-0-11-tencentos ~]# wget https://us.download.nvidia.com/tesla/535.161.07/NVIDIA-Linux-x86_64-535.161.07.run
--2024-05-21 17:29:01-- https://us.download.nvidia.com/tesla/535.161.07/NVIDIA-Linux-x86_64-535.161.07.run
```

或者您可在本地系统下载 NVIDIA 安装包，再上传到 GPU 实例的服务器。

9. 执行以下命令，对安装包添加执行权限。例如，对文件 `NVIDIA-Linux-x86_64-535.161.07.run` 添加执行权限。

```
chmod +x NVIDIA-Linux-x86_64-535.161.07.run
```

10. 依次执行以下命令，检查当前系统中是否已安装 `gcc` 和 `kernel-devel` 包。

```
rpm -qa | grep kernel-devel
```

```
rpm -qa | grep gcc
```

返回结果如下，则表示已安装 `gcc` 和 `kernel-devel`。

```
[root@VM-0-11-tencentos ~]# rpm -qa | grep kernel-devel
mlnx-ofa_kernel-devel-5.4-OFED.5.4.3.6.8.1.rhel7u9.x86_64
kernel-devel-5.4.119-19.0009.40.tl2.x86_64
[root@VM-0-11-tencentos ~]#
[root@VM-0-11-tencentos ~]#
[root@VM-0-11-tencentos ~]# rpm -qa | grep gcc
devtoolset-8-gcc-c++-8.3.1-3.2.el7.x86_64
gcc-c++-4.8.5-44.tl2.1.x86_64
gcc-gfortran-4.8.5-44.tl2.1.x86_64
compat-gcc-44-4.4.7-8.tl2.x86_64
libgcc-4.8.5-44.tl2.1.x86_64
gcc-4.8.5-44.tl2.1.x86_64
gcc-objc-4.8.5-44.tl2.1.x86_64
compat-gcc-44-c++-4.4.7-8.tl2.x86_64
devtoolset-8-gcc-8.3.1-3.2.el7.x86_64
```

如未安装，则请执行以下命令进行安装。

```
sudo yum install -y gcc kernel-devel
```

#### ⚠ 注意:

如升级了 `kernel` 版本，则需要将 `kernel-devel` 升级至与 `kernel` 相同的版本。

11. 执行以下命令，运行驱动安装程序，并按提示进行后续操作。

```
./NVIDIA-Linux-x86_64-535.161.07.run --ui=none --disable-nouveau --no-install-libglvnd --dkms -s
```

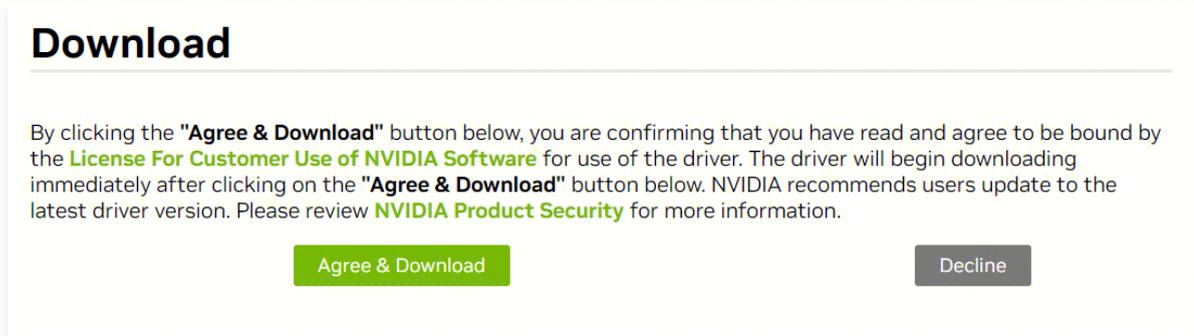
12. 安装完成后，执行以下命令进行验证。

```
nvidia-smi
```

如返回信息类似下图中的 GPU 信息，则说明驱动安装成功。



5. 如有填写个人信息的页面可选择直接跳过，当出现以下页面时，右键单击 **AGREE&DOWNLOAD** 并选择菜单中的复制链接地址。如下图所示：



6. 参见 [使用标准方式登录 Linux 实例（推荐）](#)，登录 GPU 实例。  
7. 使用 `wget` 命令，粘贴 [步骤5](#) 中复制的链接地址，下载安装包。如下图所示：

```
wget https://us.download.nvidia.com/tesla/535.161.07/NVIDIA-Linux-x86_64-535.161.07.run
```

8. 执行以下命令，修改安装包权限。请将命令中的 `xxxx` 替换为您实际的驱动版本号。

```
chmod +x NVIDIA-Linux-x86_64-xxxx.run
```

9. 由于 NVIDIA 驱动的安装需要依赖 `gcc` 和 `linux-kernel-headers`，请依次执行以下命令，检查当前系统中是否已安装 `gcc` 和 `kernel-devel` 包。

```
dpkg -l | grep -i gcc
```

```
dpkg -l | grep -i linux-headers
```

返回结果如下，则表示已安装 `gcc` 和 `kernel-devel`。

```
root@VM-0-11-ubuntu:/home/ubuntu# dpkg -l | grep -i gcc
ii gcc 4:9.3.0-1ubuntu2 amd64 GNU C compiler
ii gcc-10-base:amd64 10.3.0-1ubuntu1~20.04 amd64 GCC, the GNU Compiler Collection (base package)
ii gcc-9 9.3.0-17ubuntu1~20.04 amd64 GNU C compiler
ii gcc-9-base:amd64 9.3.0-17ubuntu1~20.04 amd64 GCC, the GNU Compiler Collection (base package)
ii libgcc1-0:amd64 10.3.0-1ubuntu1~20.04 amd64 GCC ccl plugin for GDB
ii libgcc-9-dev:amd64 9.3.0-17ubuntu1~20.04 amd64 GCC support library (development files)
ii libgcc-s1:amd64 10.3.0-1ubuntu1~20.04 amd64 GCC support library
ii libgomp1:amd64 10.3.0-1ubuntu1~20.04 amd64 GCC OpenMP (GOMP) support library
ii libquadmath0:amd64 10.3.0-1ubuntu1~20.04 amd64 GCC Quad-Precision Math Library
root@VM-0-11-ubuntu:/home/ubuntu#
root@VM-0-11-ubuntu:/home/ubuntu#
root@VM-0-11-ubuntu:/home/ubuntu#
root@VM-0-11-ubuntu:/home/ubuntu# dpkg -l | grep -i linux-headers
ii linux-headers-5.4.0-174 5.4.0-174.193 all Header files related to Linux kernel version 5.4.0
ii linux-headers-5.4.0-174-generic 5.4.0-174.193 amd64 Linux kernel headers for version 5.4.0 on 64 bit x86 SMP
```

若返回结果为空，则表明未安装，执行以下命令进行安装。

```
sudo apt-get install gcc linux-kernel-headers
```

10. 执行命令安装驱动程序，根据提示进行后续操作。请将命令中的 `xxx` 替换为您实际的驱动版本号。

```
sudo sh NVIDIA-Linux-x86_64-xxxx.run --ui=none --disable-nouveau --no-install-libglvnd --dkms -s
```

11. 安装完成后，执行以下命令进行验证。

```
nvidia-smi
```

如返回信息类似下图中的 GPU 信息，则说明驱动安装成功。



# 手动安装 Tesla 驱动-Windows

Last updated: 2024-11-01 16:48:02

## 操作场景

GPU 云服务器正常工作需提前安装正确的基础设施软件，对 NVIDIA 系列 GPU 而言，有以下两个层次的软件包需要安装：

- 驱动 GPU 工作的硬件驱动程序。
- 上层应用程序所需要的库。

若把 NVIDIA GPU 用作通用计算，则需安装 Tesla Driver + CUDA。本文介绍如何安装 Tesla Driver，如何安装 CUDA 请参见 [安装 CUDA 指引](#)。

### 说明：

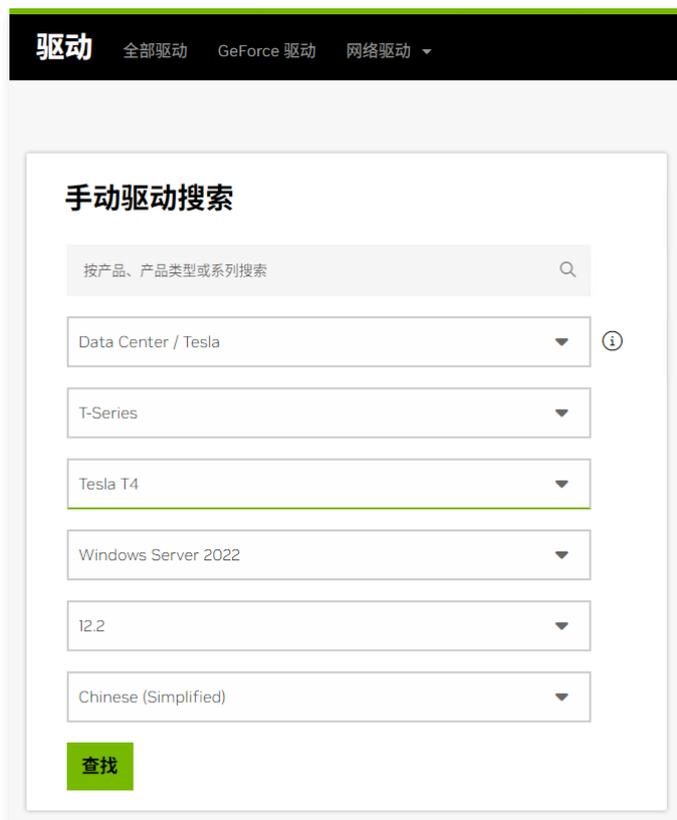
NVIDIA GPU 实例的类型包括物理直通卡实例（整卡 GPU）和 vGPU 实例（非单颗 GPU，例如1/4GPU）。

- 物理直通的 GPU 可以使用 Tesla 驱动来支持计算场景，也可以使用 GRID 驱动支持渲染场景。
- vGPU 只能使用特定版本的 GRID 驱动支持不同场景。

## 操作步骤

### Windows 驱动安装

1. 参见 [使用远程桌面连接登录 Windows 实例](#)，登录 GPU 实例。
2. 访问 [NVIDIA 驱动下载](#) 官网。
3. 根据实例操作系统及 GPU 型号，以及您需要的 CUDA 版本，选择需要的版本。GPU 型号信息请参见 [实例类型](#)。  
本文以 T4 GPU，选择支持 CUDA 12.2 的驱动安装包为例，如下图所示：



4. 按照指引下载安装包。
5. 打开下载驱动程序所在的文件夹，双击安装文件开始安装，按照界面提示安装驱动程序，并根据需要重启实例。



6. 安装完成后，打开 CMD 终端，输入 `nvidia-smi.exe`，确认新版本安装成功，且 GPU 可用。

```
管理员: 命令提示符
Microsoft Windows [版本 10.0.20348.2655]
(c) Microsoft Corporation. 保留所有权利。

C:\Users\Administrator>nvidia-smi.exe
Thu Oct 24 16:49:26 2024

+-----+
| NVIDIA-SMI 538.95                | Driver Version: 538.95          | CUDA Version: 12.2          |
+-----+-----+-----+-----+-----+-----+
| GPU  Name   Perf          TCC/WDDM    Bus-Id      Disp.A   Volatile Uncorr. ECC  |
| Fan  Temp   Perf          Pwr:Usage/Cap |      Memory-Usage   | GPU-Util  Compute M.  |
|=====+=====+=====+=====+=====+=====+
| 0   Tesla T4      P8             9W / 70W    | 00000000:00:06:0 Off |   0%      Default  |
| N/A   39C           | 0MiB / 15360MiB |           | N/A              |
+-----+-----+-----+-----+-----+-----+
| Processes:                        |
| GPU   GI   CI          PID    Type   Process name          GPU Memory  |
| ID    ID   ID                      |                     | Usage       |
+-----+-----+-----+-----+-----+-----+
| No running processes found      |
+-----+-----+-----+-----+-----+

C:\Users\Administrator>
```

# 使用预装 Tesla 驱动的云市场镜像

Last updated: 2024-11-01 16:48:02

## 操作场景

GPU 云服务器正常工作需提前安装正确的基础设施软件，对 NVIDIA 系列 GPU 而言，有以下两个层次的软件包需要安装：

- 驱动 GPU 工作的硬件驱动程序。
- 上层应用程序所需要的库。

为方便用户使用，云市场提供了预装 GPU 驱动和 AI 环境的镜像。在创建 GPU 实例或重装系统时，可以通过镜像市场选择相关镜像完成部署。

### 说明：

NVIDIA GPU 实例的类型包括物理直通卡实例（整卡 GPU）和 vGPU 实例（非单颗 GPU，例如1/4GPU）。

- 物理直通的 GPU 可以使用 Tesla 驱动来支持计算场景，也可以使用 GRID 驱动支持渲染场景。
- vGPU 只能使用特定版本的 GRID 驱动支持不同场景。

## 操作步骤

1. 在 [购买页](#) 或 [重装系统](#) 的过程中，选择 GPU 机型或实例。
2. 在镜像中选择云镜像市场，单击从云镜像市场选择。如下图所示：



3. 在弹出的镜像市场对话框的搜索框中，输入驱动并单击搜索。列出所有预装 GPU 驱动的镜像，如下图所示：



4. 根据实际需求，选择预装 GPU 驱动的镜像，单击**立即使用**。如下图所示：

本文以 Windows Server 2022 数据中心版 64位 中文版（预装538驱动）为例，您可根据实际需求进行选择。



5. 根据界面提示，完成实例配置。

# 安装 GRID 驱动

## 使用预装 GRID 驱动的镜像（推荐）

Last updated: 2024-11-01 16:48:02

### 操作场景

NVIDIA 系列 GPU 实例如用作 3D 图形渲染任务（高性能图形处理，视频编解码等），则需安装 GRID Driver 和配置 License Server，仅GNV4v、GNV4、GN7vw 实例簇提供预装 GRID 驱动的公共镜像，支持 DirectX 和 OpenGL 等图形 API。为方便用户使用，NVIDIA [渲染型 GPU 实例](#) 提供了预装 GRID 驱动的镜像，在创建 GPU 实例时，可以通过选择相关镜像完成 GRID 驱动部署。

### 操作步骤

1. 在 [购买页](#) 选择所需要的 GPU [渲染型实例机型](#)及相关配置。

#### 说明：

仅GNV4v、GNV4、GN7vw 实例簇提供预装 GRID 驱动的公共镜像，其中GNV4v、GN7vw均处于白名单开放，请您通过 [售前在线咨询](#) 进行实例购买权限开通。

2. 在镜像中选择公共镜像，单击[请选择镜像](#)如下图所示：



3. 在弹出的[选择镜像](#)下拉框的搜索框中，根据实际需求，单击您需要的镜像版本，如下图所示：

本文以 Ubuntu Server 22.04 LTS 64位 GRID16.2 为例，您可根据实际需求并参见 [渲染型实例支持镜像](#) 进行选择。



4. 根据界面提示，完成实例配置。

# 申请 License 并安装 GRID 驱动

Last updated: 2024-11-01 16:48:02

## 操作场景

腾讯云提供的NVIDIA [计算型 GPU 实例](#)搭载了 NVIDIA P4、T4、P40 以及 V100 等 GPU。Tesla 系列 GPU 可同时支持通用计算和图形图像处理，例如：

- 安装免费的 Tesla Driver 和 CUDA SDK，可用作深度学习、科学计算等通用计算场景。
- 安装 GRID Driver 并且向 NVIDIA 官网申请付费的 License，可开启 GPU 的 OpenGL 或 DirectX 图形加速能力。

使用 Tesla 系列 GPU 用作图形图像处理，需要向 NVIDIA 或其代理商购买对应的 License。此外，NVIDIA 也提供了90天试用 License 的申请。本文将介绍申请试用 License，并配置 License 服务器和安装 GRID driver。

### 说明：

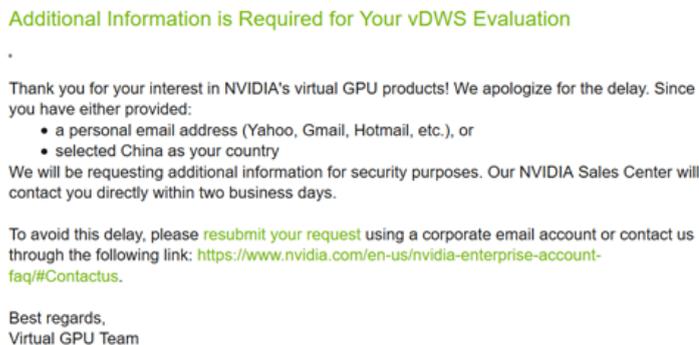
- 如在申请过程中遇到问题，则请与 NVIDIA 联系，本文仅供参考。
- 您也可直接使用腾讯云 [渲染型实例](#)，如 GN7vw、GNV4、GNV4v，免除 vDWS License 申请及搭建服务器步骤。

## 操作步骤

### 申请 License

1. 前往 [NVIDIA Enterprise Account](#)，注册账号并申请试用 License。具体步骤请参见 [注意事项](#)。
2. 成功提交申请后，将提示您于2天内设置的邮箱查收邮件。

如果您申请的地区是中国，或者使用了公共邮箱进行注册，您可能会收到如下邮件：



**Additional Information is Required for Your vDWS Evaluation**

.

Thank you for your interest in NVIDIA's virtual GPU products! We apologize for the delay. Since you have either provided:

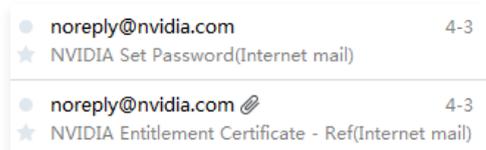
- a personal email address (Yahoo, Gmail, Hotmail, etc.), or
- selected China as your country

We will be requesting additional information for security purposes. Our NVIDIA Sales Center will contact you directly within two business days.

To avoid this delay, please [resubmit your request](https://www.nvidia.com/en-us/nvidia-enterprise-account-faq/#Contactus) using a corporate email account or contact us through the following link: <https://www.nvidia.com/en-us/nvidia-enterprise-account-faq/#Contactus>.

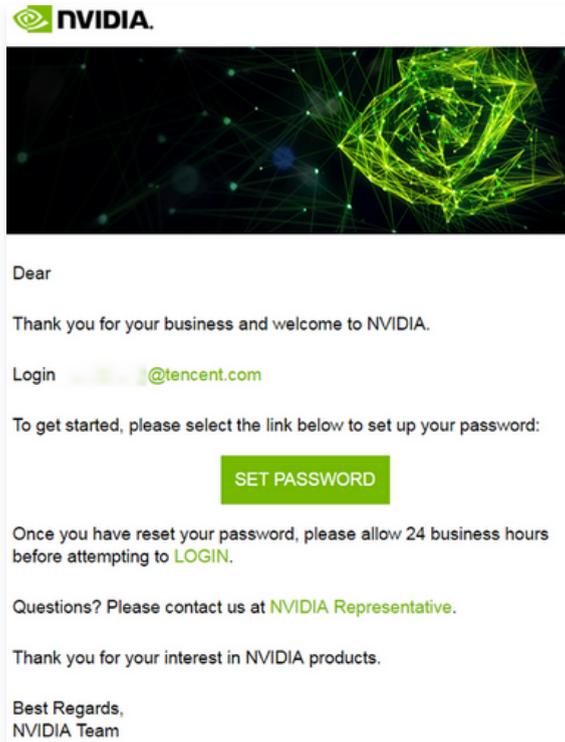
Best regards,  
Virtual GPU Team

3. 测试 License 账户需要2天的时间由后台进行账号审批，审批通过后，您会收到2封邮件。如下图所示：



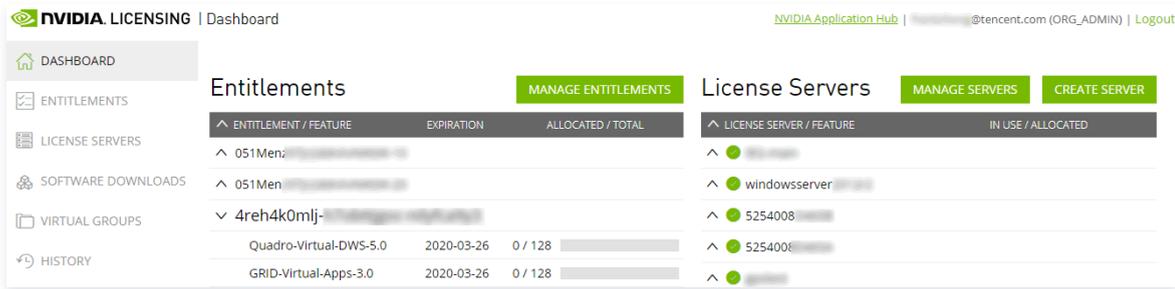
• noreply@nvidia.com	4-3
★ NVIDIA Set Password(Internet mail)	
• noreply@nvidia.com	4-3
★ NVIDIA Entitlement Certificate - Ref(Internet mail)	

4. 打开邮件 “NVIDIA Set Password”。如下图所示：



5. 选择 **SET PASSWORD** 并设置初始口令，设置完成即可单击 **LOGIN** 登录您的 NVIDIA 企业账号。也可前往 [NVIDIA 许可门户网站](#) 进行登录。

6. 登录成功后，进入 “Dashboard” 页面。如下图所示：



#### 说明：

vGPU 试用的全部资源均可由此网站获得，包括 vGPU 软件和 License Server 软件。

## 软件准备

1. 选择左侧导航栏中的 **SOFTWARE DOWNLOADS**，进入 **Software Downloads** 页面。
2. 选择下拉框中的 **Microsoft Windows Hyper-V 及 2016**，单击目标搜索结果所在行右侧的 **Download**。即使用最新的 NVIDIA Virtual GPU Software 版本，本文以 GRID 9.3 版本为例。如下图所示：



3. 创建一台普通云服务器，作为 License 服务器。详情请参见 [创建实例](#)。

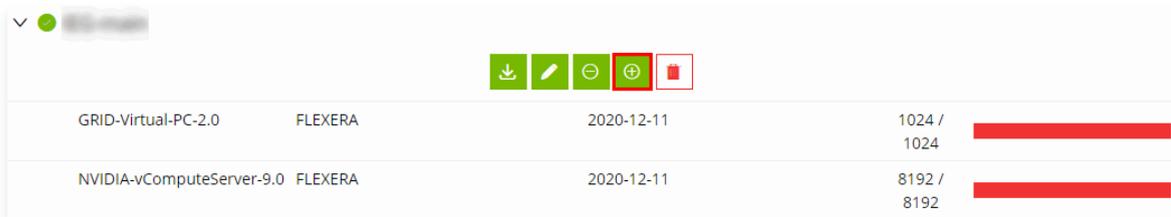
#### 说明：

云服务器操作系统建议使用 Windows Server 2012 R2 版本。

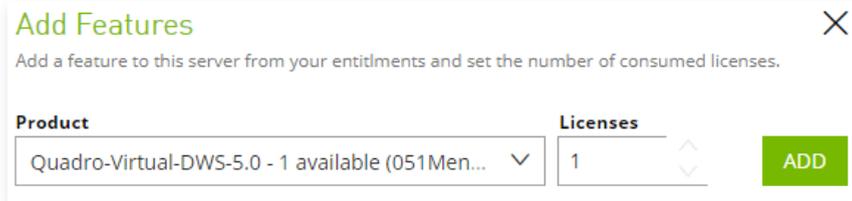
4. 选择页面右上角的 **Additional Software**，单击所需下载的 License Server 软件。本文以下载 **2019.11 64-bit License Manager for Windows** 为例，如下图所示：



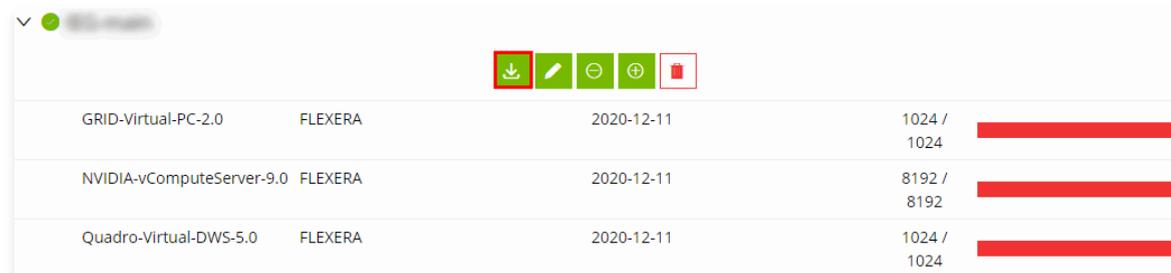
5. 将 License Server 软件安装至 **步骤3** 创建的云服务器，详情请参见 [License Server User Guide](#)。在完成安装后获取该 License 服务器的 MAC 地址。
6. 选择左侧导航栏中的 **LICENSE SERVERS**，进入 License Servers 页面。
7. 选择右上角的 **CREATE LICENSE SERVER**，在弹出的 **Create License Server** 窗口中填写相关信息注册新的 License Server。其中 MAC Address 请填写 **步骤5** 获取的 License 服务器 MAC 地址。
8. 成功创建后，在 **License Servers** 页面选择该项下的 **+**，添加多个特性。如下图所示：



9. 在弹出的 **Add Features** 页面，选择您试用的 License Feature，填写该 License 服务器的 License 数量，并单击 **ADD**。如下图所示：



10. 确认添加完毕后，单击 **ADD FEATURES**。
11. 成功添加后，可在 **License Servers** 页面查看授权该 License 服务器的状态，包含数量和 License 过期时间。
12. 选择 **↓**，下载用于该 License Server 的 License 授权文件。如下图所示：



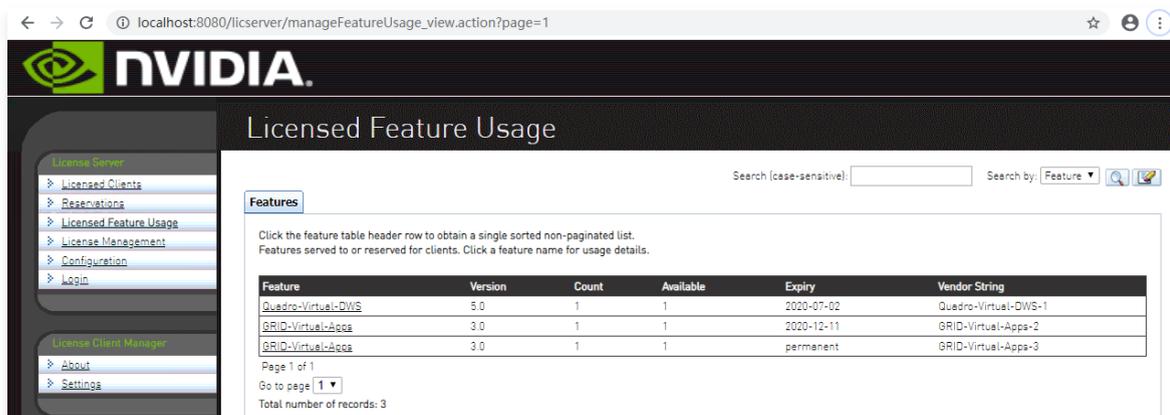
## 配置 License 服务器

### 说明：

授权文件下载完成后，需在24小时内通过 License 管理控制台导入该 License 文件。更多信息请参见 [NVIDIA官方文档](#)。

1. 使用 License 服务器访问 License 管理控制台：`http://localhost:8080/licserver`。

2. 选择左侧 License Server 栏中的 **License Management**，并导入 License 文件。
3. 选择 **Licensed Feature Usage**，查看授权数量。如下图所示：

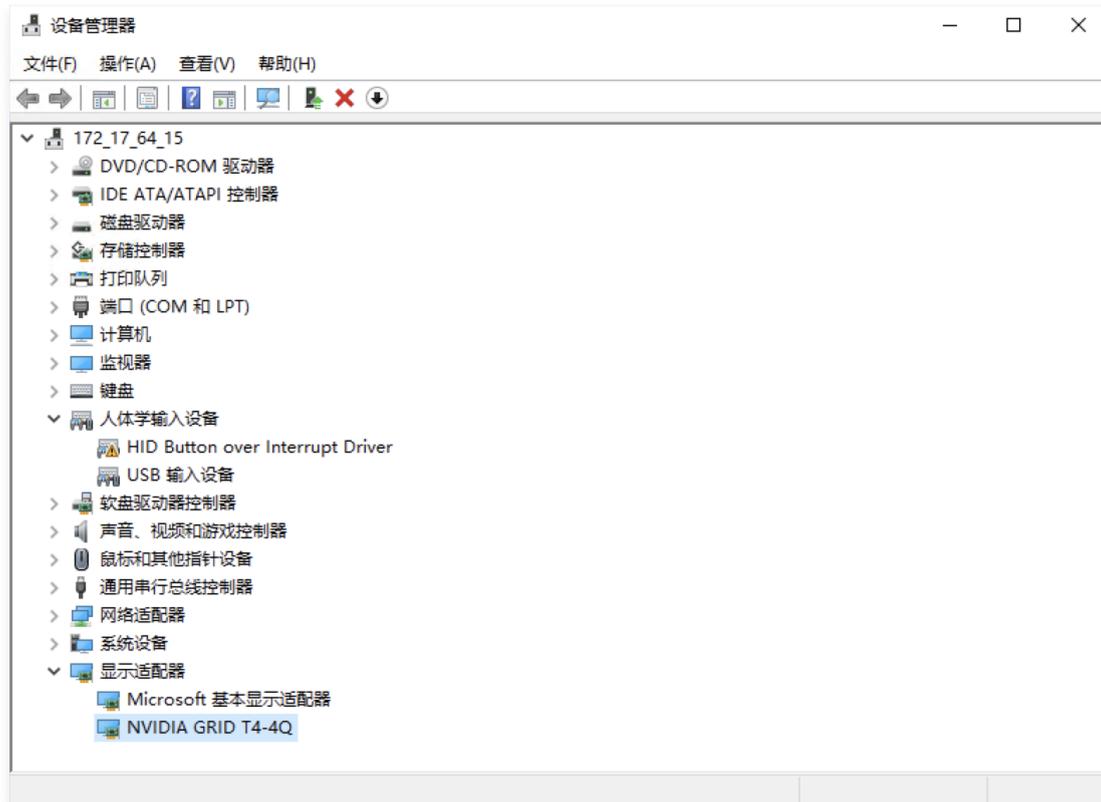


## 安装 GRID Driver

1. 购买并创建一台计算型 GPU 实例（GN6/GN6S/GN7/GN8/G10X）。详情请参见 [创建实例](#)。
2. 登录实例，详情请参见 [使用标准方式登录 Windows 实例（推荐）](#)。
3. 安装 GRID Driver 安装程序，即安装 NVIDIA vGPU for Windows 驱动程序。打开安装程序后按照界面提示完成安装，如下图所示：



完成安装后，在控制面板中查看驱动是否安装正确。如下图所示：



4. 使用远程桌面连接方式登录 GPU 实例，详情请参见 [使用远程桌面连接登录 Windows 实例](#)。

**说明：**

微软的远程桌面不支持使用 GPU 的 3D 硬件加速能力，如需使用则请安装第三方桌面协议软件，并通过对应客户端连接实例，使用 GPU 图形图像加速能力。

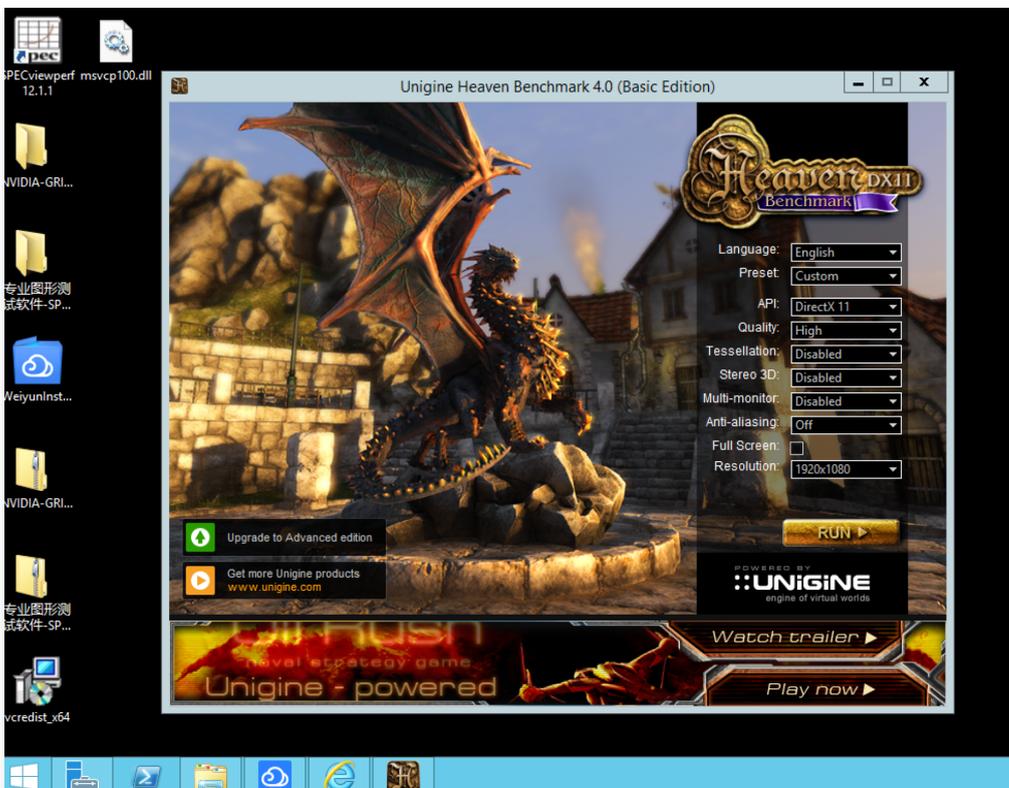
5. 在 GPU 实例中，右键单击桌面或者单击任务栏的右下角 ，打开 NVIDIA Control Panel。

6. 选择**管理许可证**，配置 License 服务器的 IP 地址和端口号，并确保 License 服务器的 IP 地址可以被访问，以及端口号已设置为开放状态。例如，将 License 服务器的 IP 地址配置为公网 IP 或者在同一个 VPC 内的内网地址，配置的 IP 地址可以被用户正常访问，且端口号已设置为开放状态。如下图

所示:



7. 完成以上配置，NVIDIA GPU 实例便即可运行图形图像处理程序。如下图所示：



# 安装 CUDA 驱动

Last updated: 2024-11-08 15:59:12

## 操作场景

CUDA ( Compute Unified Device Architecture ) 是显卡厂商 NVIDIA 推出的运算平台。CUDA™ 是一种由 NVIDIA 推出的通用并行计算架构，该架构使 GPU 能够解决复杂的计算问题。它包含了 CUDA 指令集架构 ( ISA ) 以及 GPU 内部的并行计算引擎。开发人员目前可使用 C 语言、C++ 及 FORTRAN 来为 CUDA™ 架构编写程序，所编写出的程序可在支持 CUDA™ 的处理器上以超高性能运行。

GPU 云服务器采用 NVIDIA 显卡，则需要安装 CUDA 开发运行环境。本文以目前最常用的 CUDA 12.4.0 为例，您可参考以下步骤进行安装。

## 操作步骤

### Linux 系统指引

1. 前往 [CUDA 下载](https://developer.nvidia.com/cuda-toolkit-archive) 页面或访问 <https://developer.nvidia.com/cuda-toolkit-archive>。
2. 选择 GPU 驱动对应的 [CUDA 版本](#)，本文以 CUDA Toolkit 12.4.0 为例。如下图所示：

# CUDA Toolkit Archive

Previous releases of the CUDA Toolkit, GPU Computing SDK, documentation and developer drivers can be found using the links below. Please select the below, and be sure to check [www.nvidia.com/drivers](http://www.nvidia.com/drivers) for more recent production drivers appropriate for your hardware configuration.

[Download Latest CUDA Toolkit](#)

[Learn More about CUDA Toolkit](#)

#### Latest Release

[CUDA Toolkit 12.6.2 \(October 2024\), Versioned Online Documentation](#)

#### Archived Releases

[CUDA Toolkit 12.6.1 \(August 2024\), Versioned Online Documentation](#)

[CUDA Toolkit 12.6.0 \(August 2024\), Versioned Online Documentation](#)

[CUDA Toolkit 12.5.1 \(July 2024\), Versioned Online Documentation](#)

[CUDA Toolkit 12.5.0 \(May 2024\), Versioned Online Documentation](#)

[CUDA Toolkit 12.4.1 \(April 2024\), Versioned Online Documentation](#)

[CUDA Toolkit 12.4.0 \(March 2024\), Versioned Online Documentation](#)

[CUDA Toolkit 12.3.2 \(January 2024\), Versioned Online Documentation](#)

[CUDA Toolkit 12.3.1 \(November 2023\), Versioned Online Documentation](#)

[CUDA Toolkit 12.3.0 \(October 2023\), Versioned Online Documentation](#)

[CUDA Toolkit 12.2.2 \(August 2023\), Versioned Online Documentation](#)

[CUDA Toolkit 12.2.1 \(July 2023\), Versioned Online Documentation](#)

[CUDA Toolkit 12.2.0 \(June 2023\), Versioned Online Documentation](#)

[CUDA Toolkit 12.1.1 \(April 2023\), Versioned Online Documentation](#)

[CUDA Toolkit 12.1.0 \(February 2023\), Versioned Online Documentation](#)

[CUDA Toolkit 12.0.1 \(January 2023\), Versioned Online Documentation](#)

[CUDA Toolkit 12.0.0 \(December 2022\), Versioned Online Documentation](#)

[CUDA Toolkit 11.8.0 \(October 2022\), Versioned Online Documentation](#)

[CUDA Toolkit 11.7.1 \(August 2022\), Versioned Online Documentation](#)

3. 按照页面上的提示，依次选择操作系统和安装包，并复制 `wget` 下载命令和地址。本文选择方式如下图所示：

#### ⚠ 注意

Installer Type: 推荐选择 `runfile (local)`。

- `network`: 网络安装包，安装包较小，需要在主机内联网下载实际的安装包。
- `local`: 本地安装包。安装包较大，包含每一个下载安装组件的安装包。

# CUDA Toolkit 12.4 Downloads

## Select Target Platform

Click on the green buttons that describe your target platform. Only supported platforms will be shown. By downloading and using the software, you agree to fully comply with the terms and conditions of the [CUDA EULA](#).

Operating System	Linux	Windows							
Architecture	x86_64	ppc64le	arm64-sbsa	aarch64-jetson					
Distribution	CentOS	Debian	Fedora	KylinOS	OpenSUSE	RHEL	Rocky	SLES	Ubuntu
	WSL-Ubuntu								
Version	7	8	9						
Installer Type	rpm (local)	rpm (network)	runfile (local)						

## Download Installer for Linux RHEL 9 x86\_64

The base installer is available for download below.

### > Base Installer

Installation Instructions:

```
$ wget https://developer.download.nvidia.com/compute/cuda/12.4.0/local_installers/cuda_12.4.0_550.54.14_linux.run
$ sudo sh cuda_12.4.0_550.54.14_linux.run
```

4. 参见 [使用标准方式登录 Linux 实例（推荐）](#)，登录 GPU 实例。您也可以根据实际操作习惯，选择其他不同的登录方式：

- [使用远程登录软件登录 Linux 实例](#)
- [使用 SSH 登录 Linux 实例](#)

5. 使用 `wget` 命令，粘贴步骤3中已获取的链接，下载安装包。如下图所示：

```
[root@VM-0-13-tencentos test]# wget https://developer.download.nvidia.com/compute/cuda/12.4.0/local_installers/cuda_12.4.0_550.54.14_linux.run
--2024-11-07 09:00:13-- https://developer.download.nvidia.com/compute/cuda/12.4.0/local_installers/cuda_12.4.0_550.54.14_linux.run
Resolving developer.download.nvidia.com (developer.download.nvidia.com)... 152.199.39.144
Connecting to developer.download.nvidia.com (developer.download.nvidia.com)|152.199.39.144|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 4454353277 (4.1G) [application/octet-stream]
Saving to: 'cuda_12.4.0_550.54.14_linux.run'

cuda_12.4.0_550.54.14_linux.run      5%[====>] 236.51M  11.3MB/s  eta 5m 58s
```

或者您可在本地系统下载 CUDA 安装包，再上传到 GPU 实例的服务器。下载或上传完成可使用 [11](#) 查看当前目录文件。

```
[root@VM-0-13-tencentos test]# ll
total 4349960
-rw-r--r-- 1 root root 4454353277 Feb 29 2024 cuda_12.4.0_550.54.14_linux.run
[root@VM-0-13-tencentos test]#
```

6. 依次执行以下命令，对安装包添加执行权限，请在 `root` 用户下操作。例如，对文件名为 `cuda_12.4.0_550.54.14_linux.run` 添加执行权限。

```
chmod +x cuda_12.4.0_550.54.14_linux.run
```

```
./cuda_12.4.0_550.54.14_linux.run --toolkit --samples --silent --override
```

如下图所示：

```
[root@VM-0-13-tencentos test]#
[root@VM-0-13-tencentos test]# chmod +x cuda_12.4.0_550.54.14_linux.run
[root@VM-0-13-tencentos test]# ./cuda_12.4.0_550.54.14_linux.run --toolkit --samples --silent --override
[root@VM-0-13-tencentos test]#
```

#### ⚠ 注意：

安装 CUDA 过程中，Driver 选项表示是否安装 GPU 驱动，如果您已经安装了 GPU 驱动，请取消勾选。  
GPU 驱动安装，请使用驱动安装包，不建议使用 CUDA 安装包来安装 GPU 驱动。

```
CUDA Installer
- [ ] Driver
  [ ] 550.54.14
+ [X] CUDA Toolkit 12.4
  [X] CUDA Demo Suite 12.4
  [X] CUDA Documentation 12.4
- [ ] Kernel Objects
  [ ] nvidia-fs
Options
Install
```

7. 执行以下命令，验证 CUDA 安装是否成功。

```
/usr/local/cuda-12.4/bin/nvcc --version
```

若展示出 CUDA 版本，则安装成功。

```
[root@VM-0-13-tencentos test]# /usr/local/cuda-12.4/bin/nvcc --version
nvcc: NVIDIA (R) Cuda compiler driver
Copyright (c) 2005-2024 NVIDIA Corporation
Built on Tue_Feb_27_16:19:38_PST_2024
Cuda compilation tools, release 12.4, V12.4.99
Build cuda_12.4.r12.4/compiler.33961263_0
[root@VM-0-13-tencentos test]#
```

## Windows 系统指引

1. 参见 [使用远程桌面连接登录 Windows 实例](#)，登录 GPU 实例。
2. 访问 [CUDA 官网](#)。
3. 选择对应的 CUDA 版本，本文以 CUDA Toolkit 12.4.0 为例。如下图所示：

# CUDA Toolkit Archive

Previous releases of the CUDA Toolkit, GPU Computing SDK, documentation and developer drivers can be found using the links below. Please select the below, and be sure to check [www.nvidia.com/drivers](http://www.nvidia.com/drivers) for more recent production drivers appropriate for your hardware configuration.

[Download Latest CUDA Toolkit](#)

[Learn More about CUDA Toolkit](#)

## Latest Release

[CUDA Toolkit 12.6.2 \(October 2024\), Versioned Online Documentation](#)

## Archived Releases

[CUDA Toolkit 12.6.1 \(August 2024\), Versioned Online Documentation](#)

[CUDA Toolkit 12.6.0 \(August 2024\), Versioned Online Documentation](#)

[CUDA Toolkit 12.5.1 \(July 2024\), Versioned Online Documentation](#)

[CUDA Toolkit 12.5.0 \(May 2024\), Versioned Online Documentation](#)

[CUDA Toolkit 12.4.1 \(April 2024\), Versioned Online Documentation](#)

[CUDA Toolkit 12.4.0 \(March 2024\), Versioned Online Documentation](#)

[CUDA Toolkit 12.3.2 \(January 2024\), Versioned Online Documentation](#)

[CUDA Toolkit 12.3.1 \(November 2023\), Versioned Online Documentation](#)

[CUDA Toolkit 12.3.0 \(October 2023\), Versioned Online Documentation](#)

[CUDA Toolkit 12.2.2 \(August 2023\), Versioned Online Documentation](#)

[CUDA Toolkit 12.2.1 \(July 2023\), Versioned Online Documentation](#)

[CUDA Toolkit 12.2.0 \(June 2023\), Versioned Online Documentation](#)

[CUDA Toolkit 12.1.1 \(April 2023\), Versioned Online Documentation](#)

[CUDA Toolkit 12.1.0 \(February 2023\), Versioned Online Documentation](#)

[CUDA Toolkit 12.0.1 \(January 2023\), Versioned Online Documentation](#)

[CUDA Toolkit 12.0.0 \(December 2022\), Versioned Online Documentation](#)

[CUDA Toolkit 11.8.0 \(October 2022\), Versioned Online Documentation](#)

[CUDA Toolkit 11.7.1 \(August 2022\), Versioned Online Documentation](#)

4. 按照页面上的提示，依次选择操作系统和安装包。本文选择方式如下图所示：

# CUDA Toolkit 12.4 Downloads

## Select Target Platform

Click on the green buttons that describe your target platform. Only supported platforms will be shown. By downloading and using the software, you agree to fully comply with the terms and conditions of the [CUDA EULA](#).

<b>Operating System</b>	<input type="button" value="Linux"/>	<input type="button" value="Windows"/>	
<b>Architecture</b>	<input type="button" value="x86_64"/>		
<b>Version</b>	<input type="button" value="10"/>	<input type="button" value="11"/>	<input type="button" value="Server 2022"/>
<b>Installer Type</b>	<input type="button" value="exe (local)"/>	<input type="button" value="exe (network)"/>	

## Download Installer for Windows Server2022 x86\_64

The base installer is available for download below.

> Base Installer

Download (3.0 GB)

Installation Instructions:

1. Double click cuda\_12.4.0\_551.61\_windows.exe
2. Follow on-screen prompts

Additional installation options are detailed [here](#).

5. 打开下载 CUDA 所在的文件夹，双击安装文件开始安装。



6. 若最后出现完成对话框，则表示安装成功。如下图所示：



# 卸载 NVIDIA 驱动

## 卸载 Tesla 驱动

Last updated: 2024-11-21 15:27:22

### 操作场景

GPU 云服务器正常工作需安装正确的驱动。如果您需要升级 Tesla 驱动或者由于某种原因需要卸载当前驱动，务必确保新安装的驱动与您的实例规格和操作系统、业务需求相匹配，避免因驱动不匹配而导致业务无法正常进行的风险。

#### 说明：

NVIDIA GPU 实例需要安装驱动后才可以正常使用，根据业务属于计算或渲染场景，需选择适用的驱动类型安装。

- 部署免费的 Tesla Driver 和 CUDA SDK，可用作深度学习、科学计算等通用计算场景。
- 部署 GRID Driver 并且使用付费的 License，可开启 GPU 的 OpenGL 或 DirectX 图形加速能力。

### 操作步骤

Ubuntu/TencentOS

#### 方式一、使用 TAT 自动化助手卸载驱动

- 登录云服务器控制台，选择左侧导航栏中的 自动化助手 > 公共命令。
- 在公共命令页面上方选择实例所在地域，并单击**卸载 GPU 驱动-Linux 系统**模块左下角的**执行命令**。如下图所示：



- 进入执行命令页面，勾选需执行命令的 GPU 实例，您可以使用**实例类型**筛选出 GPU 实例类型：



- 单击**执行命令**即可。

#### 方式二：手动卸载驱动

- 命令行执行以下命令，切换到 root 用户。

```
sudo su
```

2. 输入以下命令，卸载驱动。

```
nvidia-installer --uninstall -s
```

如下图所示：

```
root@VM-2-124-ubuntu:/home/ubuntu# nvidia-installer --uninstall -s
WARNING: Your driver installation has been altered since it was initially installed; this may happen, for example, if you have since installed
NVIDIA driver through a mechanism other than nvidia-installer (such as your distribution's native package management system). nvidia-installer will attempt to
uninstall it if possible. Please see the file '/var/log/nvidia-uninstall.log' for details.
```

3. 执行 `lsmod | grep nvidia` 检查驱动是否卸载干净。如果结果为空，则已经卸载干净，如下图所示：

```
root@VM-2-124-ubuntu:/home/ubuntu# sudo lsmod | grep nvidia
root@VM-2-124-ubuntu:/home/ubuntu#
```

如果出现部分文件残留，可重复执行下列命令：

```
fuser -k /dev/nvidia*; rmdir nvidia_modeset; rmdir nvidia_drm; rmdir nvidia_uvm; rmdir nvidia;
```

#### 说明：

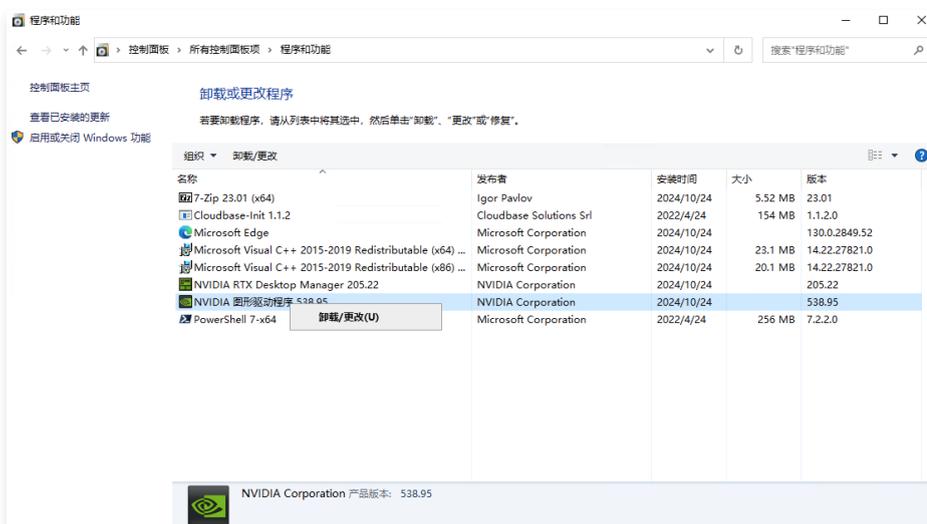
注意使用 `sudo su` 切换到 `root` 权限执行该命令，该命令可重复多次执行，无副作用。

4. 再次执行 `lsmod | grep nvidia` 检查驱动是否卸载干净，如果为空，则已卸载成功。

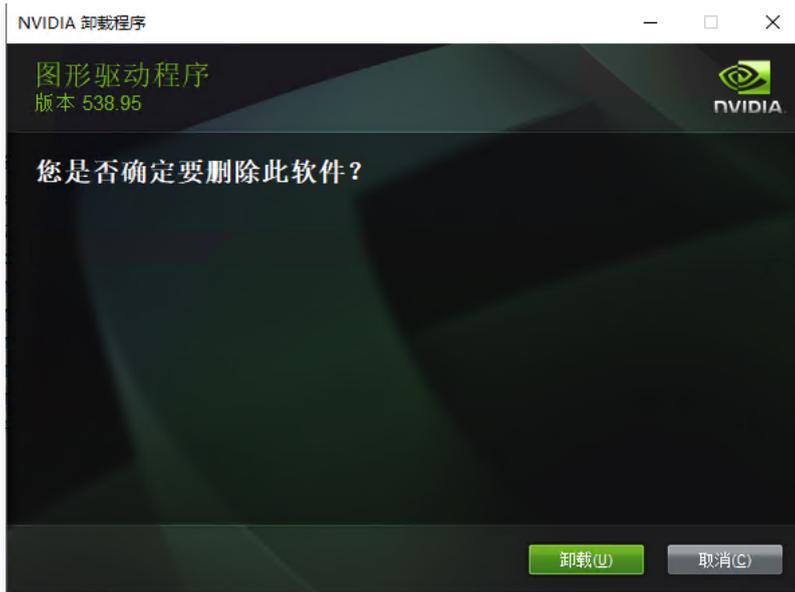
## Windows

1. 远程登录 GPU 实例，参见 [使用远程桌面连接登录 Windows 实例](#)。

2. 单击左下角 ，进入控制面板 > 程序与功能，右键选择卸载。



3. 按照界面指引执行下一步。



4. 最后一步，提示重启机器，单击**马上重新启动**。



5. 待机器重启后，进入**控制面板 > 程序与功能**，确认没有 Tesla 驱动即已经卸载干净。

# 卸载 GRID 驱动

Last updated: 2024-12-24 16:24:12

## 操作场景

GPU 云服务器正常工作需安装正确的驱动。如果您需要升级 GRID 驱动或者由于某种原因需要卸载当前驱动，务必确保新安装的驱动与您的实例规格和操作系统、业务需求相匹配，避免因驱动不匹配而导致业务无法正常进行的风险。

### 说明：

NVIDIA 渲染型 GPU 实例的类型包括物理直通卡实例（整卡 GPU）和 vGPU 实例（非单颗 GPU，例如1/4GPU）。

- 物理直通的 GPU 可以使用 GRID 驱动支持渲染场景，也可以使用 Tesla 驱动来支持计算场景。
- vGPU 只能使用特定版本的 GRID 驱动支持不同场景。

## 操作步骤

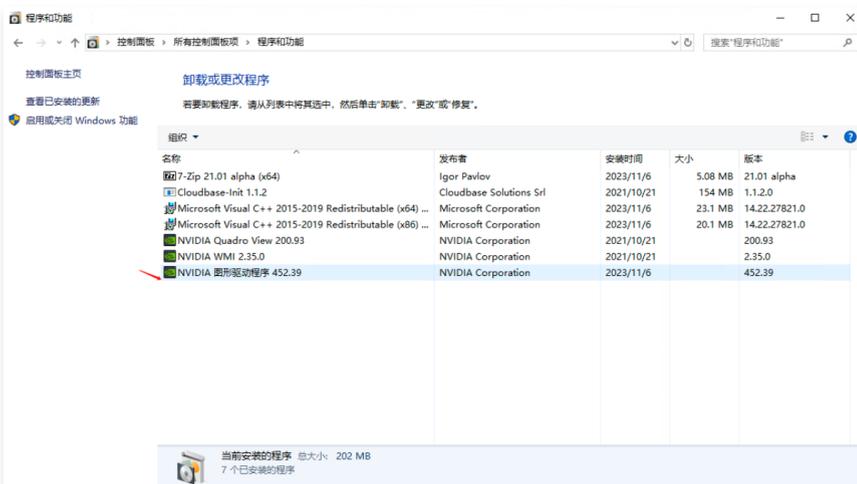
### Windows

以腾讯云操作系统 Windows Server 2019 和 Windows Server 2022 为例。

#### 1. 打开控制面板。



#### 2. 在控制面板中，选择程序和功能 > NVIDIA 图形驱动程序，右键单击需要卸载的 GRID 驱动，然后单击卸载。



#### 3. 卸载完成后，依照指引进行实例重启，重启完成后，则 GRID 驱动即卸载成功。

### TencentOS Serve/CentOS

以腾讯云操作系统 TencentOS Server 3.1 (TK4) 为例。

1. 以 root 权限执行以下命令，卸载GRID驱动并重启。

```
nvdi-a-installer --uninstall -s  
dracut --force && reboot
```

2. 重启登录实例后，执行以下命令，确认无 nvidia 内核模块。

```
lsmod | grep nvidia
```

显示如下表明卸载干净。

```
[root@VM-0-16-centos ~]# lsmod | grep nvidia  
[root@VM-0-16-centos ~]#
```

## Ubuntu

以腾讯云操作系统 TencentOS Server 3.1 (TK4) 为例。

1. 登录实例执行以下命令，卸载GRID驱动并重启。

```
sudo nvidia-installer --uninstall -s  
sudo update-initramfs -u && reboot
```

2. 重启登录实例后，执行命令以下命令，确认无 nvidia 内核模块。

```
lsmod | grep nvidia
```

显示如下表明卸载干净。

```
[root@VM-0-16-centos ~]# lsmod | grep nvidia  
[root@VM-0-16-centos ~]#
```

# 升级 NVIDIA 驱动

## 升级 Tesla 驱动

Last updated: 2024-11-01 16:48:02

### 操作场景

若您安装了与业务需求不匹配的 Tesla 或 GRID 驱动版本，或因错误的驱动类型或版本导致 GPU 实例无法正常运行，可通过卸载当前驱动并安装正确版本的方式来完成 Tesla 或 GRID 驱动的升级操作。

#### 说明：

NVIDIA GPU 实例的类型包括物理直通卡实例（整卡 GPU）和 vGPU 实例（非单颗 GPU，例如1/4GPU）。只有物理直通的 GPU 可以使用 Tesla 驱动。

### 操作步骤

#### 1. 卸载当前 GPU 驱动。

- Tesla 驱动卸载步骤请参见 [卸载 Tesla 驱动](#)。
- GRID 驱动卸载步骤请参见 [卸载 GRID 驱动](#)。

#### 注意：

请确保之前安装的驱动已成功卸载后再升级，否则会影响新驱动使用。

#### 2. 安装新的 Tesla 驱动。

- Linux 操作系统安装步骤请参见 [创建实例后快速安装 Tesla 驱动-Linux](#)。
- Windows 操作系统安装步骤请参见 [手动安装 Tesla 驱动-Windows](#)。



# 使用 GPU 监控和告警

Last updated: 2025-02-08 14:35:52

监控与告警是保证 GPU 云服务器高可靠性、高可用性和高性能的重要部分。创建 GPU 云服务器时，默认免费开通腾讯云可观测平台。您可以通过 [云服务器控制台](#) 查看监控指标，详细说明请参见 [云服务器监控内容](#)。NVIDIA GPU 系列实例另外提供了监控 GPU 使用率，显存使用量，功耗以及温度等参数的能力。

您也可以在 [腾讯云可观测平台](#) 分析监控指标和实施告警，更多详细内容可参见 [腾讯云可观测平台告警管理](#)。

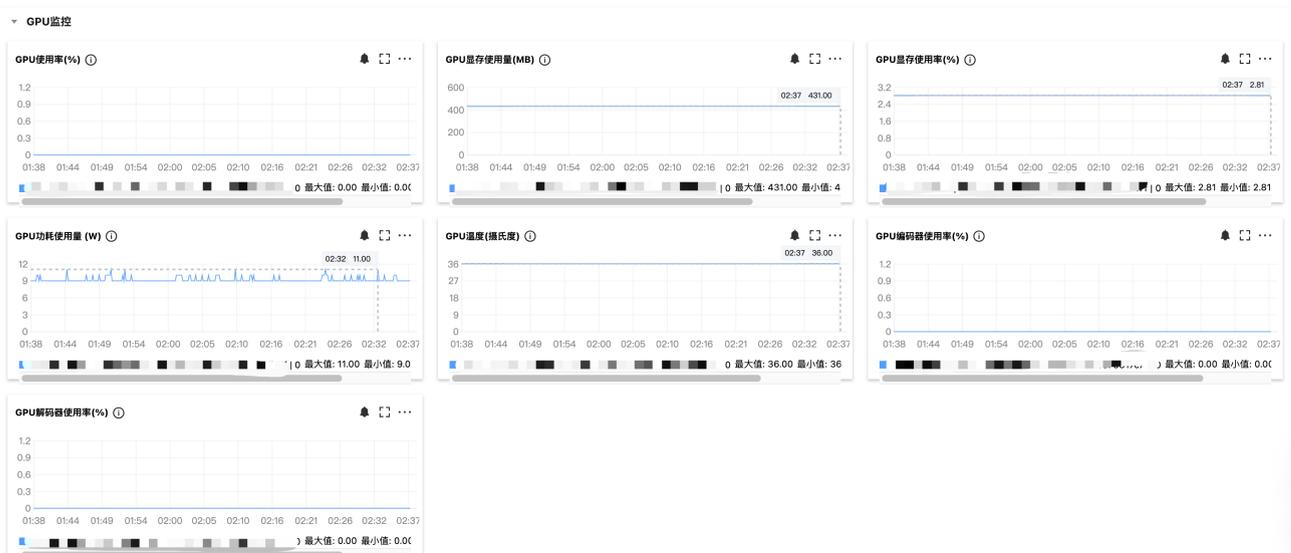
## GPU 监控工作条件

GPU 监控是通过在 GPU 云服务上部署安装相关 [GPU 驱动](#) 和 [云服务器监控组件](#) 来实现的，使用不同的镜像需要不同的处理方式：

- 使用公共镜像：公共镜像默认包含云服务器监控组件，只需安装 GPU 驱动。
- 使用云镜像市场 GPU 驱动预装镜像：无需任何安装。
- 使用导入镜像：需手动安装云服务器监控组件和 GPU 驱动。

## 在控制台查看 GPU 工作参数

单击 GPU 列表中的  监控图标，访问 [控制台](#) GPU 实例的监控页面，查看 GPU 监控，将鼠标移动到指标曲线上将显示对应 GPU 设备的 BDF（Bus、Device、Function，设备唯一地址）和监控数据。如下图所示：



参数说明：

指标名称	含义	单位	维度
GPU 使用率	评估负载所消耗的计算能力，非空闲状态百分比	%	per-GPU
GPU 显存使用量	评估负载对显存占用	MB	per-GPU
GPU 显存使用率	评估负载对显存占用百分比	%	per-GPU
GPU 功耗使用量	评估 GPU 耗电情况	W	per-GPU
GPU 温度	评估 GPU 散热状态	摄氏度	per-GPU
GPU 编码器使用率	评估编码器使用百分比	%	per-GPU
GPU 解码器使用率	评估解码器使用百分比	%	per-GPU

## 在腾讯云可观测平台查看 GPU 工作参数

[腾讯云可观测平台](#) 支持分析更丰富的 GPU 监控指标。

1. 登录 [腾讯云可观测平台](#)，左侧导航栏中选择 **Dashboard**，进入 Dashboard 列表页。

2. 在 Dashboard 列表中，单击 [新建 Dashboard](#)，在新的 Dashboard 选择新建图表，在指标处选择 GPU / 云服务器 / GPU 监控，单击您关注的指标，自定义监控面板进行多实例展示，如下图所示：



参数详情可参见 [腾讯云可观测平台 GPU 云服务器监控指标](#)，提供以下监控指标：

指标英文名	指标中文名	指标说明	单位	维度
Gpumemusage	GPU 显存使用率	GPU 显存使用率	%	per-GPU
GpuMemUsed	GPU 显存使用量	评估负载对显存占用	MB	per-GPU
Gpupowdraw	GPU 功耗使用量	GPU 功耗使用量	W	per-GPU
Gpupowusage	GPU 功耗使用率	GPU 功耗使用率	%	per-GPU
Gputemp	GPU 温度	评估 GPU 散热状态	摄氏度	per-GPU
Gpuutil	GPU 使用率	评估负载所消耗的计算能力，非空闲状态百分比	%	per-GPU
GpuEncUtil	GPU 编码器使用率	GPU 编码器使用率	%	per-GPU
GpuDecUtil	GPU 解码器使用率	GPU 解码器使用率	%	per-GPU

#### 说明：

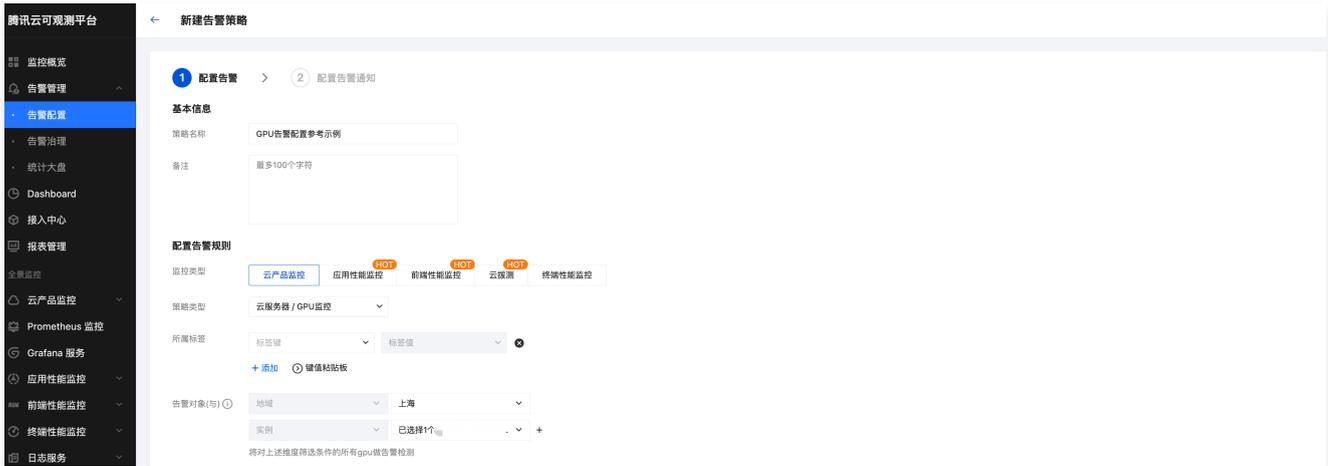
如果没有监控数据，考虑以下几种可能性：

- 只支持 NVIDIA GPU 实例。
- 没有安装 GPU 驱动或监控组件。
- 其他原因分析可参见 [云服务器无监控数据](#)。

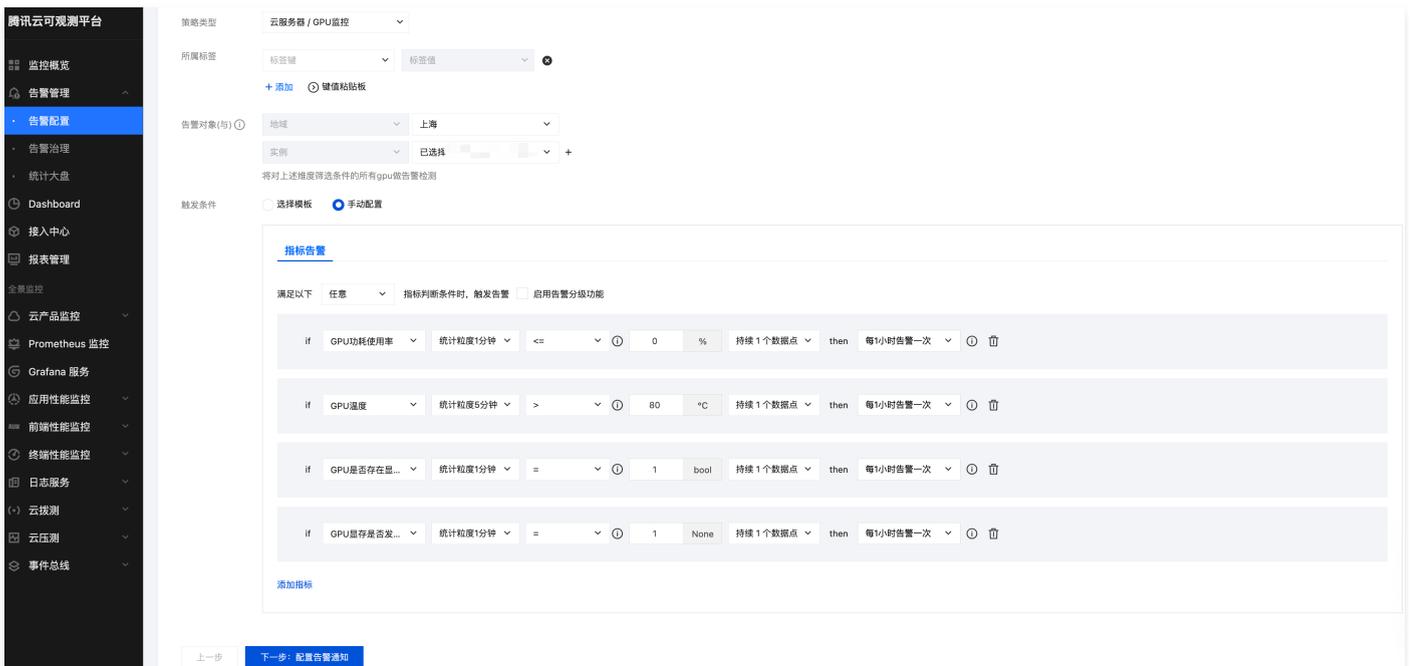
## GPU 云服务器监控指标告警配置

[腾讯云可观测平台](#) 支持分析更丰富的 GPU 监控指标。

- 登录 [腾讯云可观测平台](#)，在左侧导航栏中，选择告警管理 > 告警配置。
- 单击 [新建告警策略](#)，在监控类型选择云产品监控，策略类型中选择云服务器 / GPU 监控，选择您希望接收告警的 GPU 实例对象，触发条件选择手动配置。



3. GPU 云服务器监控支持以下指标告警：GPU 内存使用率、GPU 功耗使用率、GPU 使用率、GPU 温度、GPU 是否存在显存页需隔离、GPU 显存是否发生 UCE 等。您可以参考下图进行配置告警。告警通知的配置可参见 [新建通知模板](#)，支持通过多种渠道进行通知。



常用告警指标参考如下：

指标名称	建议告警阈值	描述	处理建议
GPU 功耗使用率	<=0	功耗小于0时可能功率出现 Unknown Error 了，会影响 GPU 的正常使用。	执行 <code>nvidia-smi</code> 命令查看 GPU 的功率是否有 ERR 或 <code>nvidia-smi -i &lt;target gpu&gt; -q  grep "Power Draw"</code> 是否为 Unknown Error，若存在该现象则尝试重启机器恢复及更新驱动观察。若重启无法恢复 <a href="#">提交工单</a> 联系腾讯云支持。
GPU 温度	持续5分钟 >80	当 GPU 温度过高时可能会导致 GPU SlowDown，影响业务性能。	可能负载过高导致 GPU 温度过高，可尝试重启实例恢复，若无法恢复 <a href="#">提交工单</a> 联系腾讯云支持。
GPU 是否存在显存页需隔离	=1	安培以下架构 GPU 出现了 ECC ERROR，应用进程被 kill，GPU卡处于pending 状态。	执行 <code>nvidia-smi -i &lt;target gpu&gt; -q -d PAGE_RETIREMENT</code> 命令查看是否有 GPU 卡处于 pending 状态，重置 GPU 卡或重启实例恢复。若重启无法恢复 <a href="#">提交工单</a> 联系腾讯云支持。
GPU 显存是否发生 UCE	=1	安培及以上架构 GPU 出现了 ECC ERROR，应用进程被	执行 <code>nvidia-smi -i &lt;target gpu&gt; -q -d ROW_REMAPPER</code> 命令查看是否有 GPU 卡处于 Pending 状态，重置 GPU 卡或重启实例恢

		kill, GPU卡处于pending 状态。	复。若重启无法恢复 <a href="#">提交工单</a> 联系腾讯云支持。
GPU 内存使用率	仅保持观察	-	评估负载对显存占用。
GPU 使用率	仅保持观察	-	评估负载对 GPU 流处理器占用。