

# NVIDIA vGPU対応 NVIDIA GPUラインナップ

ワークロードの用途からGPUを選択、サーバの型番や搭載可能枚数をご確認ください



	A100	A30	V100 V100S	A40	RTX A5000	RTX A6000	QUADRO RTX		T4	A16	M10
							8000	6000			
1ボード搭載GPU数 (GPU世代)	1 (Ampere)	1 (Ampere)	1 (Volta)	1 (Ampere)	1 (Ampere)	1 (Ampere)	1 (Turing)	1 (Turing)	1 (Turing)	4 (Ampere)	4 (Maxwell)
CUDAコア数	6,912	3,584	5,120	10,752	8,192	10,752	4,608	4,608	2,560	10,240 (2,560/GPU)	2,560 (640/GPU)
Tensorコア数	432	224	640	336	256	336	576	576	320	160 (40/GPU)	-
RTコア数	-	-	-	84	65	84	72	72	40	40 (10/GPU)	-
メモリサイズ	40GB/80GB HBM2	24GB HBM2	16GB/32GB HBM2	48GB GDDR6	24GB GDDR6	48GB GDDR6	48GB GDDR6	24GB GDDR6	16GB GDDR6	64GB GDDR6 (16GB/GPU)	32GB GDDR5 (8GB/GPU)
vGPU Profiles (GB)	4,5,8,10, 20,40,80	4,6,12,24	1,2,4,8,16,32	1,2,3,4,6,8, 12,16,24,48 ※Max32VM	1,2,3,4,6, 8,12,24	1,2,3,4,6,8, 12,16,24,48 ※Max32VM	1, 2,3,4,6,8, 12,16,24,48 ※Max32VM	1,2,3,4,6, 8,12,24	1,2,4,8,16	1,2,4,8,16	0.5,1,2,4,8
NVLink Support	Yes	Yes	Yes (SXM2 only)	Yes	Yes	Yes	Yes	Yes	No	No	No
フォームファクター 物理寸法	PCIe4.0 Dual Slot / SXM4	PCIe4.0 Dual Slot	PCIe 3.0 Dual Slot / SXM2	PCIe4.0 Dual Slot	PCIe4.0 Dual Slot	PCIe4.0 Dual Slot	PCIe3.0 Dual Slot	PCIe3.0 Dual Slot	PCIe3.0 Single Slot	PCIe4.0 Dual Slot	PCIe3.0 Dual Slot
消費電力	250W/300W /400W	165W	250W/300W	300W	230W	300W	295W	295W	70W	250W	225W
冷却	passive	passive	passive	passive	active	active	active/passive	active/passive	passive	passive	Passive
推奨 利用 シ ーン	オフィス	-	-	-	-	-	-	-	Office	Office	Office
	グラフィックス	-	-	-	CAD/RT/RENDERING		CAD/RT/RENDERING		CAD/RT	CAD/RT	-
	コンピューティング	INT4/INT8/FP16/BF16 FP32/TF32/FP64 CAE/AI/DL/HPC		FP32/FP64 CAE/AI/DL/HPC	INT4/INT8/FP16/BF16 FP32/TF32 CAE/AI/DL/HPC		FP32 CAE/DL		INT4/INT8 FP16/FP32 AI/DL	INT4/INT8 FP16/FP32 AI/DL	-

仮想化のための NVIDIA GPU : <https://www.nvidia.com/ja-jp/data-center/graphics-cards-for-virtualization/>

NVIDIA 認定サーバ : <https://www.nvidia.com/object/vgpu-certified-servers.html> ※最新サーバ構成、サポート状況は各サーバメーカー様にご確認ください。

