



**INDIAN INSTITUTE OF TECHNOLOGY BOMBAY
MATERIALS MANAGEMENT DIVISION
Powai, Mumbai 400076.**

PR No. 1000043725 Rfx No. 6100002038:

Detailed Technical Specifications of DGX Server (Qty: 1 Nos):

Processor	Dual Intel® Xeon® Platinum 8570 Processors, 112 Cores total, 2.1 GHz (Base), 4 GHz (Max Boost) or its subsequent equivalent	Compliance (Yes/No)	Additional Information if any
GPUs	8 x NVIDIA Blackwell B200 GPUs, 1,440GB total, 64TB/s HBM3e bandwidth or its subsequent equivalent		
Performance	72 petaFLOPS FP8 training and 144 petaFLOPS FP4 inference		
Benchmarks – (Machine Learning Commons Training v4.1 or latest) or its subsequent equivalent	<p>1. Task: - Language - NLP Dataset: -Wikipedia 2020/01/01 Time: - 3.795 minutes.</p> <p>2. Task: - Commerce - Recommendation Dataset: - Criteo 4TB multi-hot Time: - 2.356 minutes</p> <p>3. Task: - Graph neural network Dataset: - IGBH-Full Time: - 5.502 minutes</p> <p>4. Task:- Language - LLM Dataset :- C4 Time:- 193.738 minutes</p> <p>5. Task :- Language - LLM finetuning Dataset :- SCROLLS GovReport Time :- 12.958 minutes</p> <p>6. Task :- Vision - Object detection (light weight) Dataset :- Open Images Time :- 22.513 minutes</p> <p>7. Task :- Image Generation Dataset :- LAION-400M-filtered Time :- 19.517 minutes</p>		
Memory	2TB ECC DDR5 Memory Upgradable to 4TB DDR5 Memory Module.		

equivalent			
NVIDIA NVLink Bandwidth or its	14.4 TB/s aggregate bandwidth		
subsequent equivalent			
SYSTEM POWER Usage	~14.3kW max		
Management network	10Gb/s onboard NIC with RJ45, 100Gb/s dual-port ethernet NIC Host baseboard management controller (BMC) with RJ45		
Storage	OS: 2x 1.9TB NVMe M.2 and 8x 3.84TB NVMe U.2		
Software or its subsequent equivalent	NVIDIA AI Enterprise – Optimized AI Software NVIDIA Base Command – Orchestration, Scheduling, and Cluster Management DGX OS / Ubuntu – Operating System		
Form Factor	10 RU		
Warranty	Three-years warranty with Support for hardware and software 24/7 Enterprise Support portal access, Live agent support during local business hours		

Eligibility Criteria:

Criteria No.	Particulars	Suggested supporting document	Compliance (Yes/No)
1	The bidder should have a local registered office in Mumbai for faster response or within the radius of 200 kms. is mandatory.	Shop & Establishment Certificate required.	
2	The bidder should have local support engineers in the same Region for faster response. These engineers should be technically qualified and capable to handle the installation of various scientific, AI applications and Cloud Infra.	Submit the escalation matrix with engineer details.	
3	Power capacity and cooling BTU (British Thermal Units) for the proposed solution.	Detailed specifications to be submitted.	
4	MAF from the accelerator OEM as well as the server OEM.	Valid documents need to be submitted.	

5	The Bidders should have done the installations of minimum 3 similar GPUs in other Academic Institutes / R&D or other Government Organizations.	Purchase order copy with Installation report need to submit	
6	Bidders have to have Commercial License for Software Stack.	Relevant document	