

ハードウェアの調達からGPUによる開発環境構築までお客様のフェーズに合わせてフルスタックで支援！
仮想化および分散化技術とインフラ構築の実績をもとに、お客様のAIプロジェクトを成功に導きます。

GPU スペック比較表 2026年版 Midokura GPU Service

AMD Instinct / Radeon PRO ・ NVIDIA Hopper / Ampere / Blackwell / Ada スペック一覧

ベンダー / シリーズ	AMD Instinct / Radeon PRO						NVIDIA (Blackwell / Hopper / Ampere / Ada)															
GPU モデル名	AMD MI210	AMD PRO R9700	AMD MI300X	AMD MI325X	AMD MI350X	AMD MI355X	NVIDIA GB200 NVL72	NVIDIA B300 HGX	NVIDIA B200 HGX	NVIDIA H200 HGX	NVIDIA H200 NVL	NVIDIA H100 HGX	NVIDIA H100 NVL	NVIDIA A100	NVIDIA PRO 6000 Max-Q	NVIDIA PRO 6000	NVIDIA RTX A6000 Ada	NVIDIA RTX A6000				
画像																						
GPU タイプ	サーバー	ワークステーション	サーバー				サーバー												ワークステーション	ワークステーション		
アーキテクチャ	CDNA2	RDNA4	CDNA3	CDNA3	CDNA4	CDNA4	Blackwell	Blackwell	Blackwell	Hopper	Hopper	Hopper	Hopper	Ampere	Blackwell	Blackwell	Ada	Ampere				
販売状況	*1	販売中	*1	販売中	販売中	販売中	販売中	販売中	販売中	販売中	販売中	*1	*1	販売中	販売中	販売中	*1	*1				
■ 導入・運用要件																						
フォームファクター	PCIe Add-in Card		OAM Module				ラックスケール (NVL72ラック)	SXM (HGX baseboard)			PCIe Add-in Card	SXM (HGX baseboard)	PCIe Add-in Card	SXM / PCIe Add-in Card	PCIe Add-in Card (標準高さ、2スロット)	PCIe Add-in Card (FHFL、2スロット)	PCIe Add-in Card					
冷却方式	パッシブ	アクティブ (ブロー)	パッシブ OAM *2	パッシブ OAM *3	パッシブ OAM *2	パッシブ OAM *5	液冷必須 (ラック全体)	パッシブ *4			パッシブ (空冷ラック向け)	パッシブ *4	パッシブ (Product Brief確認済)	パッシブ *4	アクティブ (ブロー、300W)	パッシブ (600W、サーバー側空冷)	アクティブ (ブロー)					
対応プラットフォーム	標準PCIe 4.0 サーバー・WS	ワークステーション (PCIe 5.0対応)	AMD UBB 2.0 対応サーバー	AMD UBB 2.0 対応サーバー	AMD UBB 2.0 対応サーバー (空冷)	AMD UBB 2.0 対応サーバー (液冷)	NVLink Switch 対応専用ラックのみ	GPUモデル対応サーバー			標準PCIe 5.0 サーバー	GPUモデル対応サーバー	標準PCIe 5.0 サーバー	HGX A100 / 標準PCIe サーバー	ワークステーション (最大4GPU構成)	標準PCIe 5.0 サーバー (マルチGPU)	プロ向けワークステーション					
ソフトウェアエコシステム	ROCm	ROCm / OpenCL	ROCm				CUDA 12+						CUDA 11.8+			CUDA 11+	CUDA 12+	CUDA 12+	CUDA 11.8+	CUDA 11+		
■ 基本仕様																						
最大消費電力 (W)	300	300	750	1000	1000	1400	1200	1100	1000	700	600	700	400	400	300	600	300	300				
GPU メモリ (GB)	64	32	192	256	288	288	186	288	180	141	141	80	80	80	96	96	48	48				
メモリアイプ	HBM2e	GDDR6	HBM3	HBM3e	HBM3e	HBM3e	HBM3e	HBM3e	HBM3e	HBM3e	HBM3e	HBM3e	HBM3	HBM2e	GDDR7	GDDR7	GDDR6	GDDR6				
メモリ帯域幅 (TB/s)	1.6	0.64	5.3	6	8	8	8	8	7.7	4.8	4.8	3.35	3.9	2	1.792	1.792	0.96	0.768				
NVLink / Interconnect	—	—	0.9 TB/s	0.9 TB/s	1.075 TB/s	1.075 TB/s	1.8 TB/s	1.8 TB/s	1.8 TB/s	0.9 TB/s	0.9 TB/s	0.9 TB/s	0.9 TB/s	0.6 TB/s	0.6 TB/s	—	—	—	0.112 TB/s			
■ 演算性能 (TFLOPS)																						
FP64 倍精度	22.6	0.76	82	82	72	79	40	1.25	37	34	30	34	30	9.7	1.721	1.968	1.423	0.605				
FP32 単精度	22.6	47.8	163	163	144	157	80	75	75	67	60	67	60	19.5	110	126	91	39				
FP16 半精度	181	191	1307	1307	2310	2517	2500	2200	2200	989	835	989	835	312	439	504	364	155				
BF16	181	191	1307	1307	2310	2517	2500	2200	2200	989	835	989	835	312	439	504	364	155				
INT8	181	383	2615	2615	4614	5033	5000	150	4500	1979	1671	1979	1671	624	878	1008	728	310				
FP8	—	383	2615	2615	4614	5033	5000	4500	4500	1979	1671	1979	1671	—	878	1008	728	—				
FP4	—	766	—	—	9227	10066	10000	13100	9000	—	—	—	—	—	1755	2015	—	—				
■ スパース性能																						
TF32	45.3	191	1307	1307	2310	2517	2500	2200	2200	989	835	989	835	156	439	504	364	155				
FP16	181	383	2615	2615	4614	5033	5000	4500	4500	1979	1671	1979	1671	312	878	1008	728	310				
BF16	181	383	2615	2615	4614	5033	5000	4500	4500	1979	1671	1979	1671	312	878	1008	728	310				
INT8	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—				
FP8	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—				
FP6	—	766	—	—	18455	20133	10000	9000	9000	—	—	—	—	—	1755	2015	—	—				
FP4	—	1531	—	—	18455	20133	20000	18000	18000	—	—	—	—	—	3511	4030	—	—				
■ 電力効率																						
FP16 TFLOPS/W	0.6	0.64	1.74	1.31	2.31	1.8	2.08	2	2.2	1.41	1.39	1.41	2.09	0.78	1.46	0.84	1.21	0.52				
FP32 TFLOPS/kW	75.3	159.3	217.3	163	144	112.1	66.7	68.2	75	95.7	100	95.7	150	48.8	366.7	210	303.3	130				
備考							ラック単位での販売												後継:H100	低消費電力モデル	フルパワーモデル	後継:RTX A6000 Ada

*1 販売状況はお問合せください。
*2 サーバー側：空冷
*3 サーバー側：空冷/水冷
*4 サーバー側：空冷/液冷
*5 サーバー側：液冷必須

※スペックは変更になる場合があります。
記載されている会社名、製品名は各社の商標または登録商標です。

GPUの選定から調達・構築・運用まで、ワンストップでお任せください。
詳細はこちら → midokura-gpu.jp

