

GPU 云服务器 常见问题



腾讯云

【 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 】



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。

常见问题

最近更新时间：2025-03-19 20:11:32

什么是 NVIDIA Tesla?

NVIDIA Tesla（特斯拉）是 NVIDIA 继专业加速卡 QUADRO 及娱乐图形卡 GeForce 系列后，推出的一个全新的产品线，主要应用于广大科学研究的高性能计算需求。利用 NVIDIA® Tesla® GPU 加速器更快地处理要求超级严格的 HPC 与超大规模数据中心的工作负载。

什么是加速计算?

加速计算是指使用硬件加速器或协同处理器来执行浮点数计算和图形处理，比使用在 CPU 上运行的软件更高效。您可以根据业务场景选择适用于通用计算的 [GPU 计算型](#) 或适用于图形密集型应用程序的 [GPU 渲染型](#)。

GPU 相对于 CPU 有哪些优势?

GPU 比 CPU 拥有更多的逻辑运算单元（ALU），支持多线程大规模并行计算。

何时应选择使用 GPU 的实例?

GPU 实例最适用于并行度极高的应用程序。例如，使用数千个线程的工作负载。当图形处理有大量的计算要求，其中的每个任务都相对较小时，执行的一组操作形成了一个管道，而此管道的吞吐量要比单个操作的延迟更为重要。要构建能充分利用这种并行度的应用程序，用户需要掌握 GPU 设备的专项知识，了解如何针对各种图形 API（DirectX、OpenGL）或 GPU 计算编程模型（CUDA、OpenCL）进行编程。

如何根据不同实例类型和场景选择驱动?

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，也可以使用 GRID 驱动（少数卡型不支持 GRID 驱动）来实现不同场景的计算加速。
- vGPU 只能使用特定版本的 GRID 驱动来实现计算加速。

NVIDIA GPU 相关实例的驱动安装步骤请参见 [安装 NVIDIA Tesla 驱动](#) 及 [安装 NVIDIA GRID 驱动](#)。您可参见下表，结合实例类型及应用场景来选择驱动类型：

实例类型	场景	驱动类型	推荐安装方式
计算型实例 - 直通卡型	通用计算	Tesla 驱动	<ul style="list-style-type: none">购买页勾选后台自动安装GPU驱动NVIDIA 官网下载驱动后安装
	图形渲染	GRID 驱动	<ul style="list-style-type: none">在购买页镜像市场中选择已预装 GRID 驱动的付费镜像

			<ul style="list-style-type: none"> • NVIDIA 官网申请 GRID 驱动和 License 后安装
渲染型实例 – vGPU – vDWS/vWS	图形渲染	GRID 驱动	选择已预装 GRID 驱动的指定镜像

GPU 实例如何安装驱动?

您可结合实际情况，直接创建已安装 GPU 驱动的实例，或在已有实例上自行安装对应 GPU 驱动：

创建已安装 GPU 驱动的实例

您可在创建 GPU 实例时，通过以下3种方式直接使用已安装 GPU 驱动的实例：

- 在购买页的“镜像”中，选择公共镜像，勾选后台自动安装GPU驱动即可预装相应版本驱动，建议您选择该方式。该方式仅支持部分 Linux 公共镜像，详情请参见 [各实例支持的 GPU 驱动版本及安装方式](#)。
- 在购买页的“镜像”中，若选择云镜像市场，则可选择已预装 GPU 驱动的镜像，详情请参见 [使用预装 GPU 驱动的镜像](#)。
- 选择已预装 GRID 驱动的公共镜像的 vGPU 实例，无需单独安装驱动。

在已有实例上安装 GPU 驱动

若您在创建 GPU 实例时未选择自动安装 GPU 驱动，或在公共镜像中没有您需要的操作系统或版本，为确保您能够正常使用 GPU 实例，请参见 [安装 NVIDIA 驱动指引](#) 和 [安装 NVIDIA GRID 驱动指引](#) 手动安装对应驱动。关于如何选择 GPU 驱动类型请参见：

- NVIDIA 系列 GPU 实例如用作通用计算，则需安装 Tesla Driver + CUDA，详情请参见 [安装 NVIDIA Tesla 驱动指引](#) 和 [安装 CUDA 驱动指引](#)。
- 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，详情请参见 [安装 NVIDIA GRID 驱动](#)。

GPU 云服务器如何计费?

GPU 云服务器提供四种购买方式：包年包月、按量计费、竞价实例和包销计费，分别适用于不同场景下的用户需求，详情可参见 [CVM 计费模式](#)。部分 GPU 机型可能只支持其中 2-3 种计费模式，以购买页实际显示为准。

GPU 云服务器是否支持调整实例配置?

GPU 云服务器 PNV4、GT4、GN10X/GN10Xp、GN6/GN6S、GN7、GN8、GNV4v、GNV4、GN7vw、GI1 支持同实例族下调整实例配置，其中 GN7 实例不支持直通型（整卡 GPU）调整为 vGPU 型（非单颗 GPU，例如 1/4 GPU）。GI3X 暂时不支持调整实例配置。

说明：

- 配置调整的前提、注意事项及操作指南，请参见 [调整实例配置](#)。
- 调整配置费用说明请参见 [实例调整配置费用说明](#)。

什么是 SSD 本地盘？

SSD 本地盘是来自云服务器所在物理机的本地存储，该类存储为实例提供块级别的数据访问能力，具有低时延、高随机 IOPS、高吞吐量的 I/O 能力。GPU 计算型搭载 SSD 本地盘，不支持硬件（CPU、内存）的升级，仅支持带宽的升级。

GPU 云服务器是否支持访问 CVM ？

支持，GPU 云服务器具有内网 IP 和 公网 IP，支持与 CVM 等其他云产品互通访问。

哪些 GPU 机型支持 HARP 网络协议？

所有的 GPU 实例类型均支持 HARP 网络协议，配置步骤可参见 [配置 HARP 分布式训练环境](#)。

为什么 GPU 云服务器 nvidia-smi 展示显存比实际 GPU 显存小？

Tesla 系列 GPU 默认开启了 ECC（error correcting code）功能，用来检查和纠正可能会在数据传输和存储过程中引发的比特错误，开启时会使用可用显存减少，并伴随部分性能损失。为提高数据的正确性，建议您保持开启状态。

使用自定义镜像创建多卡 GPU 实例，为什么 nvidia-smi 展示的 GPU 数目少于实际 GPU 数目？

由于制作自定义镜像时驱动会被固化，创建自定义镜像时的机器 GPU 卡数会影响使用自定义镜像创建的实例 GPU 是否可被正常加载，一般建议使用同实例规格或者同实例搭载更多张 GPU 卡的规格制作镜像，可使所有的 GPU 自动加载驱动。

如何查看 GPU 的驱动版本？

- 针对 Linux 操作系统的 GPU 服务器，您可以在参考 [NVIDIA GPU驱动安装指引](#) 安装驱动后登录服务器执行 `nvidia-smi` 命令查看 GPU 驱动版本，如下图红框所示，执行 `nvidia-smi` 命令后查看到的驱动版本为 550.144.03。

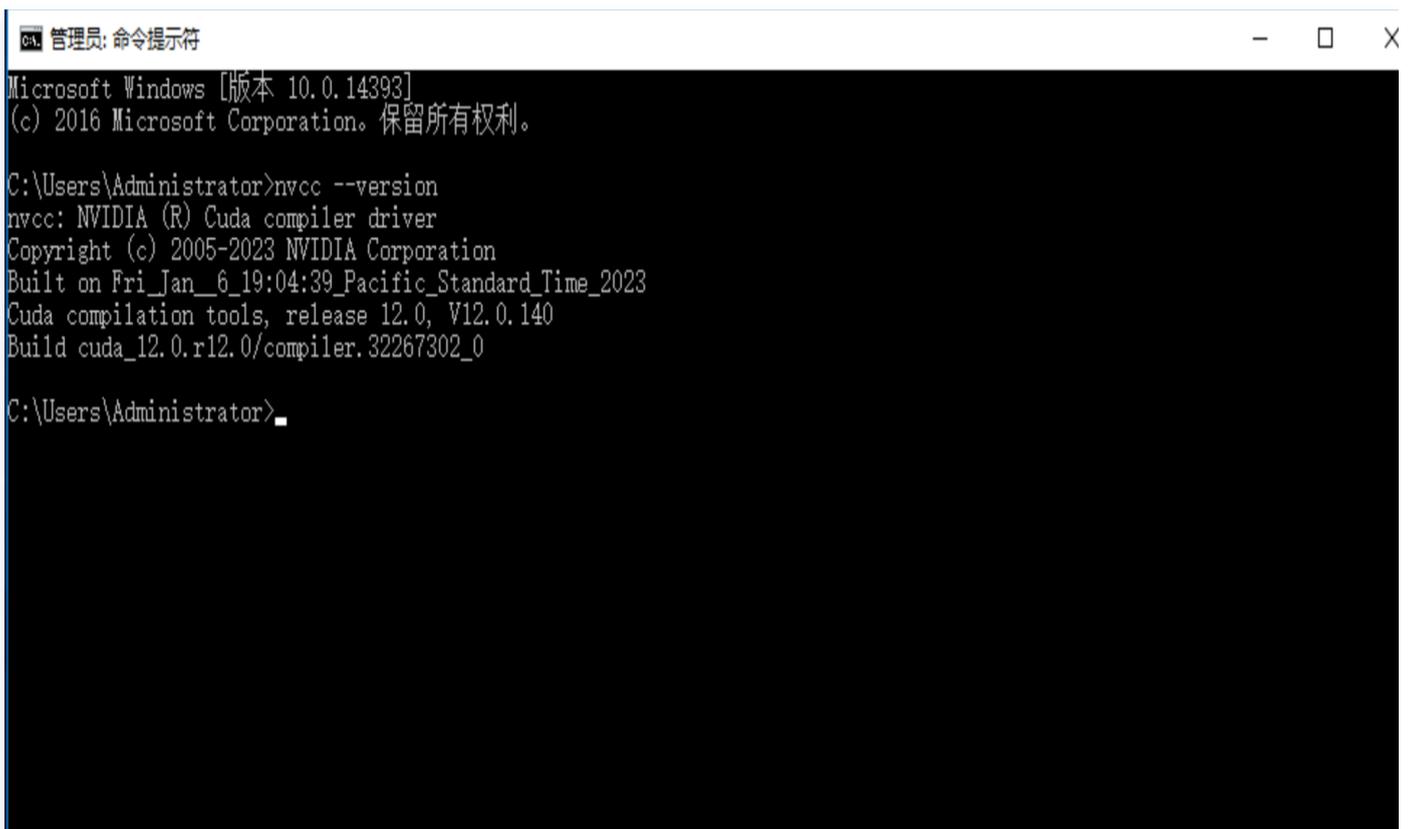

```
[root@... ~]# /usr/local/cuda/bin/nvcc --version
nvcc: NVIDIA (R) Cuda compiler driver
Copyright (c) 2005-2024 NVIDIA Corporation
Built on Thu_Mar_28_02:18:24_PDT_2024
Cuda compilation tools, release 12.4, V12.4.131
Build cuda_12.4.r12.4/compiler.34097967_0
```

若 root 用户执行如下命令配置了环境变量，则可执行 `nvcc --version` 命令获取 CUDA 版本信息。

```
echo 'export PATH=/usr/local/cuda/bin:$PATH' >/etc/profile.d/cuda.sh
source /etc/profile
```

```
[root@... ~]# echo 'export PATH=/usr/local/cuda/bin:$PATH' >/etc/profile.d/cuda.sh
[root@... ~]# source /etc/profile
[root@... ~]# nvcc --version
nvcc: NVIDIA (R) Cuda compiler driver
Copyright (c) 2005-2024 NVIDIA Corporation
Built on Thu_Mar_28_02:18:24_PDT_2024
Cuda compilation tools, release 12.4, V12.4.131
Build cuda_12.4.r12.4/compiler.34097967_0
```

- 针对 Windows 服务器可在参考文档 [GPU服务器安装 CUDA 驱动](#) 安装 CUDA 后在命令提示符中执行 `nvcc --version` 命令查看 CUDA 版本。如下图所示，CUDA 版本为12.0。



```
管理员: 命令提示符
Microsoft Windows [版本 10.0.14393]
(c) 2016 Microsoft Corporation。保留所有权利。

C:\Users\Administrator>nvcc --version
nvcc: NVIDIA (R) Cuda compiler driver
Copyright (c) 2005-2023 NVIDIA Corporation
Built on Fri_Jan_6_19:04:39_Pacific_Standard_Time_2023
Cuda compilation tools, release 12.0, V12.0.140
Build cuda_12.0.r12.0/compiler.32267302_0

C:\Users\Administrator>
```

为何在 nvidia-smi 命令查到的 CUDA 版本与 nvcc --version 命令查到的 CUDA 版本不一致？

CUDA (Compute Unified Device Architecture) 的软件堆栈由驱动层、运行时层和函数库层构成。CUDA 软件堆栈中涉及的 API 包括驱动层 API (Driver API) 和运行时层 API (Runtime Api)。

`nvidia-smi` 命令显示的 CUDA 版本是 GPU 驱动安装的驱动层 API 版本，此版本由 GPU 驱动指定。

`nvcc --version` 命令显示的 CUDA 版本是系统中实际安装的 CUDA Toolkit 的运行时层 API 版本。

两者可独立安装，所以可能存在版本不一致的情况。

为何 GPU 云服务器使用控制台中“后台自动安装 GPU 驱动功能”安装 GPU 驱动及 CUDA 后，存在 CUDA 驱动层 API 版本比 CUDA 运行时 API 版本低的情况？

CUDA 支持向前兼容，如果 GPU 驱动是 535 版本也支持 CUDA 12.4。

CUDA 的版本兼容性可参见 [CUDA 兼容性](#)。

为何 Windows 操作系统计算型 GPU 云服务器中安装驱动后在命令提示符中用 `nvidia-smi` 命令可以查看显卡信息，但在任务管理器中看不到？

因为计算型 GPU 云服务器默认不安装 GRID 驱动，GPU 模式为 TCC (Tesla Compute Cluster)，TCC 模式的 GPU 完全用于计算不支持图形显示功能，WDDM (Windows Display Driver Model) 模式的 GPU 支持计算和图形，仅在 GPU 云服务器安装了 GRID 驱动时才可以切换至 WDDM 模式，所以任务管理器中无 GPU 相关性能监控数据。Windows 操作系统 GPU 服务器中安装了 GPU GRID 驱动支持在任务管理器中查看 GPU 的性能监控数据。

Linux 操作系统 GPU 服务器如何关闭及开启 ECC Mode ？

GPU 服务器的 GPU 默认开启了 ECC Mode，ECC (Error Correcting Code) 是一种实现内存错误检查和纠正的技术，通过 GPU 驱动开启 ECC Mode 后，驱动会通过消耗一部分显存作为校验位来对数据损坏进行发现，可检测并纠正显存中的错误，显著降低因硬件故障或环境干扰导致的数据损坏风险，提升业务稳定性。

- 查看 ECC Mode

执行 `nvidia-smi -d ecc -q` 命令可查看 ECC Mode 是否开启，ECC 模式已开启则显示 "Current ECC Mode : Enabled"，未开启则显示 "Current ECC Mode : Disabled"。

```
[root@... ~]# nvidia-smi -d ecc -q

=====NVSMI LOG=====
Timestamp                : 2025-08-15 15:00:00
Driver Version           : 470.182.03
CUDA Version             : 11.4
Attached GPUs            : 1
GPU 00000000:00:08.0
Ecc Mode
  Current                 : Enabled
  Pending                 : Enabled
ECC Errors
  Volatile
    SRAM Correctable     : 0
    SRAM Uncorrectable   : 0
    DRAM Correctable     : 0
    DRAM Uncorrectable   : 0
  Aggregate
    SRAM Correctable     : 0
    SRAM Uncorrectable   : 0
    DRAM Correctable     : 0
    DRAM Uncorrectable   : 0
```

● 关闭 ECC Mode

警告：虽然关闭 ECC 模式会释放部分 GPU 显存及提升性能，但可能会增加数据错误的风险及降低稳定性，请谨慎评估操作。

可执行 `nvidia-smi -e 0` 命令关闭及重启服务器生效，参见如下代码。

警告：重启服务器可能会造成业务中断，请谨慎评估重启操作对业务的影响。

```
nvidia-smi -e 0
reboot -f
```

```
[root@... ~]# nvidia-smi -e 0
Disabled ECC support for GPU 00000000:00:08.0.
All done.
Reboot required.
[root@... ~]# reboot -f
Rebooting.
```

重启完成后登录服务器执行 `nvidia-smi -d ecc -q` 命令查看 ECC Mode，可以看到已经是 Disabled 状态。

```
[root@... ~]# nvidia-smi -d ecc -q

=====NVSMI LOG=====
Timestamp                : Wed Mar 19 10:39:52 2025
Driver Version           : 470.182.03
CUDA Version             : 11.4

Attached GPUs            : 1
GPU 00000000:00:08.0
  Ecc Mode
  Current                 : Disabled
  Pending                 : Disabled
  ECC Errors
  Volatile
    SRAM Correctable     : N/A
    SRAM Uncorrectable   : N/A
    DRAM Correctable     : N/A
    DRAM Uncorrectable   : N/A
  Aggregate
    SRAM Correctable     : N/A
    SRAM Uncorrectable   : N/A
    DRAM Correctable     : N/A
    DRAM Uncorrectable   : N/A
```

• 开启 ECC Mode

警告：调整ECC Mode 需重启服务器生效，重启服务器可能会造成业务中断，请谨慎评估重启操作对业务的影响。

可执行 `nvidia-smi -e 1` 命令开启及重启服务器生效，参见如下代码。

```
nvidia-smi -e 1
reboot -f
```

```
[root@... ~]# nvidia-smi -e 1
ECC support is already Enabled for GPU 00000000:00:08.0.
All done.
[root@... ~]# reboot -f
Rebooting.
```

重启完成后执行 `nvidia-smi -d ecc -q` 命令查看 ECC Mode，可以看到已经是 Enabled 状态。

```
[root@ ~]# nvidia-smi -d ecc -q
=====NVSMI LOG=====
Timestamp                : Wed Mar 19 10:42:13 2025
Driver Version           : 470.182.03
CUDA Version             : 11.4

Attached GPUs            : 1
GPU 00000000:00:08.0
  Ecc Mode
  Current                 : Enabled
  Pending                 : Enabled
  ECC Errors
  Volatile
    SRAM Correctable     : 0
    SRAM Uncorrectable   : 0
    DRAM Correctable     : 0
    DRAM Uncorrectable   : 0
  Aggregate
    SRAM Correctable     : 0
    SRAM Uncorrectable   : 0
    DRAM Correctable     : 0
    DRAM Uncorrectable   : 0
```

为什么在 GPU 云服务器上关闭/开启 ECC Mode 不生效?

通过执行 `nvidia-smi -e 0/1` 命令及重启服务器后调整 ECC Mode 可能会有配置失败的情况，因为 NVIDIA Tesla 535 及以上版本驱动，Legacy Persistence Mode 支持不佳，在使用 `nvidia-smi -pm 1` 命令开启常驻模式时可能会使 ECC Mode 调整失效，建议使用 `nvidia-smi -pm 0` 命令关闭常驻模式，使用 `nvidia-persistenced --persistence-mode` 命令开启常驻模式再调整 ECC Mode。

另外，也可能是驱动本身的问题，可以重装或切换驱动版本再调整 ECC Mode。

GPU 服务器如何查看 GPU 的 BusID?

参考 [NVIDIA GPU驱动安装指引](#) 安装驱动后登录服务器在 Windows 命令提示符或 Linux Shell 中执行 `nvidia-smi` 命令即可查看 GPU 的 BusID，如下图所示 BusID 为标红部分00:08.0。

```
[root@ ~]# nvidia-smi
Tue Mar 18 23:30:09 2025

+-----+
| NVIDIA-SMI 550.144.03           Driver Version: 550.144.03   CUDA Version: 12.4   |
+-----+-----+
| 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   36C    P8             9W / 70W   | 1MiB / 15360MiB |      0%      Default |
+-----+-----+

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