

AceleMax[®] AXG-828IG

8U DP Intel® Xeon® 6 GPU Server with 8x Intel® Gaudi 3 Accelerators



Features

- Increased memory for LLM efficiency: >1TB HBM capacity per universal baseboard and 29.6 TB/s HBM bandwidth. The large HBM capacity and bandwidth allow the Intel[®] Gaudi[®] 3 accelerators to achieve state-of-the-art performance for GenAl training and inference.
- Better performance: Intel[®] Gaudi[®] 3 accelerators deliver 4x BF16 compute, 2x networking bandwidth, and 1.5x memory bandwidth over Gaudi[®] 2.
- Scalability: With systems built on the Intel® Gaudi® 3 accelerators, enterprise customers can address the compute demands of GenAI with flexibility and capacity to scale from 1 node to 1000s based on open, industry-standard ethernet networking.

Applications

- Large-scale AI model training and AI inferencing
- · Generative AI applications such as LLM and RAG

| Specifications | |
|-----------------|--|
| GPU | • 8 Gaudi 3 HL-325L (air-cooled) accelerators on OAM 2.0 baseboard |
| CPU | Dual Intel® Xeon® 6 processors |
| Memory | • 24 DIMMs - up to 6TB memory in 1DPC |
| Drives | Up to 8 hot-swap PCIe 5.0 NVMe |
| Power Supplies | • 8 3000W high efficiency fully redundant (4+4) Titanium Level |
| Networking | 6 on-board OSFP 800GbE ports for scale-out |
| Expansion Slots | • 2 PCle 5.0 x16 (FHHL) + 2 PCle 5.0 x8 (FHH L) |
| Workloads | Al Training and Inference |



AMAX | 1565 Reliance Way, Fremont, CA 94539 | 1 (800) 800-6328 | www.amax.com | info@amax.com

Copyright © 2024 AMAX. All rights reserved. All trademarks are the property of their respective owners. Technical information is subject to change without notice. All company and product names are trademarks or registered trademarks of their respective owners. 090924