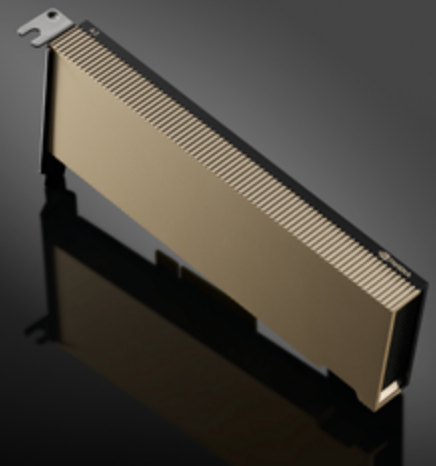




GPU Z RDZENIEM NVIDIA A2 TENSOR

Procesor graficzny klasy podstawowej, który wdraża sztuczną inteligencję NVIDIA na dowolny serwer.



Uniwersalne wnioskowanie na poziomie początkującym

Procesor graficzny **NVIDIA A2 Tensor Core** zapewnia wnioskowanie na poziomie podstawowym przy niskim poborze mocy, niewielkich rozmiarach i wysokiej wydajności dla **NVIDIA AI** na krawędzi. Wyposażony w niskoprofilową kartę PCIe Gen4 i niską konfigurowalną moc cieplną (TDP) wynoszącą 40–60 W (W), **A2** zapewnia adaptacyjne przyspieszenie wnioskowania dla każdego serwera.

Wszechstronność, niewielkie rozmiary i mała moc **A2** przewyższają wymagania dotyczące wdrożeń brzegowych na dużą skalę, umożliwiając natychmiastową modernizację istniejących podstawowych serwerów procesorowych w celu obsługi wnioskowania. Serwery przyspieszane procesorami graficznymi A2 zapewniają wyższą wydajność wnioskowania w porównaniu z procesorami CPU i wydajniejsze wdrożenia inteligentnej analizy wideo (IVA) niż poprzednie generacje procesorów graficznych – a wszystko to w cenie podstawowej.

Systemy z certyfikatem **NVIDIA™** wyposażone w procesory graficzne A2 i sztuczną inteligencję NVIDIA, w tym serwer wnioskowania **NVIDIA Triton™**, zapewniają przetomową wydajność wnioskowania na krawędzi, w centrum danych i w chmurze.

Zapewniają, że aplikacje obsługujące sztuczną inteligencję będą wdrażane z mniejszą liczbą serwerów i mniejszą mocą, co skutkuje łatwiejszymi wdrożeniami, szybszym wglądem i znacznie niższymi kosztami.

Do 20x wydajniejsze osiągi wnioskowania

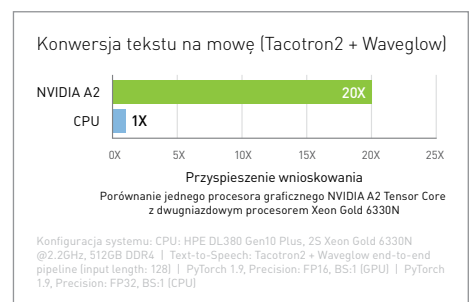
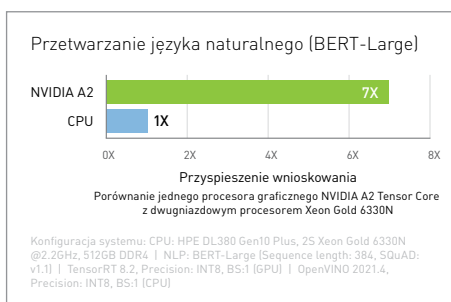
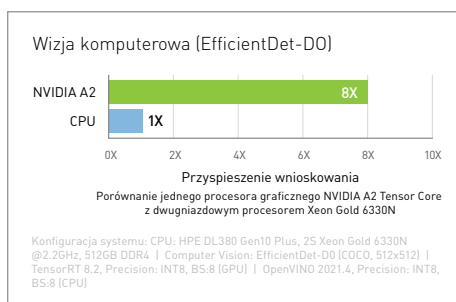
Wnioskowanie AI jest wdrażane, aby uczynić życie konsumentów wygodniejszym dzięki doświadczeniom w czasie rzeczywistym i umożliwić im uzyskanie wglądu w biliony czujników i kamer końcowych. W porównaniu z serwerami wyposażonymi wyłącznie w procesor, serwery zbudowane z procesorem graficznym **NVIDIA A2 Tensor Core** oferują do 20 razy większą wydajność wnioskowania, umożliwiając natychmiastową modernizację dowolnego serwera pod kątem obsługi nowoczesnej sztucznej inteligencji.

SPECYFIKACJE SYSTEMU

Peak FP32	4.5 TF
TF32 Tensor Core	9 TF 18 TF¹
BFLOAT16 Tensor Core	18 TF 36 TF¹
Peak FP16 Tensor Core	18 TF 36 TF¹
Peak INT8 Tensor Core	36 TOPS 72 TOPS¹
Peak INT4 Tensor Core	72 TOPS 144 TOPS¹
RT Cores	10
Silniki multimedialne	1 koder wideo 2 dekodery wideo (zawiera dekodowanie AV1)
Pamięć GPU	16GB GDDR6
Przepustowość pamięci GPU	200GB/s
Złącze	PCIe Gen4 x8
Współczynnik kształtu	1-gniazdo, niskoprofilowe PCIe
Maksymalna moc obliczeniowa cieplna (TDP)	40-60W (Konfigurowalna)
Obsługa oprogramowania vGPU ²	NVIDIA Virtual PC (vPC), NVIDIA Virtual Applications (vApps), NVIDIA RTX Virtual Workstation (vWS), NVIDIA AI Enterprise, NVIDIA Virtual Compute Server (vCS)

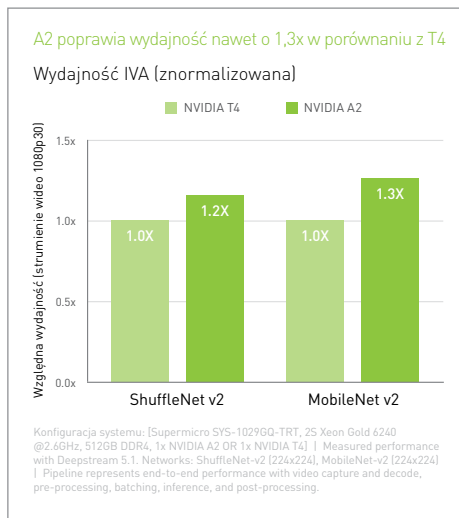
¹ Rzadko

² Obsługiwane w przyszłej wersji vGPU

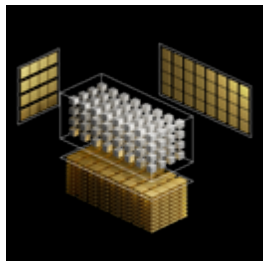


Wyższa wydajność IVA dla inteligencji krawędziowej

Serwery wyposażone w A2 oferują do 1,3 razy większą wydajność w inteligentnych zastosowaniach brzegowych, w tym w inteligentnych miastach, produkcji i handlu detalicznym. Procesory graficzne NVIDIA A2 obsługujące obciążenia IVA zapewniają bardziej wydajne wdrożenia, oferując nawet 1,6 razy lepszą wydajność w stosunku do ceny i o dziesięć procent lepszą efektywność energetyczną w porównaniu z poprzednimi generacjami procesorów graficznych.

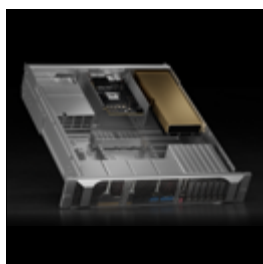


NVIDIA A2 zapewnia przetłomową technologię NVIDIA Ampere Innowacje w architekturze



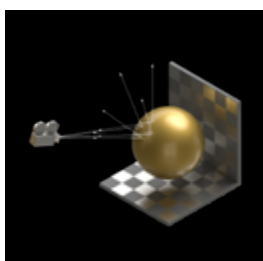
RDZENIE TENSOROWE TRZECIEJ GENERACJI

Rdzenie Tensorowe trzeciej generacji w A2 obsługują matematykę na liczbach całkowitych aż do INT4 i matematykę zmiennoprzecinkową do FP32, zapewniając wysoką wydajność uczenia i wnioskowania AI. Architektura NVIDIA Ampere obsługuje także funkcje TF32 i automatycznej mieszanej precyzji (AMP) firmy NVIDIA.



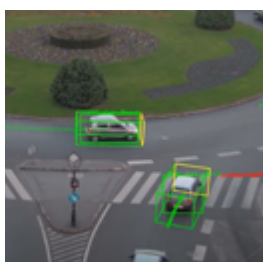
RDZEŃ BEZPIECZEŃSTWA ZAUFANIA

Zapewnienie bezpieczeństwa wdrożeń brzegowych i punktów końcowych ma kluczowe znaczenie dla operacji biznesowych przedsiębiorstwa. A2 oferuje bezpieczny rozruch dzięki uwierzytelnianiu zaufanego kodu i wzmocnionym zabezpieczeniom przed wycofywaniem w celu ochrony przed atakami złośliwego oprogramowania.



RDZENIE RT DRUGIEJ GENERACJI

A2 zawiera dedykowane rdzenie RT do śledzenia promieni, które umożliwiają przetłomowe technologie z przetłomową szybkością. Z nawet dwukrotnie większą przepustowością w porównaniu z poprzednią generacją i możliwością jednoczesnego śledzenia promieni z możliwością cieniowania lub usuwania szumów.

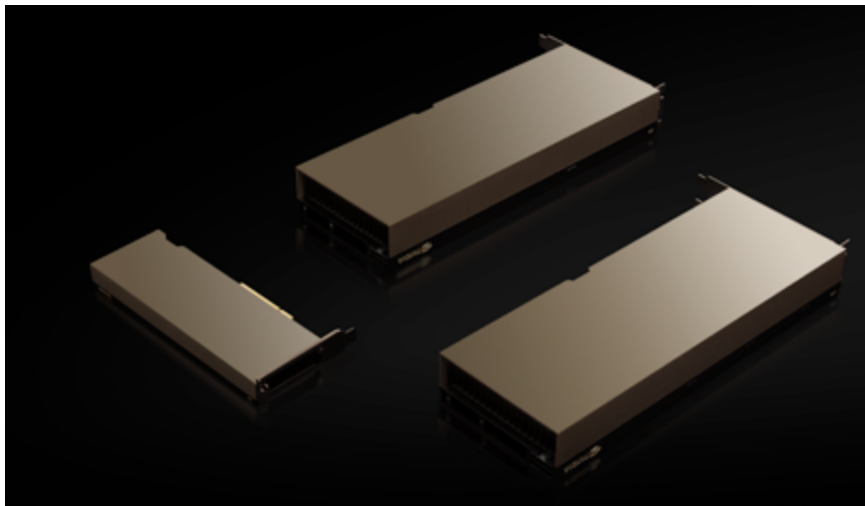


WYDAJNOŚĆ TRANSKODOWANIA SPRZĘTU

Wykładniczy rozwój aplikacji wideo wymaga skalowalnej wydajności w czasie rzeczywistym, co wymaga najnowszych możliwości sprzętowego kodowania i dekodowania. Procesory graficzne A2 wykorzystują dedykowany sprzęt do pełnego przyspieszenia dekodowania i kodowania wideo dla najpopularniejszych kodeków, w tym dekodowania H.265, H.264, VP9 i AV1.

Kompletne portfolio wnioskowania

NVIDIA oferuje pełną gamę systemów z certyfikatem NVIDIA, wyposażonych w procesory graficzne Ampere Tensor Core jako silnik wnioskowania napędzający sztuczną inteligencję NVIDIA. Procesory graficzne A2 Tensor Core dodają wnioskowanie na poziomie podstawowym w niskoprofilowej obudowie do portfolio AI NVIDIA, które obejmuje już procesory graficzne A100 i A30 Tensor Core. A100 charakteryzuje się najwyższą wydajnością wnioskowania w każdej skali, a A30 zapewnia optymalną wydajność wnioskowania dla głównych serwerów. Procesory graficzne NVIDIA A2, NVIDIA A30 i NVIDIA A100 Tensor Core zapewniają wiodącą wydajność wnioskowania na krawędziach, w centrach danych i w chmurze.



Zoptymalizowane oprogramowanie i usługi dla przedsiębiorstw

Przedsiębiorstwo oparte na sztucznej inteligencji NVIDIA

NVIDIA AI Enterprise, kompleksowy pakiet oprogramowania do sztucznej inteligencji i analizy danych, natywnie działający w chmurze, posiada certyfikat do działania na platformie A2 w infrastrukturze wirtualnej opartej na hypervisorze z VMware vSphere. Umożliwia to zarządzanie i skalowanie obciążeń AI i wnioskowania w środowisku chmury hybrydowej.

SYSTEMY Z CERTYFIKATEM NVIDIA

Systemy z certyfikatem NVIDIA z NVIDIA A2 łączą przyspieszenie obliczeniowe i szybką, bezpieczną sieć z systemami wiodących partnerów NVIDIA w konfiguracjach sprawdzonych pod kątem optymalnej wydajności, niezawodności i skali. Dzięki systemom z certyfikatem NVIDIA przedsiębiorstwa mogą śmiało wybierać rozwiązania sprzętowe zoptymalizowane pod kątem wydajności, aby obsługiwać przyspieszone obciążenia obliczeniowe — od komputerów stacjonarnych, przez centra danych, aż po urządzenia brzegowe.



[Dowiedz się więcej](#)

Aby dowiedzieć się więcej o procesorze graficznym NVIDIA A2 Tensor Core, odwiedź stronę [nvidia.com/a2](https://www.nvidia.com/a2).

© 2021 NVIDIA Corporation. All rights reserved. NVIDIA, the NVIDIA logo, Triton, NVIDIA-Certified Systems, and NGC are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. All other trademarks and copyrights are the property of their respective owners. NOV21



FORMAT



NVIDIA