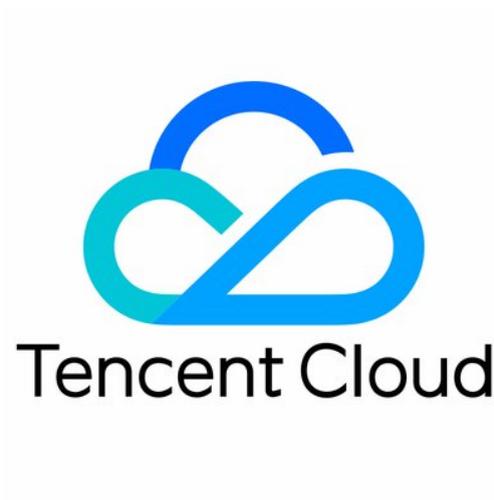


GPU 云服务器

产品简介



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



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

产品简介

产品概述

产品优势

应用场景

使用须知

产品简介

产品概述

Last updated: 2024-10-28 14:52:53

GPU 云服务器 (Cloud GPU Service) 是基于 GPU 的快速、稳定、弹性的计算服务，主要应用于深度学习训练/推理、图形图像处理以及科学计算等场景。GPU 云服务器提供和标准 CVM 云服务器一致的方便快捷的管理方式。GPU 云服务器通过其强大的快速处理海量数据的计算性能，有效解放用户的计算压力，提升业务处理效率与竞争力。

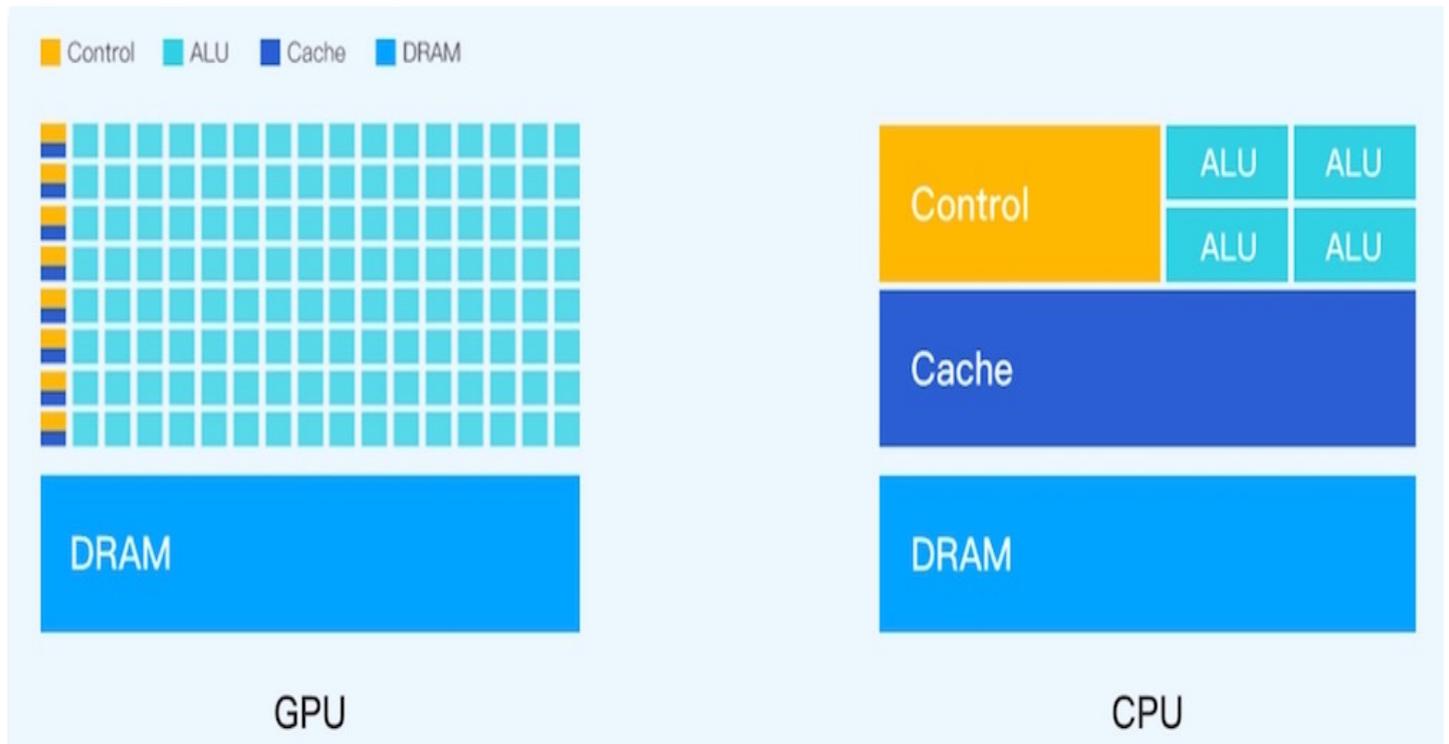
为什么选择 GPU 云服务器

GPU 云服务器与自建 GPU 服务器对比：

优势	GPU 云服务器	自建 GPU 服务器
弹性	<ul style="list-style-type: none"> 只需几分钟时间即可轻松获取一个或若干个高性能计算实例。 可按需灵活定制，一键升级到更高性能和容量的实例规格，实现快速、平滑扩容，满足业务快速发展需要。 	<p>机器固定配置，难以满足变化的需求。</p>
性能	<ul style="list-style-type: none"> 透传 GPU 性能，高效发挥 GPU 性能。 单机峰值计算能力突破 125.6T Flops 单精度浮点运算，62.4T Flops 双精度浮点运算。 	<ul style="list-style-type: none"> 用户手工容灾，依赖于硬件健壮性。 数据物理单点，数据安全不可控。
易用	<ul style="list-style-type: none"> 与腾讯云云服务器 CVM、负载均衡 CLB 等多种云产品无缝接入，内网流量免费。 和云服务器 CVM 采用一致的管理方式，无需跳板机登录，简单易用。 清晰的 GPU 驱动安装、部署指引，免去高学习成本。 	<ul style="list-style-type: none"> 购买装机管理，自行实现硬件扩展、驱动安装。 需跳板机登录，操作复杂。
安全	<ul style="list-style-type: none"> 不同用户间资源全面隔离，数据安全有保障。 完善的安全组和网络 ACL 设置让您能控制进出实例和子网的网络入出站流量并进行安全过滤。 与云安全无缝对接，享有云服务器同等的基础云安全基础防护和高防服务。 	<ul style="list-style-type: none"> 不同用户共享资源，数据不隔离。 需额外购买安全防护服务。
成本	<ul style="list-style-type: none"> 提供包月购买方式，无需大量资金投资购置物理服务器。 	<ul style="list-style-type: none"> 高服务器投资运营成本。 设备高功耗，需硬件改造适配。

- 硬件跟随主流 GPU 更新步伐，免除硬件更新带来的置换烦恼。
- 服务器运维成本低，无需预先采购、准备硬件资源，有效降低基础设施建设投入。
- 为保障服务稳定需高额的运维 IT 成本。

GPU 云服务器与 CPU 云服务器对比：



维度	GPU	CPU
核心数量	数千个加速核心（双卡 M40 高达6144个加速核心）	几十个核心
产品特点	<ol style="list-style-type: none"> 1. 高效众多的运算单元（ALU）支持并行处理 2. 多线程以到达超大并行吞吐量 3. 简单的逻辑控制 	<ol style="list-style-type: none"> 1. 复杂的逻辑控制单元 2. 强大的算数运算单元（ALU） 3. 复杂的逻辑控制
适用场景	计算密集、易于并行的程序	逻辑控制、串行运算的程序

产品优势

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

实例性能卓越可靠

实时加速计算

GPU 云服务器具有超强的计算性能：

- 采用主流的 GPU 和 CPU。
- 提供了强大的单双精度浮点运算能力，单机峰值计算能力最高突破：单精度浮点运算125.6T Flops，双精度浮点运算62.4T Flops。

服务稳定安全

GPU 云服务器提供安全可靠的网络环境和完善的防护服务：

- GPU 云服务器位于25G网络环境中（部分10G网络），内网时延低，提供优秀的计算能力。
- 支持和 [云服务器 CVM](#)、[私有网络 VPC](#)、[负载均衡 CLB](#) 等的业务对接，不增加额外的管理和运维成本，内网流量免费。
- 完善的 [安全组](#) 和 [网络 ACL](#) 设置，让您能控制进出实例和子网的网络入出站流量，并进行安全过滤。
- 与云安全无缝对接，享有云服务器同等的基础云安全基础防护和高防服务，详情请参见 [了解网络与安全性详情](#)。

实例部署迅速

- GPU 云服务器支付流程简单，即买即用，详情请参见 [购买指导](#)。
- GPU 云服务器入门简单，用户可以迅速搭建一个 GPU 实例，并且与云服务器 CVM 采用一致的管理方式，无需跳板机登录，简单易用，详情请参见 [快速入门](#)。
- GPU 云服务器与 [负载均衡 CLB](#)、[云硬盘](#) 等多种云产品无缝接入，清晰的 [NVIDIA 显卡驱动的安装、部署指引](#)，无需硬件扩展、驱动安装。

应用场景

Last updated: 2024-12-24 15:54:53

GPU 云服务器具有超强的并行计算能力，适用于 AI 训练、AI 推理、科学计算、云游戏、云渲染、视频编解码等场景。

AI 训练

GPU 计算型实例 GN10Xp 和 GT4 分别搭载了 V100 和 A100 GPU，它们具备强大的双精度浮点运算能力，并配备了 300~600GB/s 的 NVLink，适用于大规模 AI 训练加速。

AI 推理

GPU 计算型实例 PNV4、GN7 分别搭载 A10、T4 GPU，提供高性能功耗比的单精度浮点运算能力，广泛应用于在线和离线推理业务场景。

NPU 计算型 PTX1 搭载紫霄 C100 加速卡，适用于 AI 推理场景，在 CV、OCR、NLP 等场景性价比优越。

科学计算

GPU 计算型实例 GN10Xp、GT4 分别搭载 V100、A100 GPU，它们具备大显存、高算力以及低时延的卡间互联特性，能够为计算流体力学、地震分析和基因组学等科学计算场景提供大规模并行计算能力。

云游戏

GPU 渲染型 GNV4v 和 GN7vw 分别搭载 A10 和 T4 GPU，配备了适用于云游戏渲染场景的 GRID 驱动。它们提供 1/2 和 1/4 的 GPU 切分规格，旨在确保高画质和低延迟的前提下，优化单路游戏的画质成本。

GPU 渲染型 GA3 搭载腾讯云星星海 GA01，专为云游戏场景研发，可实现 1/6 细粒度切分，提供超高性价比的云游戏解决方案。

云渲染

GPU 渲染型 GNV4 和 GN7vw 分别搭载 A10 和 T4 GPU，并配备了适用于渲染场景的 GRID 驱动程序，能够实现图片和影视的快速渲染，提高影视动画、工业设计等场景的在线和离线渲染效率。

视频编解码

视频增强型 GN7vi 搭载 T4 GPU，配置了腾讯自研的明眸视频融合 AI 技术，包括极速高清编解码引擎和画质增强工具包，适用于点播和直播场景。

使用须知

Last updated: 2024-09-20 19:26:01

说明:

GPU 云服务器作为 CVM 云服务器的一类特殊实例，购买、操作、维护等方式与 CVM 云服务器一致，详细说明请参见 [云服务器文档](#)。

为了更好的使用 GPU 云服务器，请仔细阅读并关注以下使用注意事项：

1. 备份数据

GPU 云服务器提供强大的计算能力。GN8 实例可选择搭载本地 SSD 硬盘，但为避免极端情况下的数据丢失，请务必定期备份数据，以保证数据的安全性。

为保证数据更加安全可靠，也可以单独购买弹性云盘并挂载。

2. 及时续费

GPU 云服务器到期前7天开始，用户会收到到期提醒，如需继续使用请及时续费。否则，到期时实例会被关机断网并进入回收站，请务必确保在服务到期前及时续费或备份数据。

3. 外接设备

GPU 云服务器暂不支持直接加载外接硬件设备，如硬件加密狗，U 盘，外接硬盘，银行 U key 等。

4. 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](#)。

5. 禁止说明

- 禁止使用 GPU 云服务器做流量穿透服务，情节严重则处以关停并锁定实例的处罚并清退处理。
- 禁止使用 GPU 云服务器针对淘宝等电商网站从事刷单、刷销量、刷广告等虚假网站交易的网络行为。