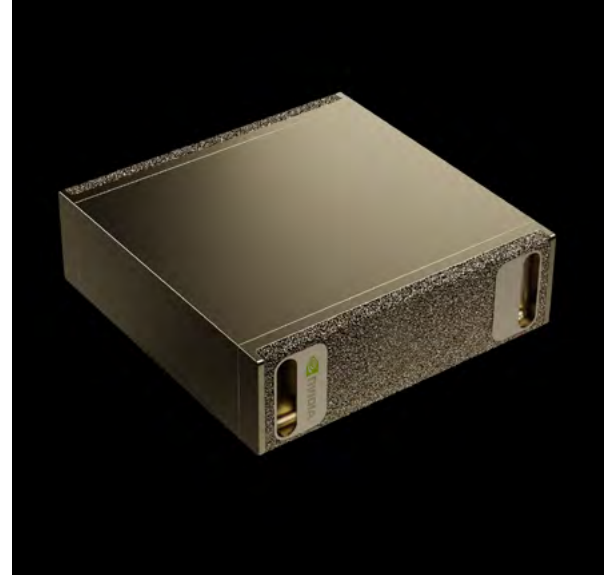




NVIDIA DGX Spark

Osobisty komputer AI DGX, zaprojektowany do tworzenia i uruchamiania AI.



Wymagania dotyczące obliczeń AI na komputerze stacjonarnym

oraz większa wielkość i złożoność modeli generatywnej AI sprawia, że prace rozwojowe na lokalnych systemach stają się coraz bardziej wyzwaniem. Tworzenie prototypów, dostrajanie i inferencja dużych modeli lokalnie wymaga dużej ilości pamięci oraz znacznej mocy obliczeniowej. W miarę rozbudowy przedsiębiorstw, dostawców oprogramowania, agencji rządowych, startupów i naukowców zaangażowanych w AI, zapotrzebowanie na zasoby obliczeniowe AI stale rośnie.

Modele z 200 miliardami parametrów na Twoim biurku

NVIDIA DGX™ Spark jest częścią nowej klasy komputerów zaprojektowanych od podstaw do budowy i uruchamiania AI. Napędzany przez NVIDIA GB10 Grace Blackwell Superchip i oparty na architekturze NVIDIA Grace Blackwell, NVIDIA DGX Spark oferuje do 1000 TOPS wydajności AI, co pozwala obsłużyć duże obciążenia AI. Z 128 GB zunifikowanej pamięci systemowej, programiści mogą eksperymentować, dostrajając lub wykonywać inferencje modeli do 200 miliardów parametrów. Dodatkowo, sieć NVIDIA ConnectX™ może połączyć dwa superkomputery NVIDIA DGX Spark, umożliwiając wykonywanie inferencji na modelach do 405 miliardów parametrów.

Aby zapewnić programistom komfortowe doświadczenie, NVIDIA DGX Spark odzwierciedla tę samą architekturę oprogramowania, która napędza przemysłowe fabryki AI. Wykorzystując system operacyjny NVIDIA DGX z Ubuntu Linux i wstępnie skonfigurowane najnowsze oprogramowanie NVIDIA AI, wraz z dostępem do programów deweloperskich NVIDIA NIM™ i NVIDIA Blueprints, programiści mogą od razu ruszyć do pracy, korzystając z popularnych narzędzi takich jak Pytorch, Jupyter i Ollama do prototypowania, dostrajania i inferencji na NVIDIA DGX Spark, a następnie bezproblemowo wdrażać je w centrach danych lub w chmurze.

Dostarczając ogromną wydajność i możliwości w kompaktowej formie, NVIDIA DGX Spark umożliwia programistom, naukowcom, data scientistom i studentom nieustanne przesuwanie granic generatywnej AI.

Kluczowe cechy

- > Zbudowany na NVIDIA GB10 Grace Blackwell Superchip
- > GPU NVIDIA Blackwell z technologią Tensor Core 5. generacji
- > CPU NVIDIA Grace z 20-rdzeniową architekturą ARM o wysokiej wydajności
- > Do 1000 TOPS wydajności AI z użyciem FP4
- > 128 GB spójnej, zunifikowanej pamięci systemowej
- > Obsługa modeli do 200 miliardów parametrów
- > Sieć NVIDIA ConnectX™ do połączenia dwóch systemów w celu obsługi modeli do 405 miliardów parametrów
- > Do 4 TB pamięci NVMe
- > Kompaktowa forma stacji roboczej na biurko

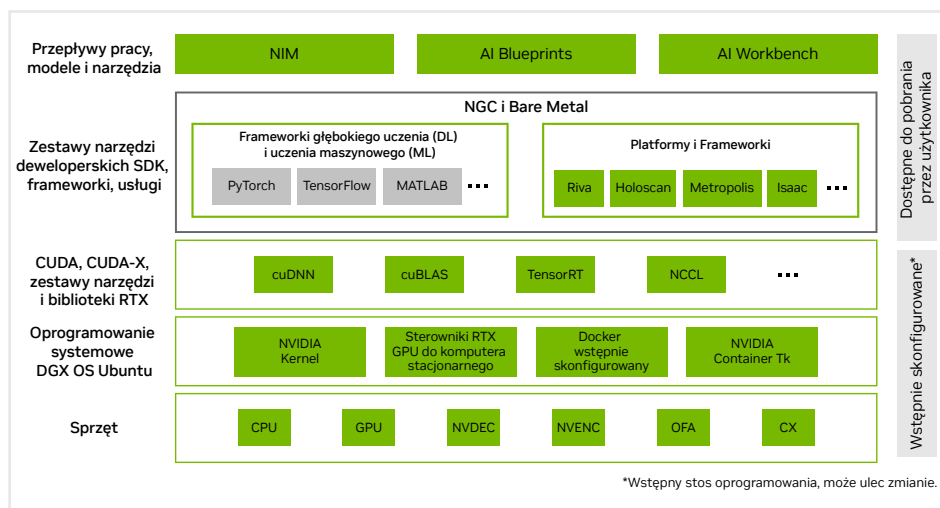
Zbudowany na architekturze NVIDIA Grace Blackwell

Serce NVIDIA DGX Spark stanowi nowy NVIDIA GB10 Grace Blackwell Superchip, oparty na architekturze NVIDIA Grace Blackwell, zoptymalizowanej pod kątem formatu desktop. GB10 wyposażony jest w potężny GPU NVIDIA Blackwell z piątą generacją Tensor Cores i wsparciem FP4, zapewniający do 1000 TOPS mocy obliczeniowej AI. GB10 zawiera wysokowydajny procesor ARM Grace z 20 rdzeniami, który przyspiesza wstępne przetwarzanie danych i orkiestrację, przyspieszając dostrajanie modeli i inferencję w czasie rzeczywistym. Superchip GB10 korzysta z NVLink™-C2C, zapewniając spójny model pamięci CPU+GPU z 5 razy większą przepustowością niż PCIe Gen 5.

Pracuj z dużymi modelami AI o wielu parametrach

Dzięki 128 GB zunifikowanej pamięci systemowej i wsparciu dla formatu danych FP4, NVIDIA DGX Spark może obsługiwać modele AI do 200 miliardów parametrów, umożliwiając deweloperom prototypowanie, dostrajanie i inferencję dużych modeli na swoim komputerze stacjonarnym. Z wbudowaną technologią sieci NVIDIA ConnectX, dwa systemy NVIDIA DGX Spark można połączyć, aby pracować nad jeszcze większymi modelami, takimi jak Llama 3.1 405B.

Twórz lokalnie, wdrażaj wszędzie na dużą skalę



Stos oprogramowania NVIDIA DGX Spark

NVIDIA DGX Spark zapewnia organizacjom i deweloperom potężną, ekonomiczną platformę do eksperymentowania z modelami prototypowymi, uwalniając cenne zasoby obliczeniowe w ich środowiskach klastrowych, które lepiej nadają się do treningu i wdrażania modeli produkcyjnych. Wykorzystując architekturę oprogramowania NVIDIA AI platformy, użytkownicy NVIDIA DGX Spark mogą bezproblemowo przenosić swoje modele z komputera stacjonarnego do chmury DGX lub dowolnej innej przyspieszonej chmury lub infrastruktury centrum danych przy praktycznie zerowych zmianach w kodzie, co czyni prototypowanie, dopracowywanie i iterację bardziej prostą niż kiedykolwiek wcześniej.

Specyfikacje techniczne*

Architektura	NVIDIA Grace Blackwell
GPU	NVIDIA Blackwell Architecture
CPU	20 rdzeni ARM, 10 Cortex-X925 + 10 Cortex-A725 ARM
Jądra CUDA	Generacja NVIDIA Blackwell
Jądra Tensor	5. generacja
Jądra RT	4. generacja
Wydajność Tensor¹	1000 TOPS AI
Pamięć systemowa	128 GB LPDDR5x, zunifikowana pamięć systemowa
Interfejs pamięci	256-bit
Przepustowość pamięci	273 GB/s
Pamięć masowa	1 lub 4 TB NVMe M.2 z własnym szyfrowaniem
USB	4x USB Type-C
Ethernet	1x RJ-45, 10 GbE
NIC	ConnectX-7 Smart NIC
Wi-Fi	WiFi 7
Bluetooth	BT 5.3 z LE
Wyjście audio	HDMI wielokanałowe
Zużycie energii	TBD
Złącza wyświetlaczy	1x HDMI 2.1a
NVENC	NVDEC
System operacyjny	NVIDIA DGX™ OS
Wymiary systemu	150 mm długości 150 mm szerokości 50,5 mm wysokości
Waga systemu	1,2 kg

* Wstępne specyfikacje, mogą ulec zmianie

1. Teoretyczna wydajność FP4 TOPS, z wykorzystaniem funkcji sparsity..

Gotowy, aby zacząć?

Aby dowiedzieć się więcej o NVIDIA DGX Spark, odwiedź:

nvidia.com/dgx-spark/

© 2025 NVIDIA Corporation i jej spółki zależne. Wszelkie prawa zastrzeżone. NVIDIA, logo NVIDIA, ConnectX, DGX, NVLink oraz NIM są znakami towarowymi i/lub zarejestrowanymi znakami towarowymi NVIDIA Corporation oraz jej spółek zależnych w USA i innych krajach. Inne nazwy firm i produktów mogą być znakami towarowymi odpowiednich właścicieli. FORMAT PL. MAY25

