

8x NVIDIA HGX H200 141GB SXM5 Enterprise AI SuperServer Barebone System

Generic Sourcing Specification: 8x NVIDIA H200 SXM5 141GB · Dual Intel Xeon Platinum 8592+ · 4,096 GB (4.0 TB) DDR5 · 61.44 TB Gen5 NVMe

✳️ Open Architecture Specification · Vendor-Neutral Enterprise Standard · True 8x 3,000W Titanium Power Delivery & Full-Speed NVLink Mesh

Official Exclusive Contact
(WeChat ID):

asianeventspk

Delivery Corridors: Hong Kong Free Port (FOB HK) | Shenzhen FTZ Warehouse

Settlement Desk: Dubai DAFZA Free Zone · Global Supply Chain Clearing Desk

1. Commercial Sourcing & Compliance Terms

Standard Procurement Guidelines

1. Authenticity & Quality

100% Factory-Sealed Brand New Original. Zero tolerance for refurbished, engineering samples (ES/QS), or third-party assembled units.

2. Serial Audit & QA

Prior to dispatch, supplier must furnish high-resolution photographic evidence of chassis serial plates, GPU Baseboard GUIDs, and component barcodes.

3. Logistics & Lead Time

Primary handover: FOB Hong Kong Free Port Warehouse or Shenzhen Free Trade Zone (FTZ). Immediate ex-stock allocation preferred; max 10-15 business days.

4. Settlement Currency

Settlement in USD or Offshore RMB (CNH). Payment via International Wire Transfer (T/T), Confirmed Irrevocable L/C, or Commercial Escrow.

2. Detailed Bill of Materials (BOM) — Part 1

Chassis Barebone, GPU Assembly, CPUs & Memory

ITEM	SUBSYSTEM	DETAILED TECHNICAL SPECIFICATION (GENERIC / OPEN STANDARD)	QTY	UNIT QUOTE (USD)
01	Server Barebone Chassis Subsystem	8U / 6U High-Density Dual-Socket Enterprise AI Server Rackmount Chassis Dual 5th Gen Intel Xeon Scalable socket platform. Optimized front-to-back isolated high-static-pressure dual-chamber wind tunnel cooling engineered specifically for 8x 700W SXM5 GPUs. Includes heavy-duty tool-less sliding rails and cable management arm. 8U / 6U Form Factor · High-Static Pressure Airflow	1 Unit	TBD
02	AI Accelerators GPU Complex	NVIDIA HGX H200 8-GPU SXM5 141GB 700W Integrated Baseboard Assembly 8x NVIDIA H200 141GB HBM3e GPUs (700W TDP per module). Aggregate cluster memory: 1,128 GB (1.128 TB). Aggregate memory bandwidth: 38.4 TB/s. Integrated 4x NVSwitch Gen3 chips providing 900 GB/s bidirectional NVLink per GPU. Optimized for vLLM, DeepSeek, and Qwen 2.5 72B FP8 (TP=8). 1,128 GB HBM3e · 900 GB/s NVLink · 4x NVSwitch	1 Set (8 GPUs)	TBD
03	Host Processors Dual Enterprise CPUs	Dual Intel Xeon Platinum 8592+ Processors (5th Gen Emerald Rapids) 64 Cores / 128 Threads per socket (Dual Socket = 128 Cores / 256 Threads total). 1.90 GHz Base / 3.90 GHz Max Turbo Frequency, 320 MB Shared L3 Cache, 350W TDP. Native PCIe Gen 5.0 lanes with high-performance copper heatsinks.	2 CPUs	TBD
04	System Memory Physical DDR5 RAM	128GB DDR5-5600 Registered ECC RDIMM Modules (Populating all 32 Slots) 32x 128GB DDR5-5600 Registered ECC memory modules. Total system physical memory: 4,096 GB (4.0 TB). Symmetrically populated across all 16 memory channels (8 channels/socket x 2 DPC). Tier-1 authentic OEM DRAM modules (Samsung / SK Hynix / Micron). 32 DIMM Slots Populated · 4,096 GB (4.0 TB) DDR5	32 DIMMs	TBD
05	Boot Storage OS Mirror Array	Dual M.2 NVMe 960GB PCIe SSDs (Hardware RAID 1 Boot Controller) 2x M.2 NVMe 960GB enterprise-grade solid-state drives configured in dedicated hardware RAID 1 mirror carrier for resilient, zero-downtime operating system booting.	1 Set (2 Drives)	TBD

2. Detailed Bill of Materials (BOM) — Part 2

Fast NVMe Storage, High-Speed Networking & Power Subsystem

ITEM	SUBSYSTEM	DETAILED TECHNICAL SPECIFICATION (GENERIC / OPEN STANDARD)	QTY	UNIT QUOTE (USD)
06	Fast NVMe Storage Primary Data Pool	15.36TB PCIe Gen 5.0 x4 U.3 / U.2 Enterprise Data Center NVMe SSDs x 4 Units 4x 15.36TB PCIe 5.0 x4 U.3 enterprise NVMe SSDs (Samsung PM1743 class or equivalent). Sequential Read up to 14,000 MB/s, Random Read up to 2.5M IOPS. Aggregate cluster fast data pool: 61.44 TB raw Gen5 NVMe for instant LLM weights and KV cache checkpoints. 61.44 TB Raw Gen5 NVMe Fast Data Pool · 14,000 MB/s Read	4 Drives	TBD
07	100G Data Fabric Cluster Ingestion	Dual-Port 100GbE QSFP56 PCIe Full-Height Network Adapters x 2 Units 2x Dual-Port 100GbE QSFP56 PCIe full-height adapters (Mellanox ConnectX-6 Dx class or equivalent). Includes 2x 100GbE QSFP28 SR4 MPO multimode optical transceivers and high-grade optical patch cords.	2 NICs	TBD
08	25G Host Fabric Cluster Control Plane	Dual-Port 10/25GbE SFP28 PCIe Full-Height Network Adapters x 2 Units 2x Dual-Port 10/25GbE SFP28 PCIe full-height adapters (Broadcom 57414 class or equivalent) for host management, telemetry, and API ingestion. Includes 2x 25GbE SFP28 SR optical transceivers.	2 NICs	TBD
09	10G Mgmt Network Out-of-Band & Telemetry	Quad-Port 10GbE Base-T RJ45 Network Adapter x 1 Unit 1x Quad-Port 10GbE Base-T RJ45 network adapter (Intel X710-T4L class, PCIe or OCP 3.0 form factor) for BMC/IPMI interconnect, administrative telemetry, and subnet isolation.	1 NIC	TBD
10	Power Subsystem Hot-Swap Titanium PSUs	8x 3,000W / 3,300W 80-Plus Titanium Hot-Plug Power Supplies (or 6x 2,800W System) 8x 3,000W/3,300W 80-Plus Titanium level (96% electrical efficiency) hot-plug redundant power supplies configured in true 4+4 or 6+2 redundant mode. Provides 24,000W reserve to eliminate GPU Power Braking under 700W peak FP8 loads. Includes 16A/250V C20/C21 high-load heavy-duty jumper cords. 8x 3,000W Titanium Level · 24,000W Reserve · Zero Throttling Risk	8 PSUs	TBD
11	Enterprise AI SW Official Production License	NVIDIA AI Enterprise (NVAIE) 3-Year Subscription Licenses (8-GPU Allocation) 8x NVIDIA AI Enterprise (NVAIE) 3-Year subscription licenses per GPU, upgraded to 24x7 Enterprise Production Support. Full official access to NVIDIA NIM microservices, NeMo Framework, Triton Inference Server, and vLLM acceleration stacks.	8 Licenses	TBD

3. Vendor Quotation Submission Guidelines

Submission Checklist & Requirements:

- Provide itemized and aggregate net commercial pricing in USD (FOB HK / Shenzhen FTZ);
- Specify physical warehouse location (Hong Kong Free Port / Shenzhen FTZ / Factory Direct);
- Confirm Earliest Ship Date (ESD) and quantity of units available in current allocation;
- State accepted commercial settlement mechanisms (T/T Wire / Commercial L/C / Escrow).

Exclusive Sourcing Communication Channel:

To ensure confidentiality and direct real-time communication, **no public telephone numbers are listed**. Verified suppliers must reach our Technology & Sourcing Desk directly via WeChat:

👉 Official WeChat ID: **asianeventspk**

Direct Sourcing & Inquiry Message Template

Ready to copy and send via WeChat / Chat

[ASIAN INTERNATIONAL INQUIRY · 8x H200 SXM5 ENTERPRISE AI SERVER]
RFQ Reference: AI-RFQ-202610-H200-EN-GENERIC
Delivery Terms: FOB Hong Kong Free Port / Shenzhen FTZ Warehouse
Detailed Requirements:
1. Barebone: 8U/6U High-Density Dual-Socket Enterprise AI Server Chassis x 1 Unit
2. GPU: NVIDIA HGX H200 8-GPU SXM5 141GB 700W Assembly (with 4x NVSwitches) x 1 Set
3. CPU: Dual Intel Xeon Platinum 8592+ 1.9G 64C/128T 350W (128 Cores total) x 2 CPUs
4. RAM: 128GB DDR5-5600 ECC RDIMM x 32 DIMMs (4,096 GB / 4.0 TB populated)
5. Boot: Dual M.2 NVMe 960GB Hardware RAID 1 Mirror x 1 Set
6. NVMe: 15.36TB PCIe 5.0 x4 U.3 Enterprise NVMe SSDs x 4 Drives (61.44 TB Fast Pool)
7. 100G NIC: Dual-Port 100GbE QSFP56 PCIe Full-Height x 2 NICs (with 4x 100G Optics)
8. 25G NIC: Dual-Port 10/25GbE SFP28 PCIe Full-Height x 2 NICs (with 4x 25G Optics)
9. 10G NIC: Quad-Port 10GbE RJ45 Base-T Network Adapter x 1 NIC
10. Power: 8x 3,000W Titanium Level Hot-Plug Power Supplies (4+4 Redundant) x 8 PSUs
11. Software: NVIDIA AI Enterprise (NVAIE) 3-Year Subscription (24x7 Support) x 8 Licenses
Condition: 100% Factory-Sealed Brand New Original. Pre-shipment SN/GUID verification required.
Please quote net USD price and earliest ship date. Exclusive WeChat Liaison: asianeventspk