

HPE ProLiant Compute XD685 (Instinct MI355X)

8x Instinct MI355X

CDNA 4

rack

2025

OVERVIEW

40.3 FP8 PFLOPS Dense	2.3 TB Aggregate Memory	64.0 TB/s Aggregate Bandwidth	— Aggregate Power
---------------------------------	-----------------------------------	---	-----------------------------

PERFORMANCE METRICS

Precision	Peak (PFLOPS)	Per GPU (TFLOPS)	TFLOPS/W
FP64	0.6	78.6	—
FP32	1.3	157.3	—
FP16	20.1	2516.6	—
BF16	20.1	2516.6	—
FP8	40.3	5033.2	—
FP4	80.5	10066.3	—
INT8	40.3	5033.2	—
FP6	80.5	10066.3	—

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

CONFIGURATION

Accelerator	Instinct MI355X OAM
Count	8
Memory per GPU	288 GB
Bandwidth per GPU	8.0 TB/s

POWER & EFFICIENCY

Aggregate Power	—
TDP per GPU	1400 W
FP8 Efficiency	—
Aggregate Bandwidth	64.0 TB/s

SYSTEM DETAILS

Vendor: HPE	Family: ProLiant Compute XD	Architecture: CDNA 4
Form Factor: rack	Release Year: 2025	Process Node: 3nm/6nm

INTERCONNECT

4th Gen AMD Infinity Fabric across the eight-OAM OCP baseboard; 4x FHHL PCIe Gen5 x16 slots for fabric, NIC or DPU

NOTES

The eight-way XD685 in its AMD Instinct MI355X configuration, liquid cooled, with accelerator kit S6J64A. This is an all-AMD machine, pairing eight MI355X OAM modules with two fifth-generation

AMD EPYC processors. No chassis part number is recorded: HPE names the accelerator kit but never states which chassis it requires, and the two candidates differ. Compute comes from AMD's eight-GPU Instinct platform, where the sparse figures cover only the four datatypes AMD publishes them for, being FP8, FP16, BF16 and INT8; AMD gives no structured-sparsity figure for MXFP4 or MXFP6. HPE publishes no system power draw for the XD685, only twelve 3,000 W Titanium supplies, so nothing is summed.