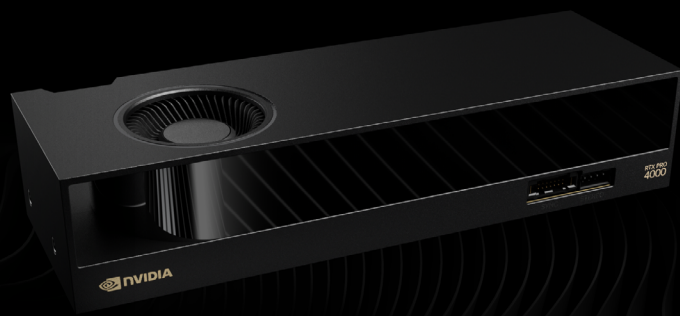




NVIDIA RTX PRO 4000 Blackwell SFF Edition

驱动 AI 新时代



NVIDIA Blackwell 助力紧凑型工作站实现强大功能

随着 AI 以惊人的速度不断发展，各行各业开始更多地依赖于即时、计算密集型任务（从生成式 AI 到超真实级渲染和基于物理的仿真），这时他们面临着双重挑战：为这些工具部署足够强大的基础设施，同时在空间受限的环境中运行。从制造车间到媒体工作室，企业需要紧凑型系统提供强大的性能，即解决方案能够兼顾原始算力与能效、散热控制，并可无缝集成到轻薄型塔式、便携式边缘系统或密集型机架式工作站。

对设计工作室、工程实验室和金融中心来说，这意味着加速大型 CAD 组件、AI 驱动式仿真和实时 3D 渲染，而不会牺牲桌面空间或可扩展性。无论是优化工厂布局，实现智慧城市可视化，还是支持 3D 动画和经 AI 升级的视觉效果，专业人员还需要解决方案能够突破传统 GPU 的局限性，同时仍提供所需性能来处理未来充满挑战的工作负载。

NVIDIA RTX PRO™ 4000 Blackwell SFF Edition 重新定义了紧凑型半高 GPU，将 NVIDIA 的革命性 Blackwell 架构与 24 GB 超高速 GDDR7 显存相结合，以小卡外形提供全面的性能。该平台旨在加速要求严苛的专业工作流，相比于同类产品，可提供令人惊叹的 AI 计算能力、新一代神经图形和无与伦比的能效。

RTX PRO 4000 SFF 搭载第五代 Tensor Core 和第四代 RT Core，可轻松处理复杂的大型数据集和多应用工作流。其 Tensor Core 可加速生成式 AI 和 LLM 推理，而 RT Core 则支持电影级光线追踪，从而助力专业人员对逼真的场景进行实时可视化。它采用时尚的半高小型设计，可确保无缝集成紧凑型工作站、边缘设备和先进的自动化系统。

RTX PRO 4000 SFF 作为功能最强大的半高工作站 GPU 而设计，可为设计工作室、工程公司、金融机构和工业应用提供非凡的通用性。

这些功能由 NVIDIA Blackwell 架构提供支持，这是加速计算领域的范式转变，融合了前所未有的 AI、光线追踪和神经网络渲染技术进步，重新定义了受真正的小型系统约束的专业工作流。

主要特性

- 专为神经网络着色器打造的增强型流式多处理器 (SM)
- 第 5 代 Tensor Core 支持 FP4 精度、DLSS 4 多帧生成
- 第 4 代 Ray Tracing Core 专为精细几何图形而打造
- 24 GB 的 GDDR7 显存
- 432 GB/s 的显存带宽
- 第 9 代 NVENC 和第 6 代 NVDEC，支持 4:2:2
- PCIe 5.0
- 四个 Mini DisplayPort 2.1b 接口
- AI 管理处理器

突破性创新

NVIDIA Blackwell 架构结合了 AI、光线追踪和神经网络渲染技术，并在性能和显存方面进行了大规模改进，可推动前沿的专业创意、设计和工程工作流程，并为未来十年的创新提供动力。

NVIDIA Blackwell 流式多核处理器：新的 SM 具有更高的处理吞吐量，并针对神经网络着色器进行了优化，支持将神经网络集成到可编程着色器中，以推动下一波 AI 增强图形创新。

第 5 代 Tensor Core：性能高达上一代的 3 倍，支持 FP4 精度，可缩短 AI 模型处理时间，同时减少显存占用。这支持 LLM 和生成式 AI 的本地微调。

第 4 代 Ray Tracing Core：RTX Mega Geometry 可将光线追踪三角形数量提高 100 倍，使光线与三角形求交计算速度比上一代产品提高了一倍，从而打造逼真、物理精准的场景和沉浸式 3D 设计。

新一代视频引擎：利用实时 AI 处理增强视频会议、视频制作和流媒体工作流程。第九代 NVENC 和第六代 NVDEC 引擎支持 4:2:2 10 位编码 / 解码，可实现高质量色彩保真度和流畅的多流 4K 工作流，满足先进创意和视频制作需求。

GDDR7 显存：最新一代 GPU 显存技术 GDDR7 在显存容量和带宽上都有显著提升，有助于更快运行应用，并处理更大、更复杂的数据集。凭借 24 GB 的 GPU 显存和 432 GB/s 的带宽，您可以处理大型 3D 和 AI 项目，在本地微调 AI 模型，探索大规模 VR 环境，并推动更大的多应用工作流。

DLSS 4：多帧生成技术可确保在逼真的模拟中实现超流畅的帧节奏。在支持的游戏引擎和 3D 渲染应用中体验快达 3 倍的帧率和令人惊叹的画质，从而获得更流畅、响应更灵敏的性能。

PCIe 5.0：支持 PCIe 5.0，带宽是 PCIe 4.0 的两倍，提高了 CPU 内存的数据传输速度，并为 AI、数据科学和 3D 建模等数据密集型任务解锁更快的性能。

DisplayPort 2.1b：实现无与伦比的视觉清晰度和性能，在 240 Hz 刷新率下驱动 8K 的高分辨率显示器，在 60 Hz 刷新率下驱动 16K 高分辨率的显示器。更高的带宽可实现无缝的多显示器设置，是多任务处理和协作的理想选择，而 HDR 和更高的色彩深度支持可确保视频编辑、3D 设计和直播等精密工作具有出色的色彩准确性。

企业级可靠性

专为需要满足最高要求的专业人员而设计，NVIDIA RTX PRO 提供无与伦比的性能、可靠性和支持。每个 GPU 都针对各种设计、工程和 AI 工作流进行了严格测试，并通过企业驱动持续优化。凭借广泛的 ISV 认证、强大的 IT 管理工具和企业级支持，RTX PRO 工作站是企业级和任务关键型工作值得信赖的选择。

规格	
GPU 架构	NVIDIA Blackwell
NVIDIA® CUDA® 核心	8,960
Tensor Cores	第 5 代
RT Core	第 4 代
GPU 显存	24 GB GDDR7 with ECC
显存接口	192 位
显存带宽	432 GB/s
系统接口	PCIe 5.0 x 8 ¹
显示器接口	4 x Mini DisplayPort 2.1b
最大同步显示器数量	➤ 4 x 3840 x 2160 @ 165 Hz ➤ 2 x 7680 x 4320 @ 100 Hz
视频引擎	2 x NVENC (第 9 代) 2 x NVDEC (第 6 代)
功耗	主板总功率：70 W
散热解决方案	主动
外形规格	6.86 厘米 x 16.77 厘米长，双插槽，半高
图形 API	Direct X 12、Shader Model 6.7、OpenGL 4.6 ² 、Vulkan 1.4 ²
计算 API	CUDA 12.8、OpenCL 3.0

准备好开始了吗？

如需了解更多，请访问：nvidia.cn/rtx-pro-4000-sff

1. 使用全长 PCIe 接口。
2. 产品基于已发布的 Khronos 规格，并有望通过日后发布的 Khronos 构象测试流程。
您可以在以下位置找到当前的一致性状态：www.khronos.org/conformance

