



NVIDIA RTX A6000 助力世界性能超强的工作站

为专业人士提供增强性能

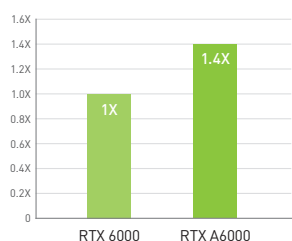
NVIDIA RTX™ A6000 在 NVIDIA Ampere 架构的基础上构建而成，可为设计师、工程师、科学家及艺术家提供处理图形和计算密集型工作流程所需的一切性能。RTX A6000 搭载最新一代的 RT Core、Tensor Core 和 CUDA® Core，可提供远超以往的渲染、AI、图形和计算性能。NVIDIA RTX 已通过众多专业应用认证，经由领先的独立软件供应商 (ISV) 和工作站制造商测试而成，并由一支覆盖全球的支持专家团队提供支持，是要求苛刻的企业部署的理想视觉计算解决方案。

规格

GPU 显存	48GB GDDR6
显存位宽	384 位
显存带宽	768 GB/s
纠错码 (ECC)	有
基于 NVIDIA Ampere 架构的 CUDA Core	10752
NVIDIA 第三代 Tensor Core	336
NVIDIA 第二代 RT Core	84
单精度性能	38.7 TFLOPS ⁷
RT Core 性能	75.6 TFLOPS ⁷
Tensor 性能	309.7 TFLOPS ⁸
NVIDIA NVLink	支持连接两个 NVIDIA RTX A6000 GPU ¹²
NVIDIA NVLink 带宽	112.5 GB/s (双向)
系统接口	PCI Express 4.0 x16
功耗	主板总功率：300 W
散热解决方案	主动冷却
外形尺寸	4.4" (高) x 10.5" (长)，双插槽，全高
显示器接口	4 个 DisplayPort 1.4a ⁹
最大支持显示器数量	4x 4096 x 2160 @ 120 Hz， 4x 5120 x 2880 @ 60 Hz， 2x 7680 x 4320 @ 60 Hz
电源连接器	1 个 8 针 CPU
编码 / 解码引擎	1 个编码引擎，2 个解码引擎 (+AV1 解码)
VR ready	支持
vGPU 软件支持	NVIDIA vPC/vApp、NVIDIA RTX 虚拟工作站、NVIDIA 虚拟计算服务器
vGPU 配置支持	1GB、2GB、3GB、4GB、6GB、8GB、12GB、16GB、24GB、48GB
图形 API	DirectX 12.0 ¹⁰ 、Shader Model 5.1 ¹⁰ 、OpenGL 4.6 ¹¹ 、Vulkan 1.18 ¹¹
计算 API	CUDA、DirectCompute、OpenCL™

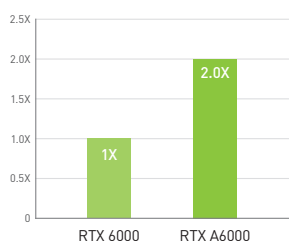
图形性能提升可高达 40%¹

SPECviewperf 2020



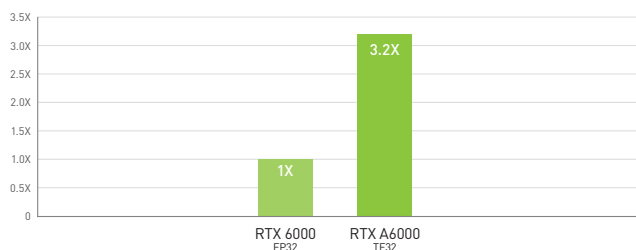
渲染性能提升可高达两倍²

Autodesk VRED

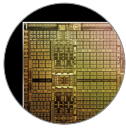


借助 TF32，开箱即用的 AI 训练性能可提升三倍以上³

BERT Large 训练

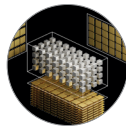


突破性创新



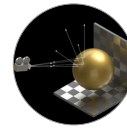
NVIDIA AMPERE 架构

NVIDIA® RTX™ 技术彻底变革了专业视觉计算。NVIDIA Ampere 架构建立于 RTX 的强大功能之上，可显著提高渲染、图形、AI 和计算工作负载的性能。NVIDIA Ampere 以追求完美为设计目标，并且包含先进的创新技术，可令 RTX 在处理专业工作负载时的表现更上一层楼。



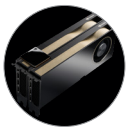
第三代 TENSOR CORE 技术

全新的 Tensor Float 32 (TF32) 精度可提供五倍于前代精度的训练吞吐量，无需更改代码即可加速 AI 及数据科学模型训练。可为结构化稀疏提供硬件支持，从而使推理吞吐量翻倍。Tensor Core 还为图形处理引入了 AI，为选定应用带来了 DLSS、AI 降噪和增强编辑等功能。



第二代 RT CORE 技术

第二代 RT Core 的吞吐量高达上一代的两倍，并能同时运行光线追踪和着色或降噪功能，从而大幅加快电影内容的逼真渲染和产品设计的虚拟原型创建等工作负载的运行速度。这项技术还可提升光线追踪动态模糊的渲染速度，从而更快获得结果，并增加视觉准确度。



第三代 NVLINK 技术

借助第三代 NVIDIA NVLink® 技术，用户可以连接两个 GPU 以共享 GPU 性能及显存。凭借高达 112 Gb/s 的双向带宽和高达 96GB 的组合显存，专业人士可以处理大型的渲染、AI、虚拟现实及视觉计算工作负载。新的 NVLink 连接器还具有更低矮的外形，可在更多型号的机箱中实现 NVLink 功能。



基于 NVIDIA AMPERE 架构的 CUDA CORE

NVIDIA Ampere 架构的 CUDA® Core 将单精度浮点 (FP32) 的运算速度提升了一倍，与 Turing GPU 相比，可带来高达两倍的能效提升。这为 3D 模型开发等图形工作流程，以及计算机辅助工程 (CAE) 桌面模拟等计算工作流程带来了显著的性能提升。



PCI EXPRESS 4.0

基于 NVIDIA Ampere 架构的 GPU 支持 PCI Express 4.0 (PCIe 4.0)，后者可提供两倍于 PCIe 3.0 的带宽。这提高了从 CPU 内存传输数据的速度，可更好地执行 AI 和数据科学等数据密集型任务。更快的 PCIe 性能还能加速 GPU 直接内存访问 (DMA) 传输，从而能让支持视频的设备通过 GPUDirect® 更快地传输视频数据，并利用 GPUDirect Storage 加快输入 / 输出 (I/O) 速度。

特性

- > PCI Express 4.0
- > 四个 DisplayPort 1.4a 接口
- > AV1 解码支持
- > 支持音频的 DisplayPort
- > VGA 支持⁴
- > 立体声接口，提供 3D 立体声支持
- > 支持 NVIDIA GPUDirect® for Video
- > 支持 NVIDIA 虚拟 GPU (vGPU) 软件
- > 兼容 NVIDIA Quadro® Sync II⁵
- > NVIDIA Quadro Experience™
- > 桌面管理软件
- > 支持 NVIDIA RTX IO
- > 支持 HDCP 2.2
- > NVIDIA Mosaic⁶ 技术

如需详细了解 NVIDIA RTX A6000，请访问 www.nvidia.com/rtx-a6000

1 已在工作站上进行测试，该工作站规格如下：配有 1 个 Xeon Gold 6154；功耗为 3GHz (3.7GHz Turbo)；搭载了 64 位 Win10 系统和 460.48 版本的 NVIDIA 驱动。SPECviewperf 2020，能效子测试。| 2 已在工作站上进行测试，该工作站规格如下：配有 2 个 Xeon Gold 6126；功耗为 2.6GHz (3.7GHz Turbo)；搭载 64 位 Win10 系统和 456.37 版本的 NVIDIA 驱动。Autodesk VRED 221.0 GA 版本。| 3 已在工作站上进行测试，该工作站规格如下：配有 AMD Ryzen 3900X；3.8GHz；4.6 Turbo；搭载 460.17 版本的 NVIDIA 驱动；使用 Pytorch 的 BERT 预训练吞吐量；第一阶段序列长度 128；使用 FP32 精度的 RTX 6000；使用 TF32 精度的 RTX A6000。| 4 通过适配器、接口或接头。| 5 Quadro Sync II 显卡单独销售。| 6 Windows 7、8、8.1、10 以及 Linux。| 7 基于 GPU 加速频率的峰值速率。| 8 使用新增稀疏性能的有效 teraFLOPS (TFLOPS)。| 9 Quadro RTX A6000 的显示屏端口默认为开启状态。显示屏端口在使用 vGPU 软件时处于未激活状态。| 10 GPU 支持 DX 12.0 API，及功能级别达 12+1 的硬件。| 11 产品根据已发布的 Khronos 规格设计而成，并且有望通过 Khronos 一致性测试流程（如果可用）。如需了解当前的一致性状况，请访问 www.khronos.org/conformance。| 12 NVIDIA NVLink 单独销售。

© 2021 NVIDIA Corporation. 保留所有权利。NVIDIA、NVIDIA 徽标、CUDA、GPUDirect、GRID、NVLink、Quadro、Quadro Experience 和 RTX 均为 NVIDIA Corporation 在美国和其他国家或地区的商标或注册商标。其他公司和产品名称可能是其各关联公司的商标。其他所有商标均为其各自所有者的资产。2021 年 1 月

