Technical Specification for GPU Cluster

S.NO	PARAMETER	SPECIFICATION	COMPLIED/ NOT COMPLIED	REFERENCE PAGE NO
1	**GPUs**	8 X NVIDIA A100 TENSOR CORE GPUs		
2	**GPU Memory**	640 GB in Total (8X80GB)		
3	**Performance**	5 petaFLOPS AI / 10 petaFLOPS INT8		
4	**Processor**	Dual AMD Rome 7742, 128 cores total, 2.25 GHz (base), 3.4 GHz (max boost)		
5	**System Memory**	1 TB		
6	**System Network**	8x Single-Port Mellanox ConnectX-6 VPI 200Gb/s HDR InfiniBand, 1x Dual-Port Mellanox ConnectX-6 VPI 10/25/50/100/200Gb/s Ethernet		
7	**Storage**	OS: 2x 1.92TB M.2 NVME drives Internal Storage: 15TB (4x 3.84TB) U.2 NVME drives		
8	**TF32 Tensor Core**	312 x 8 Tera FLOPS (including sparsity)		
9	**BFLOAT16 Performance**	BFLOAT16 Performance – 624 x 8 Tera FLOPS (including sparsity)		
10	**FP64 HPC Perf Score**	19.5 Tera FLOPS		
11	**CUDA Cores**	10368 per GPU (3,456 FP64 CUDA Cores and 6,912 FP32 Cuda Cores)		
12	**Tensor Cores**	432 per GPU		
13	**Power Requirements**	6.5 kW Max		
14	**Rack space**	6U		
15	**GPU communications protocol**	NV Switches 6		

		NV Link x 12 (600 GB/s)		
16	**Multi-instance GPUs**	Various instance sizes with up to 7MIGs @5GB (Per GPU)		
17	**Interconnect**	PCIe Gen4 64 GB/s		
18	**OS Support**	Ubuntu Linux OS		
19	**USB Port**	4		
20	**RS232 Serial Port**	1		
21	**VGA Port**	1		
22	**Ethernet (RJ45) Ports**	2		
23	**Operating Temperature Range**	Normal AC temperature		
24	**Software Support**	CUDA toolkit CUDA tuned Neural Network (cuDNN) Primitives TensorRT Inference Engine. DeepStream SDK Video Analytics CUDA tuned BLAS CUDA tuned Sparse Matrix Operations (cuSPARSE) Multi-GPU Communications (NCCL), Kubernetes Tensor Flow, caffe, pytorch, Theano, Keras, caffe2, CNTK		
25	**Throughput of pre-training aBERT model (for language modelling) on ~4700 million tokens**	1289 sequences/second		
26	**Throughput of training a Pagerank graph analytics model on 2.6TB**			

	Graph data with 4 such servers	688 billion graph edges/second		
27	**Warranty**	3 Years		

S.NO	SPECIFICATION	COMPLIED/ NOT COMPLIED	REFERENCE PAGE NO
ADDITIONAL TERMS AND CONDITIONS			
1	Vendors should provide continuous technical support and maintenance of equipment.		
2	Vendors must provide detailed documentation for the equipment		
3	Vendors may be called to visit and give a presentation/demonstration on the equipment after opening the technical bid. They need to provide the approximate date for this presentation in the bid. The time period for this presentation would be 14 days from the date of opening of the bid.		
4	Vendors must provide training to our technical staff for using the equipment		
5	All the expenses for installation, training and post sales technical support will be borne by the vendor.		

S.NO	SPECIFICATION	COMPLIED/ NOT COMPLIED	REFERENCE PAGE NO
ELIGIBILITY			
1	The bidder/OEM should have supplied at least 3 similar items to IITs, NITs, IISERs, CSIR Labs or other Govt. R&D organizations in the last 5 years, PO copies or installation certificates along with contact details of end user need to be submitted as the proof of supply. IIT Madras reserves its right to verify the claims submitted by the bidder and the feedback from the previous customers will be part of technical evaluation.		

(Note: It is mandatory for the bidders to provide the compliance statement (comply/not comply) for the Below points with document proof as required).If the compliance statement (comply/Not comply) is not furnished for the evaluation. Bidders will be disqualified