



Housed in a 2U chassis, the Inference Server can be outfitted with up to 24x NVIDIA® Jetson Orin™ NX modules.



Developed in partnership with USES Integrated Solutions, this Inference Server is an extremely low wattage, high performance deep learning inference server powered by the NVIDIA® Jetson Orin™ NX 16GB module. Within the Server, three processor module carriers house up to eight Jetson Orin™ NX modules each, and are all connected via a Gigabit Ethernet fabric through a specialized Managed Ethernet Switch developed by Connect Tech with 10G uplink capability (XDG205).



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

FEATURES

- ✓ 24x 100 TOPS, 1024 GPU CUDA cores with NVIDIA® Ampere™ architecture
- ✓ 4x 10G SFP+ uplink capability
- ✓ Operating Temperature: 0°C to +50°C (+32°F to +122°F)
- ✓ 2U ATX style redundant power supply

SPECIFICATIONS

Processing Modules	<ul style="list-style-type: none">• 24x NVIDIA® Jetson Orin™ NX• GPU: 1024-core CUDA GPU with 32 Tensor Cores, 3.7 TFLOPS (FP16), 100 TOPs (INT8)• CPU: 8-core Cortex Arm v8.2 CPU, 2MB L2 + 4MB L3• Memory: 16 GB 128-bit LPDDR5 102.4GB/s• Storage: Supports External Storage (NVMe) via x4 PCIe	Out-of-band Management Module	<ul style="list-style-type: none">• ARM based OOBM• Enabling Serial console access, power status monitoring, and power control (ON/OFF) to all 24x Orin NX modules• OOBM accessible via Ethernet or via its own integrated USB-to-Serial console
Processor Module Carriers	<ul style="list-style-type: none">• Each module carrier will allow up to 8x Orin NX modules to be installed• 3x module carriers can be installed in the system for a total of 24 modules	Internal Embedded Ethernet Switch	<ul style="list-style-type: none">• 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	<ul style="list-style-type: none">• Passive heatsinks installed on each Orin NX module• 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	<ul style="list-style-type: none">• 24x Gigabit Ethernet / 1000BASE-T / IEEE 802.3ab channels• All Orin NX modules can communicate to all other Orin NX modules
External Uplink Connections	<ul style="list-style-type: none">• 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 Orin NX module has its own M.2 NVMe interface
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