



## NVIDIA A16

### NIEBYWAŁA JAKOŚĆ UŻYTKOWANIA I GĘSTOŚĆ DLA VDI BOGATEGO W GRAFIKĘ

NVIDIA A16, zbudowany na architekturze NVIDIA Ampere, w połączeniu z oprogramowaniem NVIDIA wirtualnego GPU (vGPU), podnosi poprzeczkę w zakresie jakości użytkowania dla bogatej w grafikę infrastruktury wirtualnych pulpitów (VDI). W miarę jak coraz więcej organizacji przechodzi na zdalną pracę jako strategię długoterminową, A16 z oprogramowaniem NVIDIA Virtual PC (vPC) lub Virtual Applications (vApps) umożliwia pracownikom wiedzy w każdej branży maksymalizację wydajności z osiąganymi niemal identycznymi z natywnym komputerem PC. NVIDIA A16 zapewnia do 2 razy większą gęstość użytkowników w porównaniu do poprzedniej generacji M10, co zmniejsza ilość potrzebnych zasobów sprzętowych i obniża całkowity koszt posiadania (TCO).

W połączeniu z oprogramowaniem NVIDIA RTX Virtual Workstation (vWS), A16 umożliwia przystępne cenowo wstępne wirtualne stacje robocze idealne do obsługi obciążeń takich jak projektowanie wspomagane komputerowo (CAD). A16 wyróżnia się unikalnym projektem płyty z czterema GPU, co umożliwia przydzielanie mieszanych rozmiarów profili użytkowników, dzięki czemu dział IT może obsługiwać lekkie obciążenia wirtualnych komputerów PC, a także użytkowników o większych wymaganiach dotyczących pamięci i grafiki. Obsługiwane jest również mieszanie typów użytkowników na jednej płycie, co umożliwia przydzielanie wirtualnych komputerów, wirtualnych stacji roboczych, a nawet wirtualizowanej mocy obliczeniowej na jednej płycie.

## WYBITNA WYDAJNOŚĆ STRUMIENIOWANIA MEDIÓW

NVIDIA A16 oferuje największą liczbę enkoderów i dekodek wideo, z czterema enkoderami sprzętowymi (NVENC) i ośmioma jednostkami dekoderskimi (NVDEC) na jednej płycie A16. To zapewnia najlepszą wydajność kodowania, dekodowania i transkodowania, co przekłada się na maksymalną liczbę strumieni wideo na płycie A16 w atrakcyjnej cenie w porównaniu do alternatywnych ofert.

### NAJWAŻNIEJSZE CECHY

Architektura GPU	Architektura NVIDIA Ampere
Pamięć GPU	4x 16 GB GDDR6
Przepustowość pamięci	4x 200 GB/s
Kod korygujący błędy (ECC)	TAK
Rdzenie CUDA oparte na architekturze NVIDIA Ampere	4x 1280
Rdzenie Tensor trzeciej generacji	4x 40
Rdzenie RT drugiej generacji	4x 10
FP32   TF32   TF32 <sup>1</sup> (TFLOPS)	4x 4.5   4x 9   4x 18
FP16   FP16 <sup>1</sup> (TFLOPS)	4x 17.9   4x 35.9
INT8   INT8 <sup>1</sup> (TOPS)	4x 35.9   4x 71.8
Interfejs systemu	PCIe Gen4 (x16)
Maksymalne zużycie energii	250W
Rozwiązanie termiczne	Pasywne
Kształt obudowy	Pełna wysokość, pełna długość (FHFL) Podwójne gniazdo
Złącze zasilania	8-pin CPU
Silniki kodowania/dekodowania	4 NVENC/8 NVDEC (zawiera dekodowanie AV1)
Bezpieczny i mierzony rozruch ze sprzętowym źródłem zaufania dla procesora graficznego	Tak (opcjonalnie)
Obsługa oprogramowania vGPU	NVIDIA Virtual PC (vPC), NVIDIA Virtual Applications (vApps), NVIDIA RTX Virtual Workstation (vWS), NVIDIA AI Enterprise, NVIDIA Virtual Compute Server (vCS)
API graficzne	DirectX 12.0 <sup>2</sup> , Shader Model 5.1 <sup>2</sup> , OpenGL 4.6 <sup>3</sup> , Vulkan 1.1 <sup>3</sup>
Compute APIs	CUDA, DirectCompute, OpenCL™, OpenACC®
Wsparcie MIG	Nie

# Specyfikacja

## SPECYFIKACJE PRODUKTU

<b>Całkowite zużycie energii</b>	250 W domyślnie 250 W maksymalnie 235 W MaxQ (maksymalna wydajność na W) 195 W minimalnie
<b>Rozwiązanie termiczne</b>	Pasywne
<b>Mechaniczny format obudowy</b>	Pełny profil, pełna długość (FHFL) 10,5", dwusłotowy
<b>Taktowanie GPU</b>	Bazowe: 1312 MHz Boost: 1755 MHz
<b>VBIOS</b>	Rozmiar pamięci EEPROM: 8 Mbit UEFI: Obsługiwane
<b>Interfejs PCI Express</b>	Fizyczne 16 linii PCIe Obsługiwane PCIe Gen4 x16 lub Gen3 x16 Obsługiwane odwracanie linii i polaryzacji
<b>Zero Power</b>	Nie obsługiwane
<b>Waga</b>	Płyta: 1088 g (bez wspornika i przedłużaczy) Wspornik z wkrętami: 20 g Długi przedłużacz offsetowy: 64 g Prosty przedłużacz: 39 g

## SPECYFIKACJE PAMIĘCI

<b>Taktowanie pamięci</b>	6.25 GHz
<b>Typ pamięci</b>	GDDR6
<b>Rozmiar pamięci</b>	64 GB (16 GB na GPU)
<b>Szerokość magistrali pamięci</b>	128 bits
<b>Maksymalna przepustowość pamięci</b>	Do 4x 200 GB/s

## SPECYFIKACJE OPROGRAMOWANIA

<b>Obsługa SR-IOV</b>	Obsługiwane: 16 VF (funkcji wirtualnych)
<b>Adres BAR (fizyczna funkcja)</b>	BAR0: 16 MiB <sup>1</sup> BAR1: 16 GiB <sup>1</sup> BAR3: 32 MiB <sup>1</sup>
<b>Adres BAR (funkcja wirtualna)</b>	BAR0: 4 MiB (256 KiB na VF) <sup>1</sup> BAR1: 32 GiB, 64-bit (2 GiB na VF) <sup>1</sup> BAR3: 512 MiB, 64-bit (32 MiB na VF) <sup>1</sup>
<b>Przerwania sygnalizowane poprzez wiadomość (MSI)</b>	MSI-X: Obsługiwane MSI: Nieobsługiwane
<b>Przekazywanie ARI</b>	Obsługiwane
<b>Wsparcie sterownika</b>	R470 lub nowszy
<b>Uruchamianie zabezpieczeń</b>	Obsługiwane

<b>Firmware CEC</b>	v4.01 lub nowszy (dla kart z obsługą CEC)
<b>Wsparcie dla NVIDIA® CUDA®</b>	CUDA 11.1 lub nowsze
<b>Wsparcie oprogramowania Virtual GPU</b>	Obsługuje vGPU 13.0 lub nowszy Obsługuje NVIDIA vPC, vApps, NVIDIA RTX™ vWS, vCS
<b>NVIDIA® NGC-Ready™ test suite</b>	Certyfikacja NGC-Next 2.x lub nowsza
<b>Kod klasy PCI</b>	0x03 – Kontroler wyświetlacza
<b>Kod podklasy PCI</b>	0x02 – Kontroler 3D
<b>Wsparcie ECC</b>	Włączone (domyślnie); można wyłączyć za pomocą oprogramowania
<b>SMBus (adres 8-bitowy)</b>	GPU1: 0x9E (zapis), 0x9F (odczyt) GPU2: 0x9C (zapis), 0x9D (odczyt) GPU3: 0x9A (zapis), 0x9B (odczyt) GPU4: 0x98 (zapis), 0x99 (odczyt)
<b>Zarezerwowane adresy I2C<sup>2</sup></b>	CEC1: 0xAA (zapis) CEC2: 0xAC (zapis)
<b>Bezpośredni dostęp do SMBus</b>	Obsługiwany
<b>Interfejs SMBPBI (SMBus Post-Box Interface)</b>	Obsługiwany

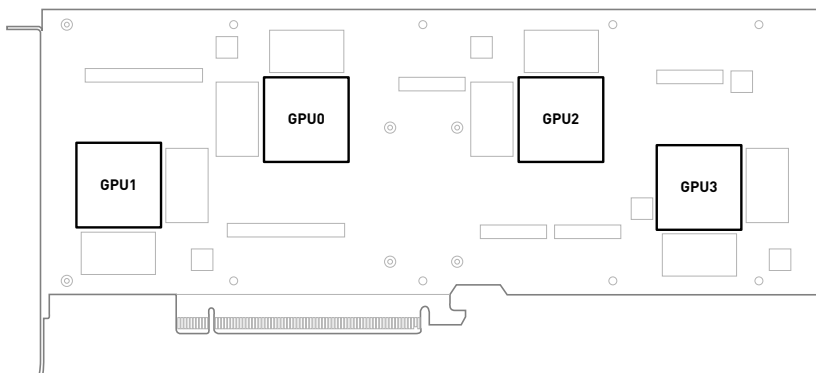
Uwaga:  
Notacja KiB, MiB i GiB podkreśla „potęgę dwójki” wartości. Zatem,  
• 256 KiB = 256 x 1024  
• 16 MiB = 16 x 1024<sup>2</sup>  
• 64 GiB = 64 x 1024<sup>3</sup>

## ŚRODOWISKOWE I NIEZAWODNOŚCIOWE SPECYFIKACJE

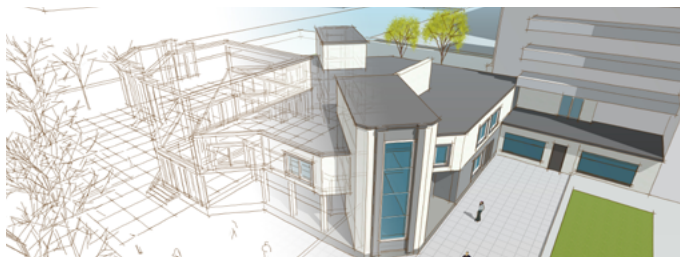
<b>Temperatura pracy otoczenia</b>	0 °C do 50 °C
<b>Temperatura pracy otoczenia (krótkoterminowa)<sup>1</sup></b>	-5 °C do 55 °C
<b>Temperatura przechowywania</b>	-40 °C do 75 °C
<b>Wilgotność robocza (krótkoterminowa)<sup>1</sup></b>	5% do 93% wilgotności względnej
<b>Wilgotność robocza</b>	5% do 85% wilgotności względnej
<b>Wilgotność przechowywania</b>	5% do 95% wilgotności względnej
<b>Średni czas między awariami (MTBF)</b>	Nieuregulowane środowisko: <sup>2</sup> 2 502 369 godzin w temperaturze 35 °C Kontrolowane środowisko: <sup>3</sup> 3 270 359 godzin w temperaturze 35 °C

Notatki:  
Specyfikacje w tej tabeli dotyczą wysokości do 6000 stóp.  
<sup>1</sup> Okres nie dłuższy niż 96 godzin ciągłych, nie więcej niż 15 dni w roku.  
<sup>2</sup> Pewne obciążenie środowiskowe z ograniczoną konserwacją (GF35).  
<sup>3</sup> Brak obciążenia środowiskowego z optymalną eksploatacją i konserwacją (GB35).

## Płyta z czterema GPU



## Obciążenia związane z grafiką i wizualizacją



Procesor graficzny NVIDIA A10 to potężny akcelerator zaprojektowany specjalnie do obciążeń graficznych i wizualizacyjnych. Wykorzystując architekturę Ampere, łączy w sobie możliwości obliczeniowe o wysokiej wydajności z zaawansowanymi funkcjami graficznymi, dzięki czemu idealnie nadaje się do zadań takich jak renderowanie 3D, projektowanie wspomagane komputerowo (CAD) i aplikacje rzeczywistości wirtualnej (VR). A10 obsługuje technologię RTX firmy NVIDIA, która umożliwia śledzenie promieni w czasie rzeczywistym i grafikę wzmocnioną sztuczną inteligencją, zapewniając oszałamiająco realistyczny obraz i przyspieszone czasy renderowania. Dzięki znacznej przepustowości pamięci i rdzeniom CUDA, A10 zapewnia płynne i wydajne przetwarzanie złożonych zbiorów danych i skomplikowanych wizualizacji. Dodatkowo jego kompatybilność z pakietem profesjonalnych narzędzi i sterowników oprogramowania NVIDIA zapewnia bezproblemową integrację z istniejącymi przepływami pracy, co czyni go doskonałym wyborem dla profesjonalistów w takich dziedzinach, jak architektura, media i rozrywka oraz wizualizacja naukowa.

## Analiza wideo i transkodowanie



The NVIDIA A16 GPU is a powerful solution tailored for demanding video analytics and transcoding applications, leveraging its advanced Ampere architecture to deliver exceptional performance and efficiency. Equipped with a high number of CUDA cores and Tensor Cores, the A16 accelerates video processing tasks, such as real-time object detection, facial recognition, and activity tracking, crucial for applications in surveillance, smart cities, and content moderation. Its specialized hardware encoders and decoders ensure efficient handling of multiple high-resolution video streams simultaneously, minimizing latency and optimizing bandwidth utilization. For video transcoding, the A16 offers rapid and high-quality conversion between different formats and resolutions, maintaining visual fidelity across various devices and platforms. The GPU's integration with NVIDIA's comprehensive software stack, including the Video Codec SDK and DeepStream SDK, enhances workflow efficiency and simplifies deployment in video analytics and transcoding environments. These features make the NVIDIA A16 an indispensable tool for enterprises seeking to enhance efficiency, scalability, and performance in video processing applications.

## Virtual Desktop Infrastructure (VDI)



Procesor graficzny NVIDIA A16 został specjalnie zaprojektowany w celu usprawnienia wdrożeń infrastruktury wirtualnych pulpitów (VDI), zapewniając niezrównaną wydajność i efektywność w zdalnych środowiskach pracy. Wykorzystując potężną architekturę Ampere, A16 oferuje dużą liczbę rdzeni CUDA i znaczną przepustowość pamięci, dzięki czemu może obsługiwać wiele wirtualnych pulpitów o wysokiej rozdzielczości jednocześnie, zapewniając płynne i responsywne doświadczenia użytkownika. Zaawansowany koder NVENC zapewnia wydajną kompresję wideo, redukując opóźnienia i wymagania dotyczące przepustowości, zachowując jednocześnie doskonałą jakość obrazu. A16 wykorzystuje także najnowocześniejszą technologię wirtualizacji firmy NVIDIA, umożliwiającą podzielenie go na wiele wirtualnych procesorów graficznych, optymalizując w ten sposób wykorzystanie zasobów i zapewniając dostosowaną wydajność dla szerokiego zakresu aplikacji i obciążeń. To sprawia, że A16 jest idealnym rozwiązaniem dla przedsiębiorstw, które chcą wspierać rozproszoną siłę roboczą za pomocą bezpiecznych, wydajnych wirtualnych pulpitów, które poradzą sobie z wymagającymi zadaniami, od codziennych aplikacji biurowych po oprogramowanie do projektowania wymagające intensywnego przetwarzania grafiki. Integracja z ekosystemem oprogramowania NVIDIA, w tym NVIDIA GRID i VMware vSphere, zapewnia bezproblemowe wdrażanie i zarządzanie w ramach istniejącej infrastruktury IT, czyniąc NVIDIA A16 kamieniem węgielnym nowoczesnych rozwiązań VDI.

# Zmodernizuj swoje VDI i zwiększ wydajność multimediiów strumieniowych



## ZAPROJEKTOWANY DO PRZYSPIESZONEJ VDI

Optymalizowany pod kątem gęstości użytkowników i w połączeniu z oprogramowaniem NVIDIA vPC, umożliwia dostęp do bogatych w grafikę wirtualnych komputerów PC z dowolnego miejsca.



## PRZYSTĘPNE WIRTUALNE STACJE ROBOCZE

Duża pamięć ramki na użytkownika dla podstawowych wirtualnych stacji roboczych z oprogramowaniem NVIDIA RTX vWS, obsługujących obciążenia takie jak projektowanie wspomaganie komputerowo (CAD).



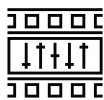
## ELASTYCZNE WSPARCIE DLA RÓŻNORODNYCH TYPOW UŻYTKOWNIKÓW

Unikalny projekt płyty z czterema GPU umożliwia przydzielanie mieszanych rozmiarów profili użytkowników i typów użytkowników, takich jak wirtualne komputery PC i wirtualne stacje robocze, na jednej płycie.



## WYBITNE DOŚWIADCZENIE UŻYTKOWNIKA

Zapewnia zwiększoną liczbę klatek na sekundę i niższe opóźnienie dla użytkowników końcowych w porównaniu do VDI opartych wyłącznie na CPU, co skutkuje bardziej responsywnymi aplikacjami i doświadczeniem użytkownika, które jest nieodróżnialne od natywnego komputera PC lub stacji roboczej natywnego komputera PC lub stacji roboczej.



## PONAD 2-KROTNA PRZEPUSTOWOŚĆ ENKODERA

Ponad dwukrotna przepustowość enkodera w porównaniu do poprzedniej generacji M10, co zapewnia wysokowydajne transkodowanie oraz wydajność wielu użytkowników potrzebną do obsługi wideo w wielu strumieniach i multimediiów.



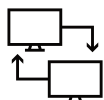
## ARCHITEKTURA NVIDIA AMPERE

Rdzenie CUDA oparte na architekturze NVIDIA Ampere, rdzenie RT drugiej generacji oraz rdzenie Tensor trzeciej generacji oferują elastyczność w hostowaniu wirtualnych stacji roboczych zasilanych oprogramowaniem NVIDIA RTX vWS lub wykorzystaniu niewykorzystanych zasobów VDI do uruchamiania obciążeń obliczeniowych z oprogramowaniem NVIDIA AI Enterprise.



## NAJWYŻSZA JAKOŚĆ WIDEO

Wsparcie dla najnowszych kodeków, w tym kodowania/dekodowania H.265, VP9 oraz dekodowania AV1, zapewniające najwyższą jakość doświadczeń wideo.



## PODWÓJNA GĘSTOŚĆ UŻYTKOWNIKÓW

Zaprojektowane z myślą o VDI bogatym w grafikę, z obsługą do 64 jednoczesnych użytkowników na płycie, w formacie dwusłotowym.



## WYSOKA ROZDZIELCZOŚĆ WYŚWIETLACZA

Obsługuje wiele monitorów o wysokiej rozdzielczości, co umożliwia maksymalną wydajność i fotorealistyczną jakość w środowisku VDI.



## PCI EXPRESS GEN 4

Wsparcie dla prędkości transferu danych PCI Express Gen 4 z pamięci CPU dla zadań wymagających dużej ilości danych.

## Właściwości

- > Zaprojektowane specjalnie dla VDI bogatego w grafikę z NVIDIA vPC
- > Zapewnia najniższy koszt na użytkownika wirtualnej stacji roboczej z NVIDIA RTX vWS4
- > Wsparcie dla wszystkich edycji oprogramowania NVIDIA vGPU: NVIDIA vPC, NVIDIA vApps, NVIDIA RTX vWS, NVIDIA AI Enterprise i NVIDIA vCS
- > PCI Express Gen 4
- > Wszechstronne wsparcie dla najnowszych kodeków: kodowanie/dekodowanie H.265, VP9 oraz dekodowanie AV<sup>1</sup>

Aby dowiedzieć się więcej o NVIDIA A16, odwiedź stronę [www.nvidia.com/a16](http://www.nvidia.com/a16)

<sup>1</sup> Włączona rzadkość strukturalna

<sup>2</sup> GPU wspiera interfejs API DX 12.0, Hardware Feature Level 12 + 1

<sup>3</sup> Produkt oparty jest na opublikowanej specyfikacji Khronosa i oczekuje się, że przejdzie proces testowania zgodności Khronosa, gdy stanie się dostępny. Aktualny status zgodności można znaleźć na stronie [www.khronos.org/conformance](http://www.khronos.org/conformance)

<sup>4</sup> Porównanie NVIDIA A16 z wirtualnymi stacjami roboczymi T4, RTX 6000, RTX 8000 i A40.