



Google TPU v3 32GB vs Google TPU v4 32GB

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

OVERVIEW

	Google TPU v3 32GB	Google TPU v4 32GB
FP32 TFLOPS	—	—
VRAM	32 GB	32 GB
Memory Bandwidth	900 GB/s	1.2 TB/s
TDP	—	—

PERFORMANCE BY PRECISION

Precision	Google TPU v3 32GB peak TFLOPS (dense)	Google TPU v4 32GB peak TFLOPS (dense)
INT8	—	275.0
BF16	123.0	275.0

SPECIFICATIONS

	Google TPU v3 32GB	Google TPU v4 32GB
Architecture	TPU	TPU
Process Node	—	—
Launch Year	2018	2020
Form Factor	—	—
Memory Type	HBM2	HBM2
Interface Width	—	—
Interconnect	—	—
TDP	—	—
FP32 Efficiency	—	—