

DATASHEET

# Vultr Cloud GPU Powered by NVIDIA A16

Powering low-latency virtual desktop  
infrastructure (VDI), transcoding, and AI inference



VULTR

[VULTR.COM](https://vultr.com)

## DATASHEET

# Vultr Cloud GPU Powered by NVIDIA A16

Powering low-latency virtual desktop infrastructure (VDI), transcoding, and AI inference.

## Introduction

Vultr Cloud GPU, powered by the NVIDIA A16, is a versatile and robust solution designed for low-latency VDI, high-efficiency transcoding, and powerful AI inference. This cutting-edge technology enables businesses to deliver high-performance, secure, scalable computing solutions across various applications. With increased computational capabilities, organizations can enhance productivity, secure data more effectively, and benefit from cost-effective infrastructure.

## Why it's important right now

The NVIDIA A16 GPU is becoming increasingly crucial in today's digital environment, where rapid technological advancements drive the need for robust, efficient computing resources. With its high-performance capabilities and efficient architecture, the NVIDIA A16 on Vultr Cloud GPU provides a highly scalable, flexible environment for businesses to tap into GPU resources on demand while reducing costs and optimizing utilization. By leveraging this industry-leading GPU on Vultr's robust cloud platform, organizations can rapidly streamline deployments, adapt to changing needs, and drive innovation more efficiently.

## Use cases

### AI-powered analytics and AI inference

Leveraging its robust computational power, the NVIDIA A16 GPU is ideal for running AI models and performing real-time data analytics and inference. This use case is vital for sectors such as healthcare for rapid diagnosis, financial services for fraud detection, and retail for personalized customer experiences. The NVIDIA A16's ability to quickly process and analyze large datasets significantly speed up decision-making and enhance the accuracy of predictive models.

### Transcoding and media streaming

The NVIDIA A16 GPU excels in media-rich environments, rapidly transcoding video to different formats, resolutions, and compressions in real time. This capability is crucial for media streaming platforms and content creators who need to deliver high-quality video content across various devices and bandwidths. By accelerating these processes, the NVIDIA A16 ensures efficient management of large video files, reducing buffering and load times, and improving the overall viewer experience.

### High-resolution multi-display setups

Suitable for environments requiring multiple high-resolution displays, the NVIDIA A16 can drive extensive digital signage solutions, large-scale visualizations, and command center setups. This use case is critical for digital marketing, urban planning, and emergency response operations, where real-time, detailed visual data representation is crucial.

### Virtual desktop infrastructure (VDI) and remote desktops

Combined with NVIDIA vPC, the NVIDIA A16 on Vultr Cloud GPU can accelerate graphics-rich virtual and remote desktops for productivity applications. By deploying virtualized desktops, users can access robust GPU resources on demand from anywhere, enabling seamless collaboration and reduced hardware costs while maintaining high-performance computing capabilities.

## Key benefits

1. **Low-latency performance:** NVIDIA vPC and the A16 GPU, optimized for VDI, deliver low-latency performance for a smooth user experience, enhancing response times and application efficiency. The NVIDIA A16's capabilities also support AI inference, enabling real-time analytics and decisions in sectors like healthcare and finance, where speed is critical.
2. **Scalability:** Vultr Cloud GPU allows for easy scalability to meet your organization's growing needs. As your business expands or user demands increase, you can quickly and effortlessly add more GPU resources to accommodate these changes without downtime or disruption to your existing infrastructure.
3. **Cost-effectiveness:** By leveraging the power of the cloud, Vultr Cloud GPU eliminates the need for expensive hardware investments and maintenance costs. Pay-as-you-go pricing ensures that you only pay for the resources you actually use, making it a cost-effective solution for organizations of all sizes.
4. **Enhanced data security:** With Vultr Cloud GPU, all data is securely stored in the cloud, reducing the risk of data loss or theft. Optional automatic backups and redundancy measures ensure that your data remains safe and accessible at all times, even in the event of hardware failure or other unforeseen issues. In addition, Vultr is a leader in delivering cloud solutions to customers, aligning with frameworks that allow customers to attain their [compliance objectives](#).
5. **Easy management and deployment:** The Vultr cloud management platform offers a user-friendly interface for effortlessly managing and deploying your virtual desktop infrastructure and AI inference operations. With just a few clicks, you can provision new virtual desktops, set up inference engines, monitor performance, and manage resources from a centralized location.
6. **Global availability:** With cloud data centers located worldwide, Vultr Cloud GPU, powered by NVIDIA, ensures that your virtual desktop infrastructure is always available and accessible to users, regardless of their location. This guarantees a consistent user experience while minimizing latency and improving overall performance.
7. **Broad application support:** The NVIDIA A16 GPU supports Windows through RDP and Linux through VNC, providing seamless cross-platform compatibility. This allows users to fully utilize this advanced technology in any preferred environment. Combining its high-performance capabilities with extensive application support, the A16 GPU enhances productivity and efficiency across diverse operating systems.

## Specifications

NVIDIA A16 GPU	
GPU Memory	4x 16GB GDDR6 with error-correcting code (ECC)
GPU Memory Bandwidth	4x 200 GB/s
Interconnect	PCI Express Gen 4 x16
vGPU Software Support	NVIDIA Virtual PC (vPC), NVIDIA Virtual Applications (vApps), NVIDIA RTX Virtual Workstation (vWS), and NVIDIA AI Enterprise
NVENC   NVDEC	4x   8x (includes AV1 decode)
Secure and Measured Boot with Hardware Root of Trust	Yes (optional)
NEBS Ready	Level 3

## Conclusion

Vultr Cloud GPU powered by the NVIDIA A16 GPU is an excellent choice for organizations deploying a low-latency virtual desktop infrastructure and AI inference applications. Its high-performance capabilities, cost-effective pricing, and enhanced data security features offer businesses a powerful solution for managing virtual desktop environments.

Learn more about [Vultr Cloud GPU powered by NVIDIA A16](#), and contact us at [vultr.com](#) to get started.