

AMD Instinct MI300X Platform

8x Instinct MI300X

CDNA 3

baseboard

2023

OVERVIEW

20,919 FP8 TFLOPS Dense	1.5 TB Aggregate Memory	42.4 TB/s Aggregate Bandwidth	6.0 kW Aggregate Power
---	---	---	--

PERFORMANCE METRICS

Precision	Peak (TFLOPS)	Per GPU (TFLOPS)	TFLOPS/W
FP64	654	82	—
FP32	1,307	163	—
TF32	5,230	654	—
FP16	10,459	1,307	—
BF16	10,459	1,307	—
FP8	20,919	2,615	—
INT8	20,919	2,615	—

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

CONFIGURATION

Accelerator	Instinct MI300X 192GB
Count	8
Memory per GPU	192 GB
Bandwidth per GPU	5.3 TB/s

POWER & EFFICIENCY

Aggregate Power	6.0 kW
TDP per GPU	750 W
FP8 Efficiency	3.49 TFLOPS/W
Aggregate Bandwidth	42.4 TB/s

SYSTEM DETAILS

Vendor: AMD	Family: Instinct Platform	Architecture: CDNA 3
Form Factor: baseboard	Release Year: 2023	Process Node: 5nm FinFET

INTERCONNECT

4th Gen AMD Infinity Fabric, eight fully connected OAM modules on an OCP UBB 2.0 baseboard, 7x 128 GB/s links per GPU, 896 GB/s ring-of-8 aggregate

NOTES

AMD's eight-GPU building block for CDNA 3, integrating eight MI300X OAM modules on an industry-standard OCP universal baseboard (UBB 2.0), each linked to the other seven over 128 GB/s Infinity Fabric links for 896 GB/s of aggregate ring bandwidth, plus one PCIe Gen 5 x16 host link per GPU. AMD's platform datasheet gives 1.5 TB of HBM3 (stored as 8 x 192 GB = 1.536 TB) and 5.3 TB/s per GPU, so 42.4 TB/s across the board, with a

750 W maximum board power per GPU (6.0 kW for the eight GPUs, not counting the host). Compute: 2,432 compute units, 155,648 stream processors at up to 2,100 MHz. AMD sells it through solution partners rather than directly. Precision figures are eight times the per-GPU MI300X row, which AMD's one-decimal platform figures round, except FP32 and FP64, where AMD's platform figures are printed more precisely and are stored as printed.