AWS Trainium2 vs NVIDIA Groq 3 LPU

Side by side datacenter GPU specifications from the Flopper.io database.

OVERVIEW

	AWS Trainium2	NVIDIA Groq 3 LPU
FP32 TFLOPS	**181.0**	–
Memory	**96 GB**	–
Memory Bandwidth	**2.9 TB/s**	–
TDP	–	–

PERFORMANCE BY PRECISION

Precision	AWS Trainium2 peak TFLOPS (dense)	NVIDIA Groq 3 LPU peak TFLOPS (dense)
FP64	–	–
FP32	181.0	–
TF32	667.0	–
BF16	667.0	–
FP16	667.0	–
FP8	1299.0	1200.0
FP6	–	–
FP4	–	–
INT8	–	–

SPECIFICATIONS

	AWS Trainium2	NVIDIA Groq 3 LPU
Architecture	NeuronCore-v3	Groq LPU
Process Node	–	–
Launch Year	2024	2026
Form Factor	–	–
Memory Type	–	On-die SRAM
Interface Width	–	–
Interconnect	NeuronLink-v3	LPU C2C

	AWS Trainium2	NVIDIA Groq 3 LPU
TDP	–	–
FP32 Efficiency	–	–