

Google TPU v3 32GB vs Google TPU v5e 16GB

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

OVERVIEW

	Google TPU v3 32GB	Google TPU v5e 16GB
FP32 TFLOPS	—	—
VRAM	32 GB	16 GB
Memory Bandwidth	900 GB/s	800 GB/s
TDP	—	—

PERFORMANCE BY PRECISION

Precision	Google TPU v3 32GB peak TFLOPS (dense)	Google TPU v5e 16GB peak TFLOPS (dense)
INT8	—	393.0
BF16	123.0	197.0

SPECIFICATIONS

	Google TPU v3 32GB	Google TPU v5e 16GB
Architecture	TPU	TPU
Process Node	—	—
Launch Year	2018	2023
Form Factor	—	—
Memory Type	HBM2	HBM
Interface Width	—	—
Interconnect	—	—
TDP	—	—
FP32 Efficiency	—	—