Connect Tech

INFERENCE SERVER
POWERED BY NVIDIA® JETSON AGX ORIN™

GPU-Accelerated Machine Learning and Inferencing Servers Bring Flexibility and Cost Efficiency to High-Powered Computing

Developed in partnership with USES Integrated Solutions, this Inference Server is an extremely low wattage, high performance AI workstation powered by the NVIDIA® Jetson AGX Orin™ 64GB module. Running NVIDIA's most powerful deep-learning software libraries, this inference server solves the challenge of deploying edge solutions at scale. The on-board 12x Jetson AGX Orin™ modules are all connected by a Gigabit Ethernet fabric through a specialized Managed Ethernet Switch with 10G uplink capability.

NVIDIA Elite Partner

Includes Out of Band Management (OOBM) via an ARM-based SMARC module to control and monitor each of the Jetson AGX Orin™ modules.

FEATURES

✓ 12x 275 TOPS, 2048-core NVIDIA® Ampere™ GPU and 64 Tensor Cores

✓ 4x 10G SFP+ uplink capability

✓ Operating Temperature: 0°C to +50°C (+32°F to +122°F)

✓ 2U ATX style redundant power supply

✓ Up to 2TB of NVMe storage per module

SPECIFICATIONS

Processing Modules	• 12x NVIDIA® Jetson AGX Orin™ • GPU: 2048-core CUDA GPU with 64 Tensor Cores, 11 TFLOPS (FP16), 275 TOPS (INT8) • CPU: 12-core Cortex ARM v8.2 64-bit CPU, 3MB L2 + 6MB L3 • Memory: 64GB 256-bit LPDDR5 | 204.8 GB/s • Storage: 64 GB eMMC 5.1	**Out-of-band Management Module**	• ARM based OOBM • Enabling Serial console access, power status monitoring, and power control (ON/OFF) to all 12x AGX Orin modules • OOBM accessible via Ethernet or via its own integrated USB-to-Serial console
Processor Module Carriers	• Each module carrier will allow up to 4x NVIDIA Jetson AGX Orin modules to be installed • 3x module carriers can be installed in the system for a total of 12 modules	**Internal Embedded Ethernet Switch**	• Vitesse/Microsemi SparX-5i VSC7558TSN-V/5CC Managed Ethernet Switch Engine (XDG205) • CPU: 1 GHz VCore • Memory: 8Gb DDR4 SDRAM • Storage: 1Gbit Serial NOR Flash • Multi 10G uplinks, with 12x 1G downstream • Complete TSN feature set (Time-Sensitive Networking)
Cooling	• Forced convection via fans on front and rear of system • Fans: Delta, 5x 70 x 70 x 28 mm, 75 CFM / 2,400 in H2O static pressure	**Internal Array Communication**	• 12x Gigabit Ethernet / 1000BASE-T / IEEE 802.3ab channels • All AGX Orin modules can communicate to all other AGX Orin modules
External Uplink Connections	• 4x SFP+ 10G uplink	**Misc / Additional IO**	1x 1GbE OOB management port via RJ-45; 1x USB UART management port; status LEDs
Input Power	100~240 VAC (dual redundant) with 1000W output each	**Internal Storage**	Each AGX Orin module has its own M.2 NVMe interface (optional)
Operating Temperature	0°C to +50°C (+32°F to +122°F)	**Dimensions**	Standard 2U rackmount height (3.5 inch / 88.9mm) 25 inch / 635mm Depth

RoHS