

Rebellions RebelServer

8x Rebel100

REBEL

rack

OVERVIEW

16.4 FP8 PFLOPS Dense	1.2 TB Aggregate Memory	38.4 TB/s Aggregate Bandwidth	7.0 kW Aggregate Power
---------------------------------	-----------------------------------	---	----------------------------------

PERFORMANCE METRICS

Precision	Peak (PFLOPS)	Per GPU (TFLOPS)	TFLOPS/W
FP16	8.2	1024.0	—
FP8	16.4	2048.0	—

All figures are dense. Vendors commonly headline the sparse number, which is twice the dense one.

CONFIGURATION

Accelerator	Rebel100
Count	8
Memory per GPU	144 GB
Bandwidth per GPU	4.8 TB/s

POWER & EFFICIENCY

Aggregate Power	7.0 kW
TDP per GPU	600 W
FP8 Efficiency	2.34 TFLOPS/W
Aggregate Bandwidth	38.4 TB/s

SYSTEM DETAILS

Vendor: Rebellions	Family: Rebel	Architecture: REBEL
Form Factor: rack	Release Year: —	Process Node: 4nm

INTERCONNECT

Backend 4 x 400G QSFP112-DD; 1 x 100G 2-port QSFP56; 1 x 10G/25G 2-port SFP28 (OCP 3.0)

NOTES

A 5U server holding eight Rebel100 cards, driven by two AMD EPYC 9355 processors with 1.5 TB of DDR5 host memory across 24 DIMMs, drawing 4 kW typically and 7 kW at maximum, and weighing 45.3 kg gross. Accelerator memory is 1.152 TB of HBM3E across the eight cards at 38.4 TB/s. Compute is 16.384 PFLOPS of FP8 and 8.192 PFLOPS of FP16, dense — Rebellions labels the Rebel100 figures "Compute (Dense)" on its own homepage. Note that Rebellions' server page states "up to 2 PFLOPS FP8 performance with 8 RebelCards per server"; that 2 PFLOPS is the per-card figure rather than the per-server one, as its own rack and pod pages confirm, since 32 cards give 65.5 PFLOPS and 256 give 524.3. Note also that the 1.5 TB on that page is DDR5 host memory, not accelerator memory.