

# GPU 云服务器

# 操作指南





#### 【版权声明】

©2013-2025 腾讯云版权所有

本文档(含所有文字、数据、图片等内容)完整的著作权归腾讯云计算(北京)有限责任公司单独所有,未经腾讯云事先明确书面许可,任何主体不得以任何 形式复制、修改、使用、抄袭、传播本文档全部或部分内容。前述行为构成对腾讯云著作权的侵犯,腾讯云将依法采取措施追究法律责任。

【商标声明】

## 🔗 腾讯云

及其它腾讯云服务相关的商标均为腾讯云计算(北京)有限责任公司及其关联公司所有。本文档涉及的第三方主体的商标,依法由权利人所有。未经腾讯云及 有关权利人书面许可,任何主体不得以任何方式对前述商标进行使用、复制、修改、传播、抄录等行为,否则将构成对腾讯云及有关权利人商标权的侵犯,腾 讯云将依法采取措施追究法律责任。

【服务声明】

本文档意在向您介绍腾讯云全部或部分产品、服务的当时的相关概况,部分产品、服务的内容可能不时有所调整。 您所购买的腾讯云产品、服务的种类、服务标准等应由您与腾讯云之间的商业合同约定,除非双方另有约定,否则,腾讯云对本文档内容不做任何明示或默示 的承诺或保证。

【联系我们】

我们致力于为您提供个性化的售前购买咨询服务,及相应的技术售后服务,任何问题请联系 4009100100或95716。



## 文档目录

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





最近更新时间: 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

最近更新时间: 2025-04-11 15:45:02

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。

## 概述

- 重启准备: 重启期间实例将无法正常提供服务,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

最近更新时间: 2024-11-05 14:53:42

> 腾讯云

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> <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	• 申请 License 并安装 GRID 驱动
這染型实例	直通卡	通用计算	Tesla 驱动	<ul> <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	● 使用预装 GRID 驱动的镜像
	vGPU	图形渲染	GRID 驱动	● 使用预装 GRID 驱动的镜像

() 说明:

1. 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。

2. 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		Tesla 驱动 无特殊要求,官方支持版本	
计算型 GT4	NVIDIA A100	GRID 驱动 Tes 535 Tes 535 Tes 470	535.161.07	CODA 12.4.0	即可。 GRID 驱动 • GRID17 驱动 • GRID16 驱动 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	
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 巡动 470.182.03			
推理型 GI3X	NVIDIA T4				挥項奓见 GRID 驱动版本说 明。	
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		_	_	Tesla 驱动 无特殊要求,官方支持版本 即可。 GRID 驱动 • GRID17 驱动 说明:安装 GRID 17 驱动需 要您前往 NVIDIA 官网下 载。请参见 GRID 驱动版本 说明。获取 License 请联 系售前在线咨询。 • GRID16 驱动 Windows-GRID16.2 驱动 537.70
渲染型 GNV4v	NVIDIA A10				Windows-GRID16.2 驱 动 537.70
渲染型 GN7vw	NVIDIA T4	GRID 驱动	不文持目动安装,可使 用预装 GRID驱动的公 共镜像	_	<ul> <li>Linux-GRID16.2 驱动 535.129.03</li> <li>Windows-GRID16.2 驱动 537.70</li> </ul>



最近更新时间:2025-04-16 15:31:02

## 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。

腾讯云

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: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。	选择公共镜像后自动安装驱动(推荐)
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。	使用脚本安装 GPU 驱动
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。	使用预装 Tesla 驱动的云市场镜像

## 方式一:选择公共镜像后自动安装驱动

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:

镜像 🕐	<b>公共镜像</b> 自定义	镜像    共享镜值	இ 云镜像⋷	5场		
	Ubuntu Ti	推荐 t encentOS	CentOS	Windows		
	👞 请选择镜像		Ň (	D		
	□ 后台自动安装GPU驱动 ⑦					
	<ul> <li>使用说明</li> <li>购买后支持更换操作系</li> <li>该实例若用于图形渲算</li> </ul>	系统,您可根据业务需要在推 验场景,需要安装 GRID 驱动	空制台重装系统 🖸 。 加并配置付费 License 。	您可直接选购 GPU 渲染型	例GN7vw 🕻 ,或使用云市场 🗹 购买已预装 GRID 驱动	力的镜像。

🕛 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。



<b>buntu@V</b> hu Oct	M <b>-2</b> -1 24 22	L <b>24-ubuntu:</b> 2:11:12 2024	∿\$ nvi 1	dia-smi					
NVIDIA	-SMI	535.161.07		l	Driver	Version:	535.161.07	CUDA Versio	on: 12.2
GPU N Fan T	lame emp	Perf	P	ersiste wr:Usag	nce–M e/Cap	Bus-Id 	Disp.A Memory-Usage	Volatile   GPU-Util 	Uncorr. ECC Compute M. MIG M.
0 T N/A	esla 30C	T4 P8		9W /	 On 70W	   0000000   2M	0:00:08.0 Off liB / 15360MiB	-+     0%  -+	0 Default N/A
Proces GPU	ses: GI ID	CI ID	PID	Туре	Proces	ss name			GPU Memory Usage
 No ru	inning	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。

镜像 🕐	公共镜像 	自定义镜像	共享镜像 <b>云镜</b>	象市场
	Ubuntu	TencentOS	CentOS	Windows
	Ubuntu Server	20.04 LTS 64位	~	0
	后台自动安装GPU驱动 建议勾选自动安装驱动, 5	动 🕜 或者您可在购买完成后手动	安装驱动程序确保 GPU 正常使用	3

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:



Proceeding in the second	an column	_	•	
网络巨黑索		✓ 法择基础配置	2 设置网络和主机	(3) 确认配置信息
网络与带风				
网络 ⑦	**********			✓ 子属靴袋可用IP 250 个
	如现有私有网络(子网不符合	1零的要求,可以去控制台 <b>新雄私有网络 2</b> 成 <b>新维子网 22</b> ,云主机构实	后可以通过控制台切映私有网络完成私有网络/子间的切映	
	□ 手动分配IP地址			
公開 19 ①	✓ 分配独立公网 IP			
线路关型	BOP			
带宽计器模式 ①	按小时带宽	按说量计数 共享带宽包		
	注意:流量使用每小时结算	一次, 当账户余额不足时, 两小时内将被停止流量服务, 腾讯云默认可非	P受不超过26bps DOOS 防护,为腾讯云上公网 IP 粉质开	应该护,减少因高度攻击产生的沈重费。
带宽值 ⑦	-			- 5 + Mbps
IPv6 地址	100ps 1000ps 2 所选的VPC/子网未开通	owops IPv6 表开通⊠ ○	100Mbpi	5
安全组				
安全蜡 ⑦	新建安全组	已有安全相		
	请选择安全组(每台)	★例最多关联5个安全组) - ○		
	<b>110</b> #3/1140200.00	3/0 表示所有中地址都可以用于该问,建议填写型来用的中地址		
甘油设置	请情保所造安全组已放进23	2第日(Linux 55H登录),3389第日(Windows记程登录),否则无法记程	登录CVM,您可以进入CVM控制台 <b>修改安全组 团</b>	
HERE				
	标弦键	标签值	8/8	
		③ 健康松弛板		
实例名称 ①	送頃, 不唱飲(J來命) 支持法量连续命名或指定廣	5、交符目间批量符名 试和命名、最多128个字符、你还可以输入128个字符		
登录方式 ①	设置密码	立即关联组织		
	登录名	ubuntu		
	密锁对	请选择密钥对		
		如當現有的密制不合适。可以 現在创建 [2]		
实例销货保护 ①	防止实例通过控制结 实例销货保护和定时销货不	全或者API误制的		
定时销股 ①	一 开启定时销毁			
	开启定时销级后,系统将在	这定时间点自动畅感机器,并且定时情感和实例情感保护不能同时使用。		
安全加固	☑ 免费开通			
	安装组件免费开通ODeS店	P和主机安全基础版 ③		
云首控	✓ 免费开递 免费开通回产品运经、分析	RIXMEN, VIRDPRIXLERIDDER ()		
自动化助手	✓ 免费开递			
	安装银件杂票开通自动化能	(子、発売時、発SSH登泉即可社量管理実例、执行命令、完成日常管理(	59 ①	

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:



4. 按照界面信息逐步操作,完成创建云服务器。



5. 登录实例,可输入 ps aux | grep auto\_install 查看后台运行自动安装脚本。

6. 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。

NVIDI	A-SMI	535.161.07		I	Driver	Version:	535.161.07	CUDA Versio	on: 12.2
GPU Fan	Name Temp	Perf	P	ersiste wr:Usag	nce–M e/Cap	Bus-Id 	Disp.A Memory-Usage	Volatile   GPU-Util 	Uncorr. ECC Compute M. MIG M.
0 N/A	Tesla 39C	T4 P8		22W /	 On 70W	   0000000   2M	0:00:08.0 Off NiB / 15360MiB	+======     0% +	0 Default N/A
Proce GPU	esses: GI ID	CI ID	PID	Туре	Proce	ss name			GPU Memory Usage

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.
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:
```

#### () 说明:

下表列举部分实例类型为裸金属云服务器、高性能计算集群实例类型。



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、 HCCPNV5v	TencentOS Server 3.1(TK4) Ubuntu Server 22.04 LTS 64位 Ubuntu Server 20.04 LTS 64位	550.90.07	12.4.0	8.9.7
GT4、PNV4、GN10Xp、 GN10X、GN8、GN7、 BMG5t、BMG5v、	TencentOS Server 3.1 ( TK4 ) TencentOS Server 2.4 ( TK4 ) Ubuntu Server 22.04 LTS 64位	535.183.06 535.161.07	12.4.0	8.9.7
HCCPNV4h、HCCG5v、 HCCG5vm、HCCPNV4sn、 HCCPNV4sne、HCCPNV5v	Ubuntu Server 20.04 LTS 64位 CentOS 7.x 64位 CentOS 8.x 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 ) Ubuntu Server 20.04 LTS 64位	525.105.17	12.0.1	8.8.0
GT4、PNV4、GN10Xp、 GN10X、GN8、GN7、 BMG5t、BMG5v、 HCCPNV4h、HCCG5v、 HCCG5vm、HCCPNV4sn、 HCCPNV4sne	Ubuntu Server 18.04 LTS 64位 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=

CUDNN VEDSTON-

#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 公网网络费用,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)

最近更新时间: 2025-04-11 15:45:02

#### 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,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。	重装公共镜像时自动安装驱动(推荐)
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。	登录实例使用脚本安装 GPU 驱动
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	使用 TAT 自动化助手安装驱动
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。	使用预装 Tesla 驱动的云市场镜像

#### 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:

青选择目标镜像					
竟像类型 当前镜像	公共镜像	自定义镜像	共享镜像	云市场镜像	
目标镜像 <mark>④</mark> Ubuntu	<b>▼</b> U	buntu Server 22.0	4 LTS 64位   img	g-487zeit 🔻	
	まGPU取売 ③				
GPU驱动版本	535.161.07 🔻	CUDA版本 12.4	.0 🔻	cuDNN版本 8.9.7	•

() 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- Thu Oct 24 22:11	ubuntu:~\$ nvidia-smi :12 2024		
NVIDIA-SMI 535	.161.07 Drive	Version: 535.161.07	CUDA Version: 12.2
   GPU Name   Fan Temp Pe 	Persistence-M rf Pwr:Usage/Cap	Bus-Id Disp.A   Memory-Usage 	Volatile Uncorr. ECC   GPU-Util Compute M.   MIG M.
=========   0 Tesla T4   N/A 30C Pi   	0n 8 9W / 70W	+==========   00000000:00:08.0 Off   2MiB / 15360MiB   +=========	+0   0% Default   N/A +
+   Processes:   GPU GI CI   ID ID	PID Type Proce	ess name	GPU Memory Usage
=====================================	ocesses found		

## 方式二: 登录实例使用脚本安装 GPU 驱动

#### 操作步骤

🦒 腾讯云

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。

→ 小I ④运行中 广州三区 GPU计算型GN/ ④ 8년 3208 5Mpcs (公) 口 - 按数计数 2024-08-06 系命名 (月) 180-0360推 周島:

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:~\$ cat > driver\_install.sh << EOF</pre> #!/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 -0 /tmp/auto\_install.sh & sudo chmod +x /tmp/auto\_install.sh && sudo /tmp/auto\_install.sh > /tmp/auto\_install.log 2>&1 & FOF ubuntu@VM-0-91-ubuntu:~\$ ls driver\_install.sh

3. 输入 bash driver\_install.sh 执行脚本。

ubuntu@VM-0-91-ubuntu:~\$ bash driver_install.sh
<pre>ubuntu@VM-0-91-ubuntu:~\$2024-08-12 21:30:27 https://mirrors.tencentyun.com/install/GPU/auto_install</pre>
sh
Resolving mirrors.tencentyun.com (mirrors.tencentyun.com) 169.254.0.3
Connecting to mirrors.tencentyun.com (mirrors.tencentyun.com) 169.254.0.3 :443 connected.
HTTP request sent, awaiting response 200 OK
Length: 10542 (10K) [application/octet-stream]
Saving to: '/tmp/auto_install.sh'

2024-08-12 21:30:27 (147 MB/s) - '/tmp/auto\_install.sh' saved [10542/10542]

#### 4. 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。

NVID:	IA-SMI	535.161.07			Driver	Version:	535.161.07	CUDA Versio	on: 12.2
GPU Fan	Name Temp	Perf	P P	ersiste wr:Usag	ence-M je/Cap	Bus-Id   	Disp.A Memory-Usage	Volatile   GPU-Util 	Uncorr. ECC Compute M. MIG M.
0 N/A	Tesla 39C	T4 P8		22W /	 On 70W	+======   0000000   2M 	0:00:08.0 Off iB / 15360MiB	+=======     0%	Default N/A
Proce GPU	esses: GI ID	CI ID	PID	Туре	Proces	ss name			GPU Memory Usage

输入 grep -i "finished" /tmp/auto\_install.log 查看驱动、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,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=535.161.0" CUDA\_VERSION=12.4.0 CUDNN\_VERSION=8.9.7 DRIVER\_URL= CUDA\_URL= CUDN\_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:

#### () 说明:

下表列举部分实例类型为裸金属云服务器、高性能计算集群实例类型。

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 ) Ubuntu Server 22.04 LTS 64位 Ubuntu Server 20.04 LTS 64位	550.90.07	12.4.0	8.9.7
GT4、PNV4、GN10Xp、 GN10X、GN8、GN7、 BMG5t、BMG5v、	TencentOS Server 3.1 (TK4) TencentOS Server 2.4 (TK4) Ubuntu Server 22.04 LTS 64位 Ubuntu Server 20.04 LTS 64位 CentOS 7.x 64位 CentOS 8.x 64位	535.183.06 535.161.07	12.4.0	8.9.7
HCCPNV4h、HCCG5v、 HCCG5vm、HCCPNV4sn、 HCCPNV4sne、HCCPNV5v		535.183.06 535.161.07	12.2.2	8.9.4
GT4、PNV4、GN10Xp、 GN10X、GN8、GN7、 BMG5t、BMG5v、 HCCPNV4h、HCCG5v、	TencentOS Server 3.1 ( TK4 ) TencentOS Server 2.4 ( TK4 ) Ubuntu Server 20.04 LTS 64位	525.105.17	12.0.1	8.8.0



HCCG5vm、HCCPNV4sn、 HCCPNV4sne、HCCPNV5v	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

#### 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=
# <b>确保实例可正常下载</b> URL <b>对应安装包</b>
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 <b>建议使用</b> tar.xz、tgz <b>格式安装包</b>
CUDNN_URL=http://mirrors.tencentyun.com/install/GPU/cudnn-linux-x86_64-8.9.7.29_cuda12-archive.tar.xz
① 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 公网网络费用,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:

	公共命令 💿 广州 🗸	
➢ OrcaTerm	命令类型	~
④ 实例自助检测		
	卸载GPU驱动-Linux系统	安装GPU驱动-Linux系统
• 个人命令	♣ 运维工具 🛕 Linux	👶 运维工具 🤷 Linux
・ 公共命令	更新时间 2024-11-05 15:19:57	更新时间 2024-11-05 15:19:42
• 执行记录	命令介绍 用于在GPU云服务器上卸载GPU驱动	命令介绍 用于在GPU云服务器上安装特定版本的 GPU驱动、CUDA、cuDNN
• 托管实例		
● 配额管理 ~	执行命令 > 克隆到个人命令 >	执行命令 > 克隆到个人命令 >

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:



命令参数	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参数均失获	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/NVID IA-Linux-x86_64-535.161.07.run
	CUDA_URL	选填,填入后VERSION参数均失资	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参数均失贫	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/cudn n-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个到	
实例类型:GPU计算型GN7	多个关键字	多个关键字用竖线"1"分隔,多个过滤标签用回车键分隔 Q						
✔ 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
✓ 3Z 后 未命名	🖂 运行中	北京六区	GPU计算型GN7 <mark> </mark>	(公)	包年包月	默认项目	0	在线
	🔿 已关机	北京五区	GPU计算型GN7 <del>《</del>	(公)	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

最近更新时间: 2025-05-08 18:40:22

#### 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。

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 Server 2.4** 

1. 执行以下命令,检查当前系统中是否已安装 dkms。

rpm -q dkms

返回结果如下图,则表示已安装 dkms。

```
[root@■ _____ ~]# rpm -q dkms
dkms-3.2.0-1.tl2.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 页面。
- 3. 根据实例操作系统及实例规格,选择对应系列。实例规格信息请参见 实例类型。本文以 V100 为例,如下图所示:

#### △注意:

选择 Linux 64-bit 即表示下载 shell 安装文件。



#### 手动驱动搜索

按产品、产品类型或系列搜索	Q	
Data Center / Tesla	•	í
V-Series	•	
Tesla V100	•	
Linux 64-bit	•	
12.8	•	
Chinese (Simplified)	•	
查找		

4. 单击**查找**搜寻驱动,选择要下载的驱动版本。

Linux 64-	bit							
Data Cente	Data Center Driver for Linux x64							
什么是 NVIDIA 推	幸驱动?							
该驱动符合适用于i Linux)的驱动可靠性	通过 Windows 硬件质量 生、稳健性和性能给予同	实验室 (WHQL) 测证 样的关注。	式的 Windows 驱动的质量	量级别,因此对非 Wi	ndows 操作系统 (例如			
驱动版 <b>本:</b> 570.124.06	<b>CUDA 工具包</b> 12.8	<b>发布日期:</b> 2025-3-3	<b>文件大小:</b> 375.73 MB	信息: ②	查看			

5. 选择特定的版本单击**查看**进入下载页面,复制**下载**按钮的链接地址。如下图所示:

驱动版本: 发布日期: 操作系统: CUDA 工具包:	570.124.06 Mon Mar 03, 2025 Linux 64-bit 12.8	单击"下载"按钮,即表示您确认已阅读并同意 NVIDIA 软件用 户使用许可。 单击"下载"按钮后,驱动程序将立即开始下载。 NVIDIA 建议用户更新到最新的驱动版本。	下载

6. 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
- 7. 使用 wget 命令,粘贴 步骤5 中复制的链接地址,下载安装包。请将命令中的 xxxx 替换为您复制的链接地址。

wget xxxx

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 kernel-devel 和 kernel-headers,请依次执行以下命令,检查当前系统中是否已安装 kernel-devel 和 kernel-headers 包,且版本要与当前内核版本一致。



rpm -qa |grep -i kernel

腾讯云

返回结果如下,则表示已安装 kernel-devel 和 kernel-headers。

```
[rootd] ______; ~]# rpm -qa |grep -i kernel
kernel-headers-5.4.119-19.0009.56.tl2.x86_64
kernel-modules-5.4.119-19.0009.56.tl2.x86_64
kernel-devel-5.4.119-19.0009.56.tl2.x86_64
kernel-core-5.4.119-19.0009.56.tl2.x86_64
kernel-debuginfo-5.4.119-19.0009.56.tl2.x86_64
tlinux-tkernel4-release-1.0-2.tl2.noarch
kernel-5.4.119-19.0009.56.tl2.x86_64
```

如果在回显信息中,您没有找到当前内核版本的 kernel-devel-\* 和 kernel-headers-\*, 您需要自行安装对应内核版本的 kernel-devel 和 kernel- headers 包。

sudo yum install -y kernel-headers-\$(uname -r).tl2.x86\_64 kernel-devel-\$(uname -r).tl2.x86\_64

10. 升级 gcc 版本, TencentOS 2.4 的默认 gcc 版本为 4.8.5, 需使用 gcc 8 编译 nvidia 内核模块。 执行如下命令可查看 gcc 版本。

gcc --version

如下图所示,gcc版本为4.8.5。

参考如下步骤使用 SCL(Software Collection) 安装 gcc 8, SCL(Software Collection) 使您能够在系统上构建并同时安装相同组件的多个版 本,而不会影响从分发版安装的软件包的系统版本。

```
sudo yum -y install tlinux-release-scl #安装TencentOS的scl源
sudo yum -y install devtoolset-8-gcc #安装scl的提供的gcc 8
sudo scl enable devtoolset-8 bash #激活scl提供的gcc
sudo gcc --version #查看gcc版本
```

11. 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 xxxx 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 -no-cc-version-check -s

12. 安装完成后,执行以下命令进行验证,如返回 GPU 信息,则说明驱动安装成功。

nvidia-smi

13. 重新生成 initramfs。

dracut -f

**TencentOS Server 3.1** 

1. 执行以下命令,检查当前系统中是否已安装 dkms。

pm -q dkms



返回结果如下图,则表示已安装 dkms。

[root@ ]# rpm —q dkms dkms—3.1.8—1.el8.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 页面。
- 3. 根据实例操作系统及实例规格,选择对应系列。实例规格信息请参见 实例类型。本文以 V100 为例,如下图所示:

Data Center / Tesla	•	(
V-Series	•	
Tesla V100	•	
Linux 64-bit	•	
12.8	•	
Chinese (Simplified)	•	
杏坮		

4. 单击**查找**搜寻驱动,选择要下载的驱动版本。

Linux 64-	·bit				
Data Cent	er Driver for Li	nux x64			
什么是 NVIDIA 抄	推荐驱动?				
该驱动符合适用于 Linux)的驱动可靠	通过 Windows 硬件质量 性、稳健性和性能给予同	实验室 (WHQL) 测试 样的关注。	t的 Windows 驱动的质	量级别,因此对非 Wir	idows 操作系统 (例如
驱动版本:	CUDA 工具包	发布日期:	文件大小:	信息:	本毛
					三日

5. 选择特定的版本单击**查看**进入下载页面,复制**下载**按钮的链接地址。如下图所示:

#### Data Center Driver for Linux x64 570.124.06 | Linux 64-bit 驱动主页 > Tesla V100 | Linux 64-bit > Data Center Driver for Linux x64 单击"下载"按钮,即表示您确认已阅读并同意 NVIDIA 软件用 户使用许可。 单击"下载"按钮后,驱动程序将立即开始下载。 NVIDIA 建议用户更新到最新的驱动版本。 驱动版本: 570.124.06 下载 Mon Mar 03, 2025 发布日期: 操作系统: Linux 64-bit 12.8 CUDA 工具包: Chinese (Simplified) 语言: 375.73 MB 文件大小:

- 6. 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

7. 使用 wget 命令,粘贴 步骤5 中复制的链接地址,下载安装包。请将命令中的 xxxx 替换为您复制的链接地址。

get xxxx

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 kernel-devel 和 kernel-headers,请依次执行以下命令,检查当前系统中是否已安装 kernel-devel 和 kernel-headers 包,且版本要与当前内核版本一致。

rpm -q |grep -i kernel

返回结果如下,则表示已安装 kernel-devel 和 kernel-headers。

如果在回显信息中,您没有找到当前内核版本的 kernel-devel-\* 和 kernel-headers-\*,您需要自行安装对应内核版本的 kernel-devel 和 kernel- headers 包。

sudo yum install -y kernel-headers-\$(uname -r).tl3.x86\_64 kernel-devel-\$(uname -r).tl3.x86\_64

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 xxxx 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 -no-cc-version-check -s

11. 安装完成后,执行以下命令进行验证,如返回 GPU 信息,则说明驱动安装成功。

nvidia-smi

12. 重新生成 initramfs。

lracut -f

13. 开启常驻模式。

sudo nvidia-persistenced

Ubuntu

() 说明:

以下步骤适用于直通卡型(GPU卡数 ≥ 1)Ubuntu 系统 Tesla 驱动和 GRID 驱动安装,若您使用 GRID 驱动,请参考文档申请和配置 GRID license,详情请参见 安装 NVIDIA GRID 驱动。

1. 执行以下命令,检查当前系统中是否已安装 dkms。



dpkg -l   grep -i dkms		
返回结果如下图,则表示已安装 dkms。		
ubuntu@I @	7–2ubun	ntu2.2 all Dynamic Kernel Module Support Framework
若返回结果为空,则表明未安装 dkms,执行以下命	令进行	安装。
sudo apt-get install <b>dkms -y</b>		
2. 登录 NVIDIA 驱动下载 页面		
3. 根据实例操作系统及实例规格,选择对应系列。实例	刘规格信	信息请参见
<u> </u>	文件。	
手动驱动搜索		
按产品、产品类型或系列搜索 C	L	
Data Center / Tesla	í	
V-Series •		
Tesla V100		
Linux 64-bit		

4. 单击**查找**搜寻驱动,选择要下载的驱动版本。

12.8

查找

Chinese (Simplified)

Linux 64	-bit				
Data Cent	er Driver for Li	nux x64			
什么是 NVIDIA 打	推荐驱动?				
该驱动符合适用于 Linux)的驱动可靠	通过 Windows 硬件质量 性、稳健性和性能给予同	实验室 <b>(WHQL)</b> 测试 样的关注。	的 Windows 驱动的质	量级别,因此对非 Wind	dows 操作系统 (例如
驱动版本:	CUDA 工具包	发布日期:	文件大小:	信息:	杏舌

•

•

5. 选择特定的版本单击**查看**进入下载页面,复制**下载**按钮的链接地址。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。请将命令中的 xxxx 替换为您复制的链接地址。

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.rur

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。

ubuntu@ ubuntu@ ubuntu@ ubuntu@ ubuntu@	-i gcc	/	
ii gcc	4:11.2.0-1ubuntu1	amd64	GNU C compiler
ii gcc-11	11.2.0-19ubuntu1	amd64	GNU C compiler
ii <mark>gcc</mark> -11-base:amd64	11.2.0-19ubuntu1	amd64	GCC, the GNU Compiler Collection (base pa
ckage)			
ii gcc-12	12.3.0-1ubuntu1~22.04	amd64	GNU C compiler
ii <b>gcc</b> -12-base:amd64	12.3.0-1ubuntu1~22.04	amd64	GCC, the GNU Compiler Collection (base pa
ckage)			
ii libcc1-0:amd64	12.3.0-1ubuntu1~22.04	amd64	GCC cc1 plugin for GDB
ii lib <mark>gcc</mark> -11-dev:amd64	11.2.0-19ubuntu1	amd64	GCC support library (development files)
ii lib <mark>gcc</mark> -12-dev:amd64	12.3.0-1ubuntu1~22.04	amd64	GCC support library (development files)
ii lib <mark>gcc</mark> -s1:amd64	12.3.0-1ubuntu1~22.04	amd64	GCC support library
ii libgomp1:amd64	12.3.0-1ubuntu1~22.04	amd64	GCC OpenMP (GOMP) support library
ii libquadmath0:amd64	12.3.0-1ubuntu1~22.04	amd64	GCC Quad-Precision Math Library
ubuntu에 🐂 🚮 🖬 🖬 🖬 v>\$ dpkg –l   grep	-i linux-headers		
ii linux-headers-5.15.0-130	5.15.0-130.140	all	Header files related to Linux kernel vers
ion 5.15.0			
<pre>ii linux-headers-5.15.0-130-generic</pre>	5.15.0-130.140	amd64	Linux kernel headers for version 5.15.0 o
n 64 bit x86 SMP			
ii linux-headers-generic	5.15.0.130.128	amd64	Generic Linux kernel headers
ubuntu@			l .

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-headers-\$(uname -r) -y

10. 安装 libgl1-mesa-glx。

sudo apt install libgl1-mesa-glx -y

11. 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 xxxx 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 -no-cc-version-check -s



12. 安装完成后,执行以下命令进行验证,如返回 GPU 信息,则说明驱动安装成功。

13.更新 initramfs。 sudo update-initramfs -u	
sudo update-initramfs -u	
14. 开启常驻模式。	
sudo nvidia-persistenced	

## 安装失败原因

Linux 系统驱动安装失败表现为 nvidia-smi 无法工作,通常原因如下:

- 1. 系统缺乏编译 kernel module 所需要的包,如 gcc、kernel-devel-xxx 等,导致无法编译,最终安装失败。
- 2. 系统里面存在多个版本的 kernel,由于 DKMS 的不正确配置,导致驱动编译为非当前版本 kernel 的 kernel module,导致 kernel module 安装失败。
- 3. 安装驱动后,升级了 kernel 版本导致原来的安装失效。
- 4. 确认是否有安装 nouveau 驱动,如果有安装需卸载。



# 手动安装 Tesla 驱动-Windows

最近更新时间: 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:

<mark>驱动</mark> 全部驱动 GeForce 驱动 网络驱	动 🚽	
手动驱动搜索		
按产品、产品类型或系列搜索	Q	
Data Center / Tesla	•	)
T-Series	•	
Tesla T4	•	
Windows Server 2022	▼	
12.2	•	
Chinese (Simplified)	•	
查找		

#### 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,按照界面上的提示安装驱动程序,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。

on 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 TCC/WDDM Bus-Id Disp.A Volatile Uncorr. ECC Fan Temp Perf Pwr:Usage/Cap Memory-Usage GPU-Util Compute M. MIG M.		
0 Tesla T4 TCC 00000000:00:06.0 0ff 0 0 N/A 39C P8 9W / 70W 0MiB / 15360MiB 0% Default N/A		
Processes: GPU GI CI PID Type Process name GPU Memory ID ID Usage		
No running processes found		
C:\Users\Administrator>_		v



## 使用预装 Tesla 驱动的云市场镜像

最近更新时间: 2025-04-11 15:45:02

#### 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:

镜像 🕐	公共镜像	自定义镜像	共享镜像	云镜像市场	
	从云镜像市场选择	]			

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:

驱动				热搜:	Ubuntu Windows		
推荐镜像	操作系统	全部  ~	系统架构 全部	~		4	综合排序
全部 操作系统 数据库与缓存 立用运行环境	J.	TensorFlow 2.8.( 预装 Al 环境: Tenso 操作系统 Ubuntu	D Ubuntu 18.04 GPU prFlow 2.8.0 + Miniconda 系统架构 x86_64 服	<b>基础镜像(预装46</b> + OpenCV 4 + Pyth 务商 腾讯云计算(北	<b>D驱动)</b> ion 3.9.12 + CUDA11.2.2 京) 有限责任公司	+ cuDNN8.2.1 + 价格: <b>免费</b>	GPU驱动版 <mark>立即使用</mark>
管理与监控 聿站镜像 安全		Windows Server Windows Server 20 操作系统 Windows	2022 数据中心版 64( 22 数据中心版 64位 中文 : 系统架构 x86_64	<b>立 中文版(预装53</b> 版(预装538驱动) 服务商 <b>腾讯云计算</b> (:	8 <b>8驱动)</b> 北京)有限责任公司		
≧业应用						价格:免费	立即使用



# 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。

镜像 ②	公共镜像	自定义镜像	共享镜像	云镜像市场	
	Windows Server 2022	数据中心版 64位 中文版	(预装538驱动)	重新选择	

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)

最近更新时间: 2025-04-11 15:45:02

## 操作场景

腾讯云

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,可以选择预装 GRID 驱动的镜像。

## 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。



2. 在镜像中选择公共镜像,单击请选择镜像如下图所示:

镜像 🕐	公共镜像	自定义镜像	共享镜像	云镜像市场
	•Ç Ubuntu	TencentOS	Windows	
	🜲 请选择镜像			~ ⑦

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

最近更新时间: 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 這染型实例,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天内在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:

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:





4. 打开邮件 "NVIDIA Set Password"。如下图所示:



NVIDIA Team

- 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:

冬 NVIDIA. LICENSING	TVIDIA, LICENSING   Dashboard (ORG_ADMIN)   Logout					
🚮 DASHBOARD						
ENTITLEMENTS	Entitlements		MANAGE ENTITLEMENTS	License Servers	MANAGE SERVERS CREATE SERVER	
•=	▲ ENTITLEMENT / FEATURE	EXPIRATION	ALLOCATED / TOTAL	▲ LICENSE SERVER / FEATURE	IN USE / ALLOCATED	
E LICENSE SERVERS	∧ 051Men;			∧ 🤣 (E) (E)		
SOFTWARE DOWNLOADS	∧ 051Men			∧ ⊘ windowsserver		
VIRTUAL GROUPS	∨ 4reh4k0mlj-	59,6,26,3		∧ 🥝 5254008		
	Quadro-Virtual-DWS-5.0	2020-03-26	0 / 128	∧ 📀 5254008		
*9 HISTORY	GRID-Virtual-Apps-3.0	2020-03-26	0 / 128	∧ 🥝		

软件准备

() 说明:

1. 选择左侧导航栏中的 SOFTWARE DOWNLOADS,进入 Software Downloads 页面。

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。

2. 选择下拉框中的 Miscrosoft 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:

	FEATURED	ALL AVAILABLE			ADDITIONAL SOFTWARE V
SOFTWARE DOWNLOADS	Microsoft Win V 2016	✓ VGPU Versio	on: All 🗸 Search		CLEAR COLUMNS V
VIRTUAL GROUPS	PLATFORM	PLATFORM VERSION	VGPU VERSION	DESCRIPTION	RELEASE DATE
<ul> <li>✓) HISTORY</li> </ul>	Microsoft Windows Hyper-V	2016	9.3	NVIDIA vGPU for Windows	2020-03-09 Download

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。

① 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:

📀 NVIDIA. LICENSING	Software Downloads				NVIDIA Application Hub   Detencent.com (ORG_ADMIN)   Logout
값 DASHBOARD		~	]		
ENTITLEMENTS	66	<b>\$</b> \$			∠ 2019.11 64-bit License Manager for Windows
ILICENSE SERVERS	FEATURED	ALL AVAILABLE			🛃 2019.11 64-bit License Manager for Linux
	Microsoft Win V 2016	V	on: All V Search		🛃 2018.10 32-bit License Manager for Windows
🛞 SOFTWARE DOWNLOADS	2010		on. Air		🕁 2018.10 License Manager for Linux
VIRTUAL GROUPS	PLATFORM	PLATFORM VERSION	VGPU VERSION	DESCRIPTION	🕁 GPUMode Change Utility for Tesla M60 and M6 GPUs
	Microsoft Windows Hyper-V	2016	9.3	NVIDIA vGF	L NVIDIA Virtual GPU Management Pack for vRealize Operations

- 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:

V 🙁 🗄 🖓 🖉		
	⊻ < ⊖ ⊕	
GRID-Virtual-PC-2.0 FLEXERA	2020-12-11	1024 / 1024
NVIDIA-vComputeServer-9.0 FLEXERA	2020-12-11	8192 / 8192

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:

Add Features Add a feature to this server from your entitlments and set	the nu	imber of cons	umed lice	X nses.
Product		Licenses		
Quadro-Virtual-DWS-5.0 - 1 available (051Men	$\mathbf{\vee}$	1	$\hat{}$	ADD

- 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:

`	🕗 🕮 main				
	GRID-Virtual-PC-2.0	FLEXERA	2020-12-11	1024 / 1024	
	NVIDIA-vComputeServer-9.0	FLEXERA	2020-12-11	8192 / 8192	
	Quadro-Virtual-DWS-5.0	FLEXERA	2020-12-11	1024 / 1024	

## 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:
  - ← → C () localhost:8080/licserver/manageFeatureUsage\_view.action?page=1

$\leftrightarrow$ $\rightarrow$ C (i) localhost:8080	//licserver/manageFeatureUsage_	view.action?page=	=1			☆ 🖰 🗄
	DIA.					
	Licensed Feat	ure Usag	je			
Licensed Olients	Features				Search [case-sensitive]:	Search by: Feature 🔻 🔍 📝
License Management     Configuration	Click the feature table header row Features served to or reserved for	to obtain a single sorte clients. Click a feature i	d non-paginated lis name for usage det	t. ails.		
Login	Feature	Version	Count	Available	Expiry	Vendor String
	Quadro-Virtual-DWS	5.0	1	1	2020-07-02	Quadro-Virtual-DWS-1
	GRID-Virtual-Apps	3.0	1	1	2020-12-11	GRID-Virtual-Apps-2
License Client Manager	GRID-Virtual-Apps	3.0	1	1	permanent	GRID-Virtuel-Apps-3
> About	Page 1 of 1					
Settings	Go to page 1 🔻					
	Total number of records: 3					

#### 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:

	NVIDIA 🕏	<del>、</del> 装程序		<b>-</b>   <b>-</b>   <b>x</b>
NVIDIA 图形 版本 432.33	驱动程序			
<ul> <li>✓ 系统检查</li> <li>✓ 许可协议</li> </ul>	NVIDIA 安装和	呈序己完成		
<ul> <li>◇ 迭顶</li> <li>◇ 安装</li> <li>结束</li> </ul>	组件 NVIDIA NGX NVIDIA WMI nView 图形驱动程序	版本 1.3.0.42 2.33.0 149.77 432.33	状态 已安装 已安装 已安装	
			1	关闭( <u>C</u> )

## 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:

腾讯云

La 设备管理器	_	×
文件(F) 操作(A) 查看(V) 帮助(H)		
∨ ≞ 172_17_64_15		
> 🔐 DVD/CD-ROM 驱动器		
> 📷 IDE ATA/ATAPI 控制器		
> 🔜 磁盘驱动器		
> 🍇 存储控制器		
> 💼 打印队列		
> 📮 端口 (COM 和 LPT)		
> 🛄 计算机		
> 🛄 监视器		
> 📖 線盘		
✔ 쨰 人体学输入设备		
ଲ HID Button over Interrupt Driver		
📊 USB 输入设备		
> 🚽 软盘驱动器控制器		
> 🕠 声音、视频和游戏控制器		
> 🕕 鼠标和其他指针设备		
> 🏺 通用串行总线控制器		
> 🚽 网络适配器		
> 🛅 系统设备		
▶ 🔄 显示适配器		
🔙 Microsoft 基本显示适配器		
🕞 NVIDIA GRID T4-4Q		

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。

# ① 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



<b>5</b>	NVIDIA 控制面板	
文件(E) 编辑(E) 桌面(K) 帮助(H)		
🔇 File 🔹 🐑 🐔		
选择一项任务 日 30 设置 一通过指约调整图像设置 一管理 30 设置 日 许可 一直过新可加	<ul> <li>管理许可证</li> <li>参可以通过应用许可证未启用附加功能。</li> <li>许可证类型:</li> <li>@ Quadro Virtual Data Center Workstation</li> <li>GRID Virtual Applications</li> <li>() 你的系統目前的零先 CRID Virtual Applications</li> </ul>	
	- 銀江可证服务器: 118.24. 選口号: 8080	
Tuble	ta la	(EI(A) EDGA





# 安装 CUDA 驱动

最近更新时间: 2025-04-11 15:45:02

## 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 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:



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。



	📀 NVIDIA. DEVELOPER	Home Blog Forums Docs Downloads	is Training	٩	Join			
	Topics - Platforms - Industries -		10 4 D					
	Select Target Platform	JDA TOOIKIT	12.4 Down	loads				
	Click on the green buttons that describe and conditions of the <u>CUDA EULA</u> .	your target platform. Only supported platforms w	will be shown. By downloading and using th	ne software, you agree to fully comply v	vith the terms			
	Operating System	Linux Windows						
	Architecture	x86_64 ppc64le arm6	i4-sbsa aarch64-jetson					
	Distribution	CentOS Debian Fedor	ra KylinOS OpenSUSE	RHEL Rocky SLES	Ubuntu			
	Distribution	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 dow	vnload below.						
	> Base Installer							
	Installation Instructions:	wnload nyidia com/compute/cuda/12.4.0/loc	ral installers/curia 12.4 0 550 54.14	1 linux-run				
	\$ sudo sh cuda_12.4.0_550.5	1.14_linux.run	La (_113ta ttel 3) tuda_12.4.0_330.34.14					
4.	参见 使用标准方式登录	t Linux 实例(推荐),到	登录 GPU 实例。您也	可以根据实际操作习	惯,选择其他不	下同的登录方式:		
	○ 使用远程登录软件	登录 Linux 实例						
	○ 使用 SSH 登录 L	inux 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	乐示:				
	[root@VM-0-13-tencentos 2024-11-07 09:00:13 Resolving developer down	test]# wget https://develop https://developer.download load pyidia com (developer	per.download.nvidia.com/ d.nvidia.com/compute/cuc .download_pvidia_com)	/compute/cuda/12.4.0/l da/12.4.0/local_instal 152 100 30 144	ocal_installers lers/cuda_12.4.	/cuda_12.4.0_550.54. 0_550.54.14_linux.ru	.14_linux.ru un	n
	Connecting to developer. HTTP request sent, await	download.nvidia.com (develo ing response 200 OK	oper.download.nvidia.com	n)  152.199.39.144 :443	connected.			
	Length: 4454353277 (4.1G Saving to: 'cuda_12.4.0_	) [application/octet-stream 550.54.14_linux.run'	n]					
	cuda_12.4.0_550.54.14_li	nux.run 5%[===>				] :	236 <b>.</b> 51M 11.	3MB/s e
	或者您可在本地系统下	载 CUDA 安装包,再上传	专到 GPU 实例的服务	器。下载或上传完成词	可使用 11 查	看当前目录文件。		
	[root@VM-0-13-tencen	tos test]# ll						
	total 4349960 -rw-rr 1 root ro	ot 4454353277 Feb 29 2	2024 cuda_12.4.0_550	.54.14_linux.run				
~		tos test]#				0 54 44 11	注:+0	
6.	依次执行以下命令, <u>为</u> :	女装巴添加执行权限,谓伯	王 root 用户下操作。		la_12.4.0_5	0.54.14_linux	.run אאת	动们了权限。
	chmod +x cuda_1	2.4.0_550.54.14_lin	ux.run					
	./cuda_12.4.0_5	50.54.14_linux.run	toolkitsamj	plessilent -	-override			
	如卜图所示:	tos tostl#						
	[root@VM-0-13-tencent [root@VM-0-13-tencent [root@VM-0-13-tencent [root@VM-0-13-tencent	test]# tos test]# chmod +x cuda tos test]# ./cuda_12.4. tos test]# _	la_12.4.0_550.54.14_1 0_550.54.14_linux.ru	linux.run un —-toolkit —-sar	nples —silen	t — override		
	▲ 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。 eta 5m 58s





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 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:





5. 打开下载 CUDA 所在的文件夹,双击安装文件开始安装。

NVIDIA 安装程序	- 🗆 X
<b>NVIDIA CUD/</b> 版本 12.4	
◇ 系统检查	NVIDIA 软件许可协议
许可协议	请仔细阅读以下 NVIDIA软件许可协议。
选项	End User License Agreement
安装	
结束	NVIDIA Software License Agreement and CUDA Supplement to Software License Agreement.
	The CUDA Toolkit End User License Agreement applies to the NVIDIA CUDA Toolkit, the NVIDIA CUDA Samples, the
	如1225-74 如果您接受此协议条款,请单击同意并继续。 ————————————————————————————————————
	同意并继续( <u>A)</u> 取消( <u>C</u> )

6. 若最后出现完成对话框,则表示安装成功。如下图所示:



NVIDIA 安装程序			-	□ ×
<b>NVIDIA CUDA</b> 版本 12.4	1			
✓ 系统检查	NVIDIA 安装程用	亨己完成		
✓ 许可协议	组件	版本	状态	^
🛛 🕹 选项	CUSPARSE Runtime	12.4	已安装	
✓ 安装	CUBLAS Runtime	12.4 12.4	日安装	
	Compute Sanitizer	12.4	ロシ表 已安装	
結果	CURAND Development	12.4	已安装	
	CUDART Runtime	12.4		× .
	<			>
			l	关闭( <u>C</u> )

# 卸载 NVIDIA 驱动 卸载 Tesla 驱动

最近更新时间: 2025-04-11 15:45:02

## 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,以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

Linux

## 方式一、使用 TAT 自动化助手卸载驱动

- 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,并单击卸载 GPU 驱动-Linux 系统模块左下角的执行命令。如下图所示:

概览	公共命令 💿 广州 🗸
OrcaTerm	命令类型 🗸 命令分类
② 实例自助检测	
⑦ 自动化助手 ^	卸载GPU驱动-Linux系统
・ 个人命令	よ 运维工具   ▲ Linux
・ 公共命令	更新时间   2024-11-05 15:19:57
・执行记录	命令介绍 用于在GPU云服务器上卸载GPU驱动
・ 托管实例	
● 配额管理 ~	执行命令 > 克隆到个人命令 >

3. 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:

选择执行实例								已选择1个§
实例类型:GPU计算型GN	7 多个关键字	『用竖线 " " 分隔	,多个过滤标签用回车银	建分隔		Q		
✔ 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
✓ 3Z □ 未命名	🖂 运行中	北京六区	GPU计算型GN7 ④	(公)	包年包月	默认项目	$\bigtriangledown$	在线
1000	()已关机	北京五区	GPU计算型GN7 <del>《</del> 》	(公)	包年包月	-		未安装

4. 单击执行命令即可。

#### 方式二: 手动卸载驱动

1. 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。

如下图所示:						
root@VM-2-124-ubu	n <mark>tu:/home/ubuntu#</mark> nvidia-installer -	–uninstall –s				
WARNING: Your driv VIDIA driver throu than nvic t can. Please sec '/var/log	ver installation has been altered si ugh a mechanism other dia-installer (such as your distribu e the file y/nvidia-uninstall.log' for details.	nce it was initially tion's native package	installed; this may	/ happen, for ex	ample, if you have sir	nce inst Ininstal
3.执行 lsmod	grep nvidia 检查驱动是否卸载	战干净。如果结果为3	空,则已经卸载干	净,如下图所表	<b>⊼</b> :	
root@VM-2-1 root@VM-2-1	L24-ubuntu:/home/ubuntu L24-ubuntu:/home/ubuntu	# sudo lsmod #	grep nvidi	a		
如果出现部分文件	<b>持残留,可重复执行下列命令</b> :					
fuser -k /d	lev/nvidia*; rmmod nvidia_	_modeset; rmmod	. nvidia_drm;	rmmod nvidi	ia_uvm; rmmod nv	/idia;
<ul> <li>说明:</li> <li>注意使用</li> </ul>	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。		
I. 再次执行 lsmod	grep nvidia <b>检查驱动是否</b>	卸载干净,如果为空	≥,则已卸载成功。	,		
1. 再次执行 lsmod Vindows	grep nvidia <mark>检查驱动是否</mark>	卸载干净,如果为5	≧,则已卸载成功。	,		
4. 再次执行 1smod Vindows 1. 远程登录 GPU \$ 2. 单击左下角 € , ◎ 199400%	grep nvidia 检查驱动是否 在例,参见 使用远程桌面连接登录 进入控制面板 > 程序与功能,右行	卸载干净,如果为至 Windows 实例 。 建选择 <b>卸载。</b>	≧,则已卸载成功。	X		
4. 再次执行 1smod Vindows 1. 远程登录 GPU \$ 2. 单击左下角 <sup>■</sup> ,	grep nvidia 检查驱动是否 实例,参见 使用远程桌面连接登录 进入控制面板 > 程序与功能,右键 R、所相控制面版, 图序和功能	卸载干净,如果为玄 : Windows 实例 。 建选择 <b>卸载。</b>	2,则已卸载成功。	- □ × 時100時: ♪		
<ol> <li>再次执行 1smod</li> <li>Vindows</li> <li>远程登录 GPU \$</li> <li>单击左下角 ■</li> <li>● 程序印功器</li> <li>◆ → ~ ↑ ● → 控制回路</li> <li>☆ → ~ ↑ ● → 控制回路</li> <li>☆ → ~ ↑ ● → 控制回路</li> </ol>	grep nvidia 检查驱动是否 に例,参见 使用远程桌面连接登录 进入控制面板 > 程序与功能,右 a、所有控制面板, 医序和功能	卸载干净,如果为S Windows 实例。 建选择 <b>卸载。</b>	≥,则已卸载成功。 → δ 腹変態	- П X Жиллав: Р		
<ol> <li>再次执行 1 smod</li> <li>Windows</li> <li>远程登录 GPU \$</li> <li>・ 逆程登录 GPU \$</li> <li>・ 単击左下角 ●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●</li></ol>	grep nvidia 检查驱动是否 它例,参见 使用远程桌面连接登录 进入控制面板 > 程序与功能,石铭 R、所有控制圆版项,图序和功能 和载或更改程序 和赛和联系度, 课从列集中将其选中,然后单击"知敏"、"更必	:卸载干净,如果为≤ :Windows 实例。 建选择 <b>卸载。</b>	≥,则已卸载成功。 ~ δ 渡索砲	- — X жалласт р		
<ol> <li>再次执行 1 smod</li> <li>Vindows</li> <li>近程登录 GPU \$</li> <li>単击左下角 ●● ,</li> <li>● 程序和功能</li> <li>◆ → ◆ ◆ ● ,</li> <li>使制面版主页</li> <li>● 程序和功能</li> <li>◆ → ◆ ◆ ● ,</li> <li>● 成用或关闭 Windows 功能</li> </ol>	grep nvidia         检查驱动是否           在例,参见使用远程桌面连接登录           进入控制面板>程序与功能,石结           进入控制面板>程序与功能,石结           建筑或型改程序           建筑或型改程序           建筑或型改程序           建筑和集集表           MR3成型改程序           通数成型改程序           通数成型改程序           通知表示的表示。           通知表示的表示。           通知表示的表示。           建築           Microsoft Visual C++ 2015-2019 Redistributable (x64)           Provershell 7-x64 <b>PRU/PRX(U)</b>	御載干净,如果为名 Windows 实例。 建选择卸載。 建选择卸載。 建造でPavlov Cloudbase Solutions Srl Microsoft Corporation Microsoft Corporation NVIDIA Corporation NVIDIA Corporation NVIDIA Corporation Microsoft Corporation Microsoft Corporation Microsoft Corporation NVIDIA Corporation	を、则已印載成功。	ー □ × 附口功能・ ♪ 下口力能・ ♪		





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

最近更新时间: 2025-04-11 15:45:02

## 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,以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。

Linux



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。

nvidia-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-0-13-tencentos ~]# lsmod | grep nvidia
[root@VM-0-13-tencentos ~]#

# 升级 NVIDIA 驱动 升级 Tesla 驱动

最近更新时间: 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。



# 升级 GRID 16 驱动

最近更新时间: 2024-11-01 16:48:02

## 背景信息

NVIDIA 在2023年7月31日后不再对旧版 GRID License Server 提供技术支持,且腾讯云采购的旧版 GRID License 也将于2024年3月30日到期,为 了您的业务不受驱动到期失效影响,建议您尽快将目前使用的 GRID 驱动版本升级到由新版 GRID License Server 提供支持的 GRID 16。

#### () 说明:

关于 NVIDIA 官网不再对旧版 GRID License Server 提供技术支持的更多信息,请参见: NVIDIA Virtual GPU Software License Server End of Life Notice。

## 升级方案

根据不同卡类型不同,升级方案有所差异。

#### GPU 直通卡实例规格

如果您的实例属于 GNV4v、GNV4、GN7vw 整卡 GPU 规格,您可以根据实际情况,选择以下方式升级 GRID 驱动。

- 使用自定义镜像重装:如果您需要保留原始镜像中的业务环境,推荐您在原始镜像中卸载旧驱动,升级为 GRID 16驱动,制作为 自定义镜像 后重装系统,驱动升级操作步骤将通过站内信通知。
- 更换公共镜像:如果您不需要保留原始镜像中的业务环境,您可以使用 GPU 云服务器控制台 重装系统 功能,选择预装 GRID 16.2 的公共镜像重装。

#### GPU 切分卡实例规格

如果您的实例属于 GNV4v、GN7vw 切分 GPU 规格,腾讯云会通过 <mark>实例维修任务</mark> 进行升级,具体步骤如下:

- 1. 升级前一周会发出站内信通知,告知预计升级时间。
- 2. 根据站内信指引制作自定义镜像,如果您选择使用预装 GRID 16.2 的公共镜像重装系统,可跳过本步。
- 3. 升级时间前48小时,登录云服务器控制台,选择左侧导航栏中的维修任务 > 任务列表,如下图所示:

任务ID	产品类型	主机ID	类型	任务说明	状态	创建时间 ↓	计划维护时间 \$	结束时间 \$	操作
	云服务器		实例维护升级	监控到您的云服务器因…	已结束	2024-01-09 18:14:40	2024-01-09 19:47:19	2024-01-10 16:40:20	授权/预约

4. 在维修任务列表页面,查看类型为实例维护升级的任务列表。

- 5. 单机授权/预约,在弹出的窗口中,完成授权和时间预约操作即可完成授权。
- 6. 实例维护升级任务完成后,使用 步骤2 中制作的自定义镜像或选择预装 GRID 16.2 的公共镜像重装系统即完成升级。

#### () 说明:

- 如果您没有完成授权操作,实例维修任务下发后48小时会默认升级。
- 如果您对升级节奏有特殊需求,可以咨询您的腾讯云架构师 / 售后经理提供支持。

# 🔗 腾讯云

# 使用 GPU 监控和告警

最近更新时间: 2025-06-25 14:10:22

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,移动鼠标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	评估负载使用的显存量	MB	per-GPU
GPU 显存使用率	评估负载使用的显存量占显存总量的百分比	%	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:

云产品监控     应用性能监控     前端性能监控     云拨测     告警数据源     Prometheus       指标 ()     GPU / 云服务器 / GPU监控     ▼     GPU监控 / GPU编码器使用率(% ▼       筛选 ()     实例     ▼     16个(ins-h4gjsrs3,ins-jccunj ▼		New	New	ew -		
指标 (i) GPU / 云服务器 / GPU监控 ▼ GPU监控 / GPU编码器使用率(% ▼ 筛选 (i) 实例 ▼ 16个(ins-h4gjsrs3,ins-jccunj ▼	云产品监控	应用性能监控 前端	性能监控	告警数据源	Prometheus	
筛选 🗊 🛛 案例 💌 16个(ins-h4gjsrs3,ins-jccunj 🔻	指标 🛈	GPU / 云服务器 / GPU监控	▼ GPU监控 / GF	PU编码器使用率(% ▼		
	筛选 🚺	实例 🔻	16个(ins-h4gjsrs3,ins-	jccunj 🔻		

#### 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	评估负载使用的显存量	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。

1. 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。

2. 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。



腾讯云可观测平台	← 新建告警策■			
监控概览				
○ 生整等理 。	1 配置告警	> 2 配置告警通知		
	基本信息			
<ul> <li>告警配置</li> </ul>				
<ul> <li>告警治理</li> </ul>	策略名称	GPU告誓配置参考示例		
<ul> <li>・ 统计大盘</li> </ul>	备注	最多100个字符		
Dashboard				
◎ 援入中心				
─ 报表管理	配置告警规则			
全景监控	监控类型	任07 云产品监控 应用性能监控 前端性	HOT HOT 能监控 云拨测	终端性能监持
△ 云产品监控 ∨	策略类型	云服务器 / GPU监控 >		
♀ Prometheus 监控				
G Grafana 服务	所属标签	标签键 🖌 标签值	× 0	
() 应用性能收达 >		+ 添加 ③ 罐值粘贴板		
⊙ 应用性能监控 ◇				
™ 前端性能监控 ~	告警对象(与) ()	地域 ~ 上海	Ý	
② 终端性能监控 ~		实例 × 已选择1个。	. ×	+
		将对上述维度筛选条件的所有gpu做告警检测		

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 nvidia-smi 命令查看 GPU 的功率是否有 ERR 或 nvidia-smi -i <target gpu=""> -q  grep "Power Draw" 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 提交工单 联系腾讯云支持。</target>
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 <mark>提</mark> <mark>交工单</mark> 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 nvidia-smi -i <target gpu=""> -q -d PAGE_RETIREMENT 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 提交工单 联系腾讯云支持。</target>
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 nvidia-smi -i <target gpu=""> -q -d ROW_REMAPPER 命令 查看是否有 GPU 卡处于 Pending 状态,重置 GPU 卡或重启实例恢</target>



		kill,GPU卡处于pending 状 态。	复。若重启无法恢复 <mark>提交工单</mark> 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。