

Intel Gaudi 3 HLB-325 Baseboard

8x Gaudi 3

Gaudi 3

baseboard

2024

OVERVIEW

—	1.0 TB	—	7.6 kW
Peak Dense	Aggregate Memory	Aggregate Bandwidth	Aggregate Power

CONFIGURATION

Accelerator	Gaudi 3 128GB
Count	8
Memory per GPU	128 GB
Bandwidth per GPU	3.7 TB/s

POWER & EFFICIENCY

Aggregate Power	7.6 kW
TDP per GPU	900 W
Efficiency	—
Aggregate Bandwidth	—

SYSTEM DETAILS

Vendor: Intel	Family: Gaudi	Architecture: Gaudi 3
Form Factor: baseboard	Release Year: 2024	Process Node: 5nm

INTERCONNECT

Direct all-to-all between the eight accelerators at 4.2 TB/s bidirectional with no switching IC, plus 1.2 TB/s bidirectional scale-out over 6 OSFP connectors as 24x 200 Gbps RDMA (RoCE v2); host attach is 8x PCIe Gen 5.0 x16

NOTES

Intel's eight-accelerator universal baseboard for Gaudi 3, and the unit Intel offers OEMs to build servers around. It carries eight HL-325L OAM mezzanine cards and connects them all-to-all at 4.2 TB/s bidirectional without a switching chip, which is the design's distinguishing choice: scale-up needs no switch silicon, and scale-out runs over standard Ethernet at 1.2 TB/s through six OSFP connectors rather than a proprietary fabric. Intel rates the whole assembly at 7.6 kW including the eight OAMs, draws it from a 54 V main input with 12 V standby, and sizes the board at 417 by 585 mm to drop into 19-inch server designs. Sold either as the bare HLB-325 or, as the HLFB-325L, pre-populated with eight cards and tested as one subsystem. Intel publishes no aggregate throughput figure for the baseboard, so none is shown: per-accelerator compute is on the Gaudi 3 chip entry, sourced to Table 5 of Intel's technical paper. Total memory of 1.024 TB is the eight cards' published 128 GB apiece.