Qualcomm AI250 vs Qualcomm Cloud AI 100 Ultra

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

OVERVIEW

	Qualcomm AI250	Qualcomm Cloud AI 100 Ultra
FP32 TFLOPS	—	—
Memory	768 GB	128 GB
Memory Bandwidth	—	548 GB/s
TDP	—	150 W

PERFORMANCE BY PRECISION

Precision	Qualcomm AI250 peak TFLOPS (dense)	Qualcomm Cloud AI 100 Ultra peak TFLOPS (dense)
FP64	—	—
FP32	—	—
TF32	—	—
BF16	—	—
FP16	—	—
FP8	—	—
FP6	—	—
FP4	—	—
INT8	—	870.0

SPECIFICATIONS

	Qualcomm AI250	Qualcomm Cloud AI 100 Ultra
Architecture	Dragonfly	Cloud AI 100
Process Node	—	7nm
Launch Year	2027	2023
Form Factor	—	PCIe
Memory Type	HBC Gen 1	LPDDR4X
Interface Width	—	—

	Qualcomm AI250	Qualcomm Cloud AI 100 Ultra
Interconnect	PCIe 6.0 scale-up / Ethernet with RoCE scale-out	—
TDP	—	150 W
FP32 Efficiency	—	—