



NVIDIA Groq 3 LPU

Groq LPU

2026

OVERVIEW

—	—	—	—
FP32 TFLOPS	VRAM On-die SRAM	Memory Bandwidth	TDP

PERFORMANCE METRICS

Precision	Peak TFLOPS	Bits	Type
FP8	1200.0	8	Standard

MEMORY SPECIFICATIONS

Capacity	—
Type	On-die SRAM
Bandwidth	—
Interface Width	—

POWER & EFFICIENCY

TDP	—
Est. Max Power	—
FP32 Efficiency	—
FP16 Efficiency	—

HARDWARE DETAILS

Vendor: NVIDIA	Architecture: Groq LPU	Process Node: —
Form Factor: —	Launch Year: 2026	

INTERCONNECT

GPU-to-GPU: LPU C2C	Bandwidth: 2.5 TB/s
---------------------	---------------------