



NVIDIA DGX Spark

DGX personal AI computer, designed to build and run AI.



ローカルシステムにおける AI計算需要

生成AIモデルのサイズと複雑さが増大するにつれて、ローカルシステムでの開発作業が困難になっています。大規模モデルのプロトタイプリング、チューニング、推論をローカルで行うには、大量のメモリと高い計算性能が必要です。企業、ソフトウェアプロバイダ、政府機関、スタートアップ、研究機関がAIへの取り組みを強化するにつれて、AI計算リソースの必要性は増大し続けています。

机上で200B パラメーターモデル

NVIDIA DGX Sparkは、AI の構築と実行のためにゼロから設計されたパーソナルAI コンピュータです。NVIDIA GB10 Grace Blackwell Superchipを搭載し、NVIDIA Grace Blackwellアーキテクチャに基づいたNVIDIA DGX Sparkは、最大1000TOPSのAI性能を提供し、大規模な AIワークロードを強化します。128GBのユニファイドシステムメモリにより、開発者は最大 200Bパラメータのモデルを実験、ファインチューニング、または推論することができます。さらに、NVIDIA ConnectXネットワーキングにより、2台のNVIDIA DGX Sparkスーパーコンピュータを接続して、最大405Bパラメータのモデルで推論を行うことができます。

NVIDIA Grace Blackwell を搭載

NVIDIA DGX Spark の中核をなすのは、デスクトップフォームファクタに最適化されたNVIDIA Grace Blackwell アーキテクチャに基づく新しいNVIDIA GB10 Grace Blackwell Superchipです。GB10は、第5世代TensorコアとFP4をサポートする強力なNVIDIA Blackwell GPUを搭載しており、最大1000 TOPSのAI計算性能を提供します。GB10には、データの前処理とオーケストレーションを大幅に高速化し、モデルのチューニングとリアルタイム推論を加速する高性能なGrace 20コアArm CPUが含まれています。GB10 SuperchipはNVLink-C2Cを使用しており、PCIe Gen 5の5倍の帯域幅を持つCPU+GPUコヒーレントメモリモデルを実現します。

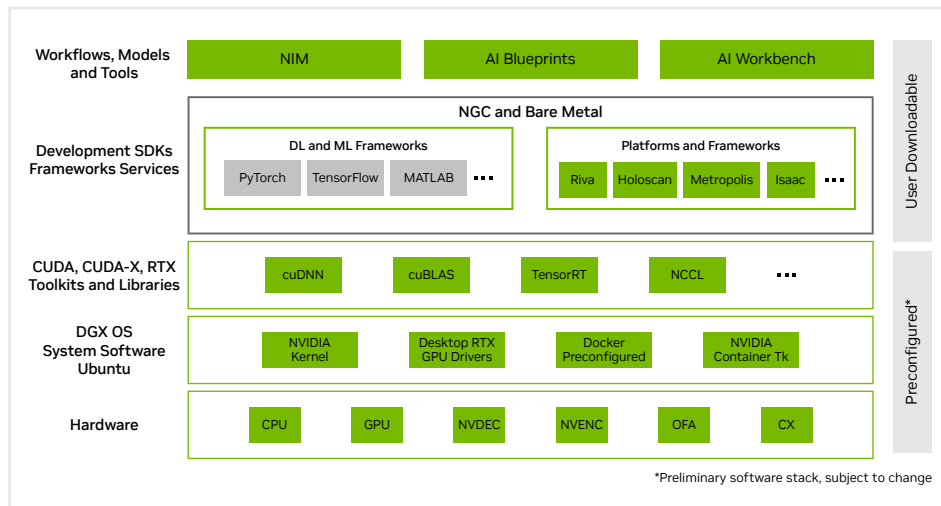
主な特長

- NVIDIA GB10 Grace Blackwell Superchipを搭載
- 第5世代Tensorコアテクノロジーを搭載したNVIDIA Blackwell GPU
- 20コア高性能Armアーキテクチャを搭載したNVIDIA Grace CPU
- FP4使用時、最大1000 TOPSのAI性能
- 128 GBのコヒーレントな統合システムメモリ
- 最大2000億パラメータモデルをサポート
- NVIDIA ConnectXネットワーキングにより、2つのシステムを接続し、最大4050億パラメータモデルに対応
- 最大4 TBのNVMeストレージ
- コンパクトなデスクトップフォームファクタ

大規模パラメータAIモデルでの作業

128 GBの統合システムメモリとFP4データ形式のサポートにより、NVIDIA DGX Sparkは最大2000億パラメータのAIモデルをサポートし、AI開発者はデスクトップ上で大規模モデルのプロトタイプ作成、ファインチューニング、および推論を行うことができます。

ローカルで開発し、どこでも大規模に展開



NVIDIA DGX Spark software stack

NVIDIA DGX Sparkは、組織や開発者に対し、プロトタイプモデルのための強力かつ経済的な実験環境を提供します。これにより、クラスター環境における貴重な計算リソースを、生産モデルのトレーニングや展開により適した用途に解放します。

NVIDIA AIプラットフォームのソフトウェアアーキテクチャを活用することで、NVIDIA DGX Sparkのユーザーは、事実上コードを変更することなく、デスクトップからDGX Cloud、または任意のアクセラレーションクラウドやデータセンターインフラストラクチャにモデルをシームレスに移行できます。これにより、プロトタイプ作成、ファインチューニング、イテレーションがこれまで以上に容易になります。

Technical Specifications*

Architecture	NVIDIA Grace Blackwell
GPU	NVIDIA Blackwell Architecture
CPU	20 core Arm, 10 Cortex-X925 + 10 Cortex-A725 Arm
CUDA Cores	NVIDIA Blackwell Generation
Tensor Cores	5th Generation
RT Cores	4th Generation
Tensor Performance ¹	1000 AI TOPS
System Memory	128 GB LPDDR5x, unified system memory
Memory Interface	256-bit
Memory Bandwidth	273 GB/s
Storage	1 or 4 TB NVME.M2 with self-encryption
USB	4x USB TypeC
Ethernet	1x RJ-45 connector 10 GbE
NIC	ConnectX-7 Smart NIC
Wi-Fi	WiFi 7
Bluetooth	BT 5.3 w/LE
Audio-output	HDMI multichannel audio output
Power Consumption	TBD
Display Connectors	1x HDMI 2.1a
NVENC NVDEC	1x 1x
OS	NVIDIA DGX™ OS
System Dimensions	150 mm L x 150 mm W x 50.5 mm H
System Weight	1.2 kg

* preliminary specifications, subject to change

1. Theoretical FP4 TOPS using the sparsity feature.

Ready to Get Started?

To learn more about NVIDIA DGX Spark, visit
nvidia.com/dgx-spark/

© 2025 NVIDIA Corporation and affiliates. All rights reserved. NVIDIA, the NVIDIA logo, ConnectX, DGX, NVLink, and NIM are trademarks and/or registered trademarks of NVIDIA Corporation and affiliates in the U.S. and other countries. Other company and product names may be trademarks of the respective owners with which they are associated. 3841551. APR25

