

GPU 云服务器

实例类型

产品文档



腾讯云

【版权声明】

©2013-2024 腾讯云版权所有

本文档著作权归腾讯云单独所有，未经腾讯云事先书面许可，任何主体不得以任何形式复制、修改、抄袭、传播全部或部分本文档内容。

【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。

【服务声明】

本文档意在向客户介绍腾讯云全部或部分产品、服务的当时的整体概况，部分产品、服务的内容可能有所调整。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模式的承诺或保证。

文档目录

实例类型

 计算型实例

 渲染型实例

实例类型

计算型实例

最近更新时间：2024-02-07 16:13:49

GPU 计算型实例能够提供强大的计算能力，从容应对高实时、高并发的海量计算场景，广泛适用于深度学习、科学计算等 GPU 通用计算场景。腾讯云 GPU 云服务器以和 [云服务器 CVM](#) 一致的管理方式，提供快速、稳定、弹性的计算服务。

适用场景

适用于 AI 计算、高性能计算场景，例如：

AI 计算

深度学习推理

深度学习训练

科学计算/高性能计算

流体动力学

分子建模

气象工程

地震分析

基因组学等

说明：

若您的 GPU 实例用于 3D 图形渲染任务，则建议您使用已配置 vDWs/vWs License 并安装 GRID driver 的 [渲染型实例](#)，以此免除手动配置 GPU 图形图像处理基础环境步骤。

计算型实例总览

GPU 云服务器计算型系列提供以下实例：

售卖情况	实例	GPU 类型	可用镜像	可用区域
主售	PNV4	NVIDIA A10	CentOS 7.2 及以上 Ubuntu 16.04 及以上	广州、上海、北京
	GT4	NVIDIA A100 NVLink 40G	Windows Server 2016 及以上	广州、上海、北京、南京
	GN10Xp	NVIDIA Tesla	CentOS 7.2 及以上	广州、上海、北京、南京、成都、

		V100 NVLink 32G	Ubuntu 14.04 及以上 Windows Server 2012 及以上	重庆、新加坡、孟买、硅谷、法兰克福
	GN7	NVIDIA Tesla T4		广州、上海、南京、北京、成都、重庆、香港、新加坡、曼谷、雅加达、孟买、首尔、东京、硅谷、弗吉尼亚、法兰克福、圣保罗
	GN7vi	NVIDIA Tesla T4	CentOS 7.2 - 7.9 Ubuntu 14.04 及以上	上海、南京
在售	GI3X	NVIDIA Tesla T4		广州、上海、北京、南京、成都、重庆
在售	GN10X	NVIDIA Tesla V100 NVLink 32G	CentOS 7.2 及以上 Ubuntu 14.04 及以上 Windows Server 2012 及以上	广州、上海、北京、南京、成都、重庆、新加坡、硅谷、法兰克福、孟买
在售	GN8	NVIDIA Tesla P40		广州、上海、北京、成都、重庆、香港、硅谷
在售	GN6GN6S	NVIDIA Tesla P4		GN6：成都 GN6S：广州、上海、北京

说明：

可用区域：精确到城市级，细分区域详见下文中的实例配置信息。

计算型实例选型推荐

腾讯云提供了类型丰富的 GPU 计算实例，可满足不同业务应用场景的需求。请参考下表，并结合实际需求选择合适的计算实例。

GPU 云服务器计算型实例选型推荐如下表，其中 ✓ 为支持，★ 为推荐。

功能\实例	PNV4	GT4	GN10Xp	GN7	GN7vi	GI3X	GN10X	GN8	GN6GN6S
图形图像处理	✓	-	✓	✓	✓	✓	✓	✓	✓
视频编解码	✓	-	✓	★	★	★	✓	✓	✓
深度学习训练	✓	★	★	✓	✓	✓	★	★	✓

深度学习推理	★	✓	★	★	★	★	★	✓	✓
科学计算	-	★	★	-	-	-	★	-	-

注意：

以上推荐用途仅供参考，请根据实际需要进行选择。

NVIDIA 系列 GPU 实例如用作通用计算，则需安装 Tesla Driver + CUDA，安装方法请参考 [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。

支持范围

支持 [按量计费](#)。

支持在 [私有网络](#) 中启动。

支持与 [负载均衡](#) 等产品的业务对接，不增加额外的管理和运维成本，内网流量免费。

实例规格

计算型 PNV4

计算型 PNV4 不仅适用于深度学习等 GPU 通用计算场景，也适用于图形图像处理（3D 渲染，视频编解码）场景。

适用场景

性价比高，适用于如下场景：

深度学习的推理场景和小规模训练场景。例如：

大规模部署的 AI 推理

深度学习小规模训练

图形图像处理场景。例如：

图形图像处理

视频编解码

图形数据库

可用区

PNV4 实例支持可用区为：广州七区、上海四/五区、北京六区。

硬件规格

CPU：2.55GHz AMD EPYCTM Milan 处理器，睿频3.5GHz。

GPU：NVIDIA® A10（62.5 TFLOPS 单精度浮点计算，250 INT8 TOPS，500 INT4 TOPS）。

存储：可选择 [云硬盘类型](#)，如需 [扩容](#) 可新建弹性云盘进行挂载。

网络：默认网络优化，实例网络性能与规格对应。[公网网络](#) 可按需配置。

PNV4 实例提供以下配置：

型号	GPU (NVIDIA A10)	GPU 显存 (GDDR6)	vCPU	内存 (DDR4)	内网带宽	网络收发包 (PPS)	队列数
PNV4.7XLARGE116	1颗	1 * 24GB	28核	116GB	13Gbps	230万	28
PNV4.14XLARGE232	2颗	2 * 24GB	56核	232GB	25Gbps	470万	48
PNV4.28XLARGE466	4颗	4 * 24GB	112核	466GB	50Gbps	950万	48
PNV4.56XLARGE932	8颗	8 * 24GB	224核	932GB	100Gbps	1900万	48

计算型 GT4

计算型 GT4 适用于深度学习、科学计算等 GPU 通用计算场景。

适用场景

GT4 具有强大的双精度浮点运算能力，适用于大规模深度学习训练、推理和科学计算场景。例如：

深度学习

高性能数据库

计算流体动力学

计算金融

地震分析

分子建模

基因组学及其他

可用区

GT4 实例支持可用区为：广州三/四/六区、上海四/五区、北京五/六区、南京一区。

硬件规格

CPU：GT4 配置 AMD EPYCTM ROME 处理器，主频2.6GHz。

GPU：NVIDIA® A100 NVLink 40GB（19.5TFLOPS单精度浮点计算，9.7TFLOPS双精度浮点计算，600GB/s NVLink）。

内存：DDR4，内存计算性能稳定。

存储：可选择 [云硬盘类型](#)，如需 [扩容](#) 可新建弹性云盘进行挂载。

网络：最高可支持50Gbps内网带宽，超高网络收发包能力，实例网络性能与规格对应。[公网网络](#) 可按需配置。

GT4 实例提供以下配置：

型号	GPU (NVIDIA Tesla A100 NVLink 40G)	GPU 显 存 (HBM2)	vCPU	内存 (DDR4)	内网带宽	网络收 发包 (PPS)	队列数
GT4.4XLARGE96	1颗	1 * 40GB	16核	96GB	5Gbps	120万	4
GT4.8XLARGE192	2颗	2 * 40GB	32核	192GB	10Gbps	235万	8
GT4.20XLARGE474	4颗	4 * 40GB	82核	474GB	25Gbps	600万	16
GT4.41XLARGE948	8颗	8 * 40GB	164核	948GB	50Gbps	1200万	32

说明：

GPU 驱动：NVIDIA A100 系列需要安装 NVIDIA Tesla 450 版本以上驱动，推荐您安装 460.32.03(Linux)/461.33(Windows) 版本驱动，驱动版本信息请参见 [NVIDIA 官方文档](#)。

计算型 GN10Xp

计算型 GN10Xp 不仅适用于深度学习、科学计算等 GPU 通用计算场景，也适用于图形图像处理（3D 渲染，视频编解码）场景。

适用场景

GN10Xp 具有强大的双精度浮点运算能力，适用于如下场景：

大规模深度学习训练，推理和科学计算场景。例如：

深度学习

高性能数据库

计算流体动力学

计算金融

地震分析

分子建模

基因组学及其他

图形图像处理场景。例如：

图形图像处理

视频编解码

图形数据库

可用区

GN10Xp 实例支持可用区为：广州三/四区、上海二/三区、南京一区、北京四/五/七区、成都一区、重庆一区、新加坡一区、孟买二区、硅谷二区、法兰克福一区。

硬件规格

CPU：GN10Xp 配置 Intel® Xeon® Platinum 8255C CPU，主频2.5GHz。

GPU：NVIDIA® Tesla® V100 NVLink 32GB（15.7TFLOPS 单精度浮点计算，7.8TFLOPS 双精度浮点计算，125TFLOPS Tensor Core 深度学习加速，300GB/s NVLink）。

内存：DDR4，内存带宽达2666MT/s。

存储：可选择 [云硬盘类型](#)，如需 [扩容](#) 可新建弹性云盘进行挂载。

网络：默认网络优化，实例网络性能与规格对应。[公网网络](#) 可按需配置。

GN10Xp 实例提供以下配置：

型号	GPU (NVIDIA Tesla V100 NVLink 32G)	GPU 显 存 (HBM2)	vCPU	内存 (DDR4)	内网带 宽	网络收 发包 (PPS)	队列 数
GN10Xp.2XLARGE40	1颗	1 * 32GB	10核	40GB	3Gbps	80万	2
GN10Xp.5XLARGE80	2颗	2 * 32GB	20核	80GB	6Gbps	150万	5
GN10Xp.10XLARGE160	4颗	4 * 32GB	40核	160GB	12Gbps	250万	10
GN10Xp.20XLARGE320	8颗	8 * 32GB	80核	320GB	24Gbps	490万	16

计算型 GN7

NVIDIA 实例 GN7 不仅适用于深度学习等 GPU 通用计算场景，也适用于图形图像处理（3D 渲染，视频编解码）场景。

适用场景

性价比高，适用于如下场景：

深度学习的推理场景和小规模训练场景。例如：

大规模部署的 AI 推理

深度学习小规模训练

图形图像处理场景。例如：

图形图像处理

视频编解码

图形数据库

可用区

GN7 实例支持可用区为：广州三/四/六/七区、上海二/三/四/五区、南京一/二/三区、北京三/五/六/七区、成都一区、重庆一区、香港二区、新加坡一/二/三区、曼谷二区、雅加达二区、孟买二区、首尔一/二区、东京二区、硅谷二区、法兰克福一区、弗吉尼亚二区、圣保罗一区。

硬件规格

CPU：Intel® Xeon® Platinum 8255C CPU，主频 2.5 GHz。

GPU：NVIDIA® Tesla® T4（8.1 TFLOPS 单精度浮点计算，130 INT8 TOPS，260 INT4 TOPS）。

内存：DDR4，内存带宽达2666MT/s。

存储：可选择 [云硬盘类型](#)，如需 [扩容](#) 可新建弹性云盘进行挂载。

网络：默认网络优化，实例网络性能与规格对应。[公网网络](#) 可按需配置。

GN7 实例提供以下配置：

型号	GPU (NVIDIA Tesla T4)	GPU 显存 (HBM2)	vCPU	内存 (DDR4)	内网带宽	网络收发包 (PPS)	队列数
GN7.2XLARGE32	1颗	1 * 16GB	8核	32GB	3Gbps	60万	8
GN7.5XLARGE80	1颗	1 * 16GB	20核	80GB	7Gbps	140万	10
GN7.8XLARGE128	1颗	1 * 16GB	32核	128GB	10Gbps	240万	16
GN7.10XLARGE160	2颗	2 * 16GB	40核	160GB	13Gbps	280万	20
GN7.20XLARGE320	4颗	4 * 16GB	80核	320GB	25Gbps	560万	32

视频增强型 GN7vi

NVIDIA 实例 GN7vi 是在 GN7 基础上配置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、直播场景。使用该实例，您可在实例内部使用腾讯云自研的极速高清编解码和 AI画质增强功能。

可用区

GN7vi 实例支持可用区为：上海二/三/四/五区、南京一/二区。

硬件规格

CPU：Intel® Xeon® Platinum 8255C CPU，主频2.5GHz。

GPU：NVIDIA® Tesla® T4（8.1TFLOPS 单精度浮点计算，130 INT8 TOPS，260 INT4 TOPS）。

内存：DDR4，内存带宽达2666MT/s。

存储：可选择 [云硬盘类型](#)，如需 [扩容](#) 可新建弹性云盘进行挂载。

网络：默认网络优化，实例网络性能与规格对应。[公网网络](#) 可按需配置。

GN7vi 实例提供以下配置：

型号	GPU (NVIDIA Tesla T4)	GPU 显存 (HBM2)	vCPU	内存 (DDR4)	内网带宽	网络收发包 (PPS)	队列数
GN7vi.5XLARGE80	1颗	1 * 16GB	20核	80GB	6Gbps	140万	20
GN7vi.10XLARGE160	2颗	2 * 16GB	40核	160GB	13Gbps	280万	32
GN7vi.20XLARGE320	4颗	4 * 16GB	80核	320GB	25Gbps	560万	32

推理型 GI3X

NVIDIA 实例 GI3X 适用于深度学习等 GPU 通用计算场景，也适用于图形图像处理（3D 渲染，视频编解码）场景。

适用场景

性价比高，适用于如下场景：

深度学习的推理场景和小规模训练场景。例如：

大规模部署的 AI 推理

深度学习小规模训练

图形图像处理场景。例如：

图形图像处理

视频编解码

图形数据库

可用区

GI3X 实例支持可用区为：广州三区、上海四/五区、南京一/二区、北京五/六区、成都一区、重庆一区。

硬件规格

CPU：2.6GHz AMD EPYC™ ROME 处理器，睿频3.3GHz。

GPU：NVIDIA® Tesla® T4（8.1 TFLOPS 单精度浮点计算，130 INT8 TOPS，260 INT4 TOPS）。

内存：搭配最新八通道 DDR4，内存计算性能稳定。

存储：可选择 [云硬盘类型](#)，如需 [扩容](#) 可新建弹性云盘进行挂载。

网络：默认网络优化，实例网络性能与规格对应。[公网网络](#) 可按需配置。

GI3X 实例提供以下配置：

型号	GPU (NVIDIA Tesla T4)	GPU 显存 (GDDR6)	vCPU	内存 (DDR4)	内网带宽	网络收发包 (PPS)	队列数
GI3X.8XLARGE64	1 颗	1 * 16GB	32核	64GB	5Gbps	140万	8
GI3X.22XLARGE226	2颗	2 * 16GB	90核	226GB	13Gbps	375万	16
GI3X.45XLARGE452	4颗	4 * 16GB	180核	452GB	25Gbps	750万	32

计算型 GN10X

计算型 GN10X 不仅适用于深度学习、科学计算等 GPU 通用计算场景，也适用于图形图像处理（3D 渲染，视频编解码）场景。

适用场景

GN10X具有强大的双精度浮点运算能力，适用于如下场景：

大规模深度学习训练，推理和科学计算场景。例如：

深度学习

高性能数据库

计算流体动力学

计算金融

地震分析

分子建模

基因组学及其他

图形图像处理场景。例如：

图形图像处理

视频编解码

图形数据库

可用区

GN10X 实例支持可用区为：广州三/四区、上海二/三区、南京一区、北京四/五/七区、成都一区、重庆一区、新加坡一区、硅谷二区、法兰克福一区、孟买二区。

硬件规格

CPU：GN10X 配置 Intel® Xeon® Gold 6133 CPU，主频2.5GHz。

GPU：NVIDIA® Tesla® V100 NVLink 32GB（15.7TFLOPS 单精度浮点计算，7.8TFLOPS 双精度浮点计算，125TFLOPS Tensor Core 深度学习加速，300GB/s NVLink）。

内存：DDR4，内存带宽达2666MT/s。

存储：可选择 [云硬盘类型](#)，如需 [扩容](#) 可新建弹性云盘进行挂载。

网络：默认网络优化，实例网络性能与规格对应。[公网网络](#) 可按需配置。

GN10X 实例提供以下配置：

型号	GPU (NVIDIA Tesla V100 NVLink 32G)	GPU 显 存 (HBM2)	vCPU	内存 (DDR4)	内网带 宽	网络收 发包 (PPS)	队列 数
GN10X.2XLARGE40	1颗	1 * 32GB	8核	40GB	3Gbps	80万	2
GN10X.9XLARGE160	4颗	4 * 32GB	36核	160GB	13Gbps	250万	9
GN10X.18XLARGE320	8颗	8 * 32GB	72核	320GB	25Gbps	490万	16

计算型 GN8

NVIDIA 实例 GN8 不仅适用于深度学习等 GPU 通用计算场景，也适用于图形图像处理（3D 渲染，视频编解码）场景。

适用场景

适用于如下场景：

深度学习的推理和训练场景。例如：

大吞吐量的 AI 推理

深度学习

图形图像处理场景。例如：

图形图像处理

视频编解码

图形数据库

可用区

GN8 实例支持可用区为：广州三区、北京二/四区、成都一区、香港二区、上海三区、重庆一区、硅谷一区。

硬件规格

CPU：Intel® Xeon® E5-2680 v4 CPU，主频2.4GHz。

GPU：NVIDIA® Tesla® P40（12TFLOPS 单精度浮点计算，47INT8 TOPS）。

内存：DDR4，内存带宽达2666MT/s。

存储：可选择 [云硬盘类型](#)，如需 [扩容](#) 可新建弹性云盘进行挂载。

网络：默认网络优化，实例网络性能与规格对应。[公网网络](#) 可按需配置。

GN8 实例提供以下配置：

型号	GPU (NVIDIA Tesla P40)	GPU 显 存 (HBM2)	vCPU	内存 (DDR4)	内网带宽	网络收 发包 (PPS)	队列 数
GN8.LARGE56	1颗	24GB	6核	56GB	1.5Gbps	45万	8
GN8.3XLARGE112	2颗	48GB	14核	112GB	2.5Gbps	50万	8
GN8.7XLARGE224	4颗	96GB	28核	224GB	5Gbps	70万	14
GN8.14XLARGE448	8颗	192GB	56核	448GB	10Gbps	70万	28

计算型 GN6/GN6S

NVIDIA 实例 GN6/GN6S 不仅适用于深度学习等 GPU 通用计算场景，也适用于图形图像处理（3D 渲染，视频编解码）场景。

适用场景

性价比高，适用于如下场景：

深度学习的推理场景和小规模训练场景。例如：

大规模部署的 AI 推理

深度学习小规模训练

图形图像处理场景。例如：

图形图像处理

视频编解码

图形数据库

可用区

GN6/GN6S 实例支持可用区为：

GN6：成都一区。

GN6S：广州三区、上海二，三，四区、北京四，五区。

硬件规格

CPU：GN6 配置 Intel® Xeon® E5-2680 v4 CPU，主频2.4GHz。GN6S 配置 Intel® Xeon® Silver 4110 CPU，主频2.1GHz。

GPU：NVIDIA® Tesla® P4（5.5TFLOPS 单精度浮点计算，22INT8 TOPS）。

内存：DDR4，内存带宽达2666MT/s。

存储：可选择 [云硬盘类型](#)，如需 [扩容](#) 可新建弹性云盘进行挂载。

网络：默认网络优化，实例网络性能与规格对应。[公网网络](#) 可按需配置。

GN6/GN6S 实例提供以下配置：

型号	GPU (NVIDIA Tesla P4)	GPU 显存 (HBM2)	vCPU	内存 (DDR4)	内网带宽	网络收发包 (PPS)	队列数
GN6.7XLARGE48	1颗	8GB	28核	48GB	5Gbps	120万	14
GN6.14XLARGE96	2颗	16GB	56核	96GB	10Gbps	120万	28
GN6S.LARGE20	1颗	8GB	4核	20GB	5Gbps	50万	8
GN6S.2XLARGE40	2颗	16GB	8核	40GB	9Gbps	80万	8

渲染型实例

最近更新时间：2024-01-11 17:11:13

GPU 渲染型实例 适用于 GPU 传统图形图像处理（3D 渲染）应用场景。腾讯云以和 [云服务器 CVM](#) 一致的管理方式，提供快速、稳定、弹性的计算服务。

适用场景

适用于高性能图形处理，3D 渲染。例如：

- 非线性编辑
- 云游戏
- 云手机
- 云桌面
- CloudXR
- 图形图像处理等

渲染型实例总览

GPU 云服务器渲染型系列提供以下实例：

实例	GPU 类型	可用镜像	可用区域
GNV4v	NVIDIA A10	Windows Server 2019 数据中心版 64位 中文版 GRID13	北京、上海、广州
GNV4	NVIDIA A10	CentOS 7.2 及以上 Ubuntu 16.04 及以上 Windows Server 2019 数据中心版 64位 中文版 GRID13	北京、上海、广州、重庆
GN7vw	NVIDIA Tesla T4	CentOS 8.0 64位 GRID11.1 Windows Server 2019 数据中心版 64位 中文版 GRID11.1	北京、上海、广州、南京、成都、重庆、香港、新加坡、孟买、硅谷、弗吉尼亚、法兰克福
GI1	Intel SG1	CentOS 7.6 64位 + SG1-pv1.3 CentOS 7.6 64位 + SG1-pv1.4	北京、上海、广州、南京、重庆

说明：

可用区域：精确到城市级，细分区域详见下文中的实例配置信息。

渲染型实例选型推荐

腾讯云提供了类型丰富的 GPU 计算实例，可满足不同业务应用场景的需求。请参考下表，并结合实际需求选择合适的计算实例。

GPU 云服务器渲染型实例选型推荐如下表，其中 ✓ 为支持，★为推荐。

功能\实例	GNV4v	GNV4	GN7vw	GI1
图形图像处理	★	★	★	★
视频编解码	★	★	★	★
深度学习训练	-	✓	-	-
深度学习推理	-	✓	-	-
科学计算	-	-	-	-

注意：

以上推荐用途仅供参考，请根据实际需要进行选择。

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。GNV4v、GNV4、GN7vw 实例可选择已预装 GRID 驱动的指定镜像，无需单独安装 GRID Driver 和配置 License Server。

GNV4v、GNV4、GN7vw 实例簇提供支持 vDWs/vWs 类型的 vGPU 实例类型，支持 DirectX 和 OpenGL 等图形 API。

支持范围

支持 [竞价实例](#) 和 [按量计费](#)。

支持在 [私有网络](#) 中启动。

支持与 [负载均衡](#) 等产品的业务对接，不增加额外的管理和运维成本，内网流量免费。

实例规格

渲染型 GNV4v

NVIDIA 实例 GNV4v 配置了 vDWS License 服务器并安装 GRID driver 的渲染型实例，适用于图形图像处理（3D 渲染，视频编解码）场景。使用该实例，您可免除手动配置 GPU 图形图像处理基础环境。

注意：

该实例暂时处于白名单开放，请您通过 [提交工单](#) 进行实例购买权限开通。

可用区

GNV4v 实例支持可用区为：广州七区、上海五区、北京六区。

硬件规格

CPU：2.55GHz AMD EPYCTM Milan 处理器，睿频3.5GHz。

GPU：NVIDIA® A10（62.5 TFLOPS 单精度浮点计算，250 INT8 TOPS，500 INT4 TOPS）。

存储：可选择 [云硬盘类型](#)，如需 [扩容](#) 可新建弹性云盘进行挂载。

网络：默认网络优化，实例网络性能与规格对应。[公网网络](#) 可按需配置。

GNV4v 实例提供以下配置：

型号	GPU (NVIDIA A10)	GPU 显存 (HBM2)	vCPU	内存 (DDR4)	内网带宽	网络收发包 (PPS)	队列数
GNV4v.XLARGE24	1/4颗	6GB vGPU	6核	24GB	3Gbps	50万	6
GNV4v.3XLARGE58	1/2颗	12GB vGPU	14核	58GB	7Gbps	110万	14
GNV4v.7XLARGE116	1颗	1 * 24GB	28核	116GB	13Gbps	230万	28

渲染型 GNV4

NVIDIA 实例 GNV4 配置了 vDWS License 服务器并安装 GRID driver 的渲染型实例，适用于图形图像处理（3D 渲染，视频编解码）场景。使用该实例，您可免除手动配置 GPU 图形图像处理基础环境。

注意：

该实例暂时处于白名单开放，请您通过 [提交工单](#) 进行实例购买权限开通。

可用区域

GNV4 实例支持可用区为：北京六区、上海五区、广州六区、重庆一区。

硬件规格

CPU：Intel® Xeon® Cooper Lake 处理器，基频 3.4 GHz，睿频3.8 GHz。

GPU：NVIDIA® A10（31.2 TFLOPS 单精度浮点计算，250 INT8 TOPS，500 INT4 TOPS）。

存储：可选择 [云硬盘类型](#)，如需 [扩容](#) 可新建弹性云盘进行挂载。

网络：默认网络优化，实例网络性能与规格对应。[公网网络](#) 可按需配置。

GNV4实例提供以下配置：

型号	GPU (NVIDIA A10)	GPU 显存 (GDDR6)	vCPU	内存 (DDR4)	内网带宽	网络收发包 (PPS)	队列数
GNV4.3XLARGE44	1颗	24GB	12核	44GB	2Gbps	53万	4

渲染型 GN7vw

NVIDIA 实例 GN7vw 是在 GN7 基础上配置 vDWS License 服务器并安装 GRID driver 的渲染型实例，适用于图形图像处理（3D 渲染，视频编解码）场景。使用该实例，您可免除手动配置 GPU 图形图像处理基础环境。

注意：

GPU 渲染型 GN7vw 现处于限量购买阶段。

可用区

GN7vw 实例支持可用区为：广州三，四区、上海二，四，五区、南京一，二区、北京五区、成都一区、重庆一区、香港二区、新加坡一区、孟买二区、硅谷二区、弗吉尼亚二区、法兰克福一区。

硬件规格

CPU：Intel® Xeon® Platinum 8255C CPU，主频 2.5 GHz。

GPU：NVIDIA® Tesla® T4（8.1 TFLOPS 单精度浮点计算，130 INT8 TOPS，260 INT4 TOPS）。

内存：DDR4，内存带宽达2666MT/s。

存储：可选择 [云硬盘类型](#)，如需 [扩容](#) 可新建弹性云盘进行挂载。

网络：默认网络优化，实例网络性能与规格对应。[公网网络](#) 可按需配置。

GN7vw实例提供以下配置：

型号	GPU (NVIDIA Tesla T4)	GPU 显存 (GDDR6)	vCPU	内存 (DDR4)	内网带宽	网络收发包 (PPS)	队列数
GN7vw.LARGE16	1/4颗	4GB vGPU	4核	16GB	2Gbps	50万	8
GN7vw.2XLARGE32	1/2颗	8GB vGPU	8核	32GB	4Gbps	80万	8
GN7vw.4XLARGE64	1颗	1 * 16GB	16核	64GB	7Gbps	150万	8

渲染型 GI1

GPU 渲染型 G11 实例搭载 H3C XG310 加速卡，单张加速卡包含 4 颗 Intel SG1 芯片，适用于 Android 云游戏、Android 云应用、视频转码处理等场景。

注意：

该实例暂时处于白名单开放，请您通过 [提交工单](#) 进行实例购买权限开通。

适用场景

- Android 云手机
- Android 云游戏
- Android 云应用
- 视频转码

可用区

G11 实例支持可用区为：北京六区、上海五区、广州七区、南京三区、重庆一区。

硬件规格

CPU： Intel® Xeon® Platinum 8255c CPU，主频2.5GHz。

GPU： Intel® SG1，采用 H3C XG310 加速卡，单张加速卡包含4颗 SG1 芯片。

存储： 可选择 [云硬盘类型](#)，如需 [扩容](#) 可新建弹性云盘进行挂载。

网络： 默认网络优化，实例网络性能与规格对应。[公网网络](#) 可按需配置。

G11 实例提供以下配置：

型号	GPU (Intel SG1)	GPU 显存	vCPU	内存 (DDR4)	内网带宽	网络收发包 (PPS)	队列数
G11.10XLARGE160	1*H3C XG310 (4颗 Intel SG1 芯片)	32GB (4* 8GB)	42核	160GB	13Gbps	250万	32
G11.21XLARGE320	2*H3C XG310 (8颗 Intel SG1 芯片)	64GB (8 * 8GB)	84核	320GB	25Gbps	600万	32