



INDRAPRASTHA INSTITUTE of
INFORMATION TECHNOLOGY
DELHI

**TENDER FOR SUPPLY & INSTALLATION OF AI
CLUSTER FOR CAMPUS OF THE INSTITUTE AT
OKHLA PHASE-III, NEW DELHI-110020**

(IIIT-D/IT/Servers/016/2024-25)

INVITATION FOR BIDS

Indraprastha Institute of Information Technology -Delhi (IIITD), a State University created by an Act of Govt. of NCT of Delhi, invites sealed bids for **Supply and installation of AI Cluster** (as per specifications mentioned under Scope of Work below) for its campus at Okhla Phase-III New Delhi-110020.

1. An amount of Rs. 2,50,000- (Two Lakhs Fifty Thousand Only) towards earnest money (EMD) must be deposited in the form of demand draft in favour of "IIIT-Delhi Collections" account, payable at New Delhi. No interest will be paid on the earnest money deposited by the bidder. Tender Document without earnest money will be summarily rejected. EMD is exempted for MSMEs/NSIC registered suppliers
2. The tender document can be downloaded from the Institute's website. Those wishing to get the copy of the document from the Institute may please deposit a non-refundable Tender Document Fee of Rs.1, 180/- (Rupees One Thousand One Hundred Eighty Only) in the form of Demand draft drawn in favour of IIIT-Delhi Collections", payable at New Delhi or may deposit Rs.1, 180/-with the F&A division of the Institute and enclose the receipt with the filled up tender document. The tender fee is exempted for MSMEs/NSIC registered suppliers.
3. The last date for submission of Bid is 27th September 2024 up to 3:00 PM. The Technical Bids shall be opened on the same day, i.e., on 27th September 2024 at 3.30 PM. The Tender Document should be addressed to:

**Registrar,
Indraprastha Institute of Information Technology-Delhi
Okhla Phase-III
(Behind Govind Puri Metro Station)
New Delhi-110020.**

The document should be deposited in the Tender Box kept in the Store & Purchase department in Room no. A 108, First Floor, Old Academic Block of the Institute. Bids received after 3:00 PM will not be accepted or considered under any circumstances.

Bidding Procedure:

1. Bids are invited in Two Bids System, i.e. (1) Technical and (2) Financial.

Technical and Financial bids should be sealed separately and enclosed in a sealed envelope clearly indicating separately Technical Bid for "**Supply and installation of AI Cluster**" and Financial Bid for "**Supply and installation of AI Cluster**" addressed to Registrar IIIT-Delhi, Okhla Industrial Area Phase-III, New Delhi-110020.

2. Sealed quotations shall be received no later than 3.00 PM on 27th September 2024. No bids will be accepted after this date & time under any circumstances. The Institute will not be responsible for any postal/courier delay and reasons beyond the Institute's control.
3. Technical bids must contain the EMD for specified amount, along with complete technical details as desired by this tender. Technical bids of all the tenderers will be opened on pre scheduled date, time & venue. Technical bids without EMD will be summarily rejected. The financial bid will be opened after evaluation of the technical bid. Financial bid of only those meeting the requirement of the Institute will be opened and no representation in this regard will be entertained. The date, time of opening of financial bid will be communicated later. The EMD is exempted for MSMEs/NSIC registered suppliers.

Scope of Work

Specification of the Master Server (Qty:01)

S.no	Specifications	Parameter
1	Processor	Dual Intel Xeon Platinum 4 th Gen processor having at least with 32 cores each and at least 2.0 GHz base clock or Equivalent AMD 9000 Series processor
2	Motherboard	Appropriate Motherboard and chipset.
3	Memory	Minimum of 512 GB DDR5 ECC RAM (Upgradeable to 1TB)
4	Storage (OS)	2 x 1.92 TB NVMe M.2 SSD in RAID 1 array
5	Network 1	400Gbps
6	Network 2	10Gb/s onboard NIC with RJ45
7	I/O ports	At least: 2x/4x USB 2.0/3.0, and,1xVGA/HDMI
8	Cooling	Appropriate cooling must be provided within the chassis to cool all the parts of the system. There must not be any failures and performance issues due to heating.
9	Remote Management	Server should have a dedicated 1Gbps remote management port. It should support remote power on/off, virtual media, virtual console, virtual KVM, hardware monitoring.
10	Power Supply	Appropriate energy efficient redundant (N+N) hot swappable power supply to support the configuration and the GPUs.
11	Form Factor	1U/2U
12	OS Support	RHEL 8/RHEL9, Ubuntu 20.04 LTS Server, Ubuntu server 22.04 LTS
13	Warranty	5 Years Onsite Comprehensive warranty, NBD support (Parts & Labor)
14	Note	a). All Cables such as power cables for C-19 socket, GPU power cables, rack mounting kits or any other accessories should be provided. b) Supplied server must show quoted OEM make, model and Serial number of the server in system bios. c) The Offered System must not be End of Life, End of support, End of Sale.

Specification of the GPU Rack Server (Qty:01)

S.no	Specifications	Parameter
1	Processor	Dual Intel Xeon Platinum 4 th Gen processor having at least with 56 cores each and at least 2.0 GHz base clock or Equivalent AMD 9000 Series processor
2	Motherboard	Appropriate Motherboard and chipset.
3	Memory	Minimum of 2TB GB DDR5 ECC RAM
4	Storage (OS)	2 x 1.92 TB NVMe M.2 SSD in RAID 1 array
5	GPUs	8 x NVIDIA H100 GPUs that provide 640 GB total GPU memory
6	NVSwitch	4 x 4th generation NVLink that provide 900 GB/s GPU-to-GPU bandwidth
7	Network 1	4x OSFP ports serving 8x single-port NVIDIA ConnectX-7 VPI > Up to 400Gb/s InfiniBand/Ethernet 2x dual-port QSFP112 NVIDIA ConnectX-7 VPI > Up to 400Gb/s InfiniBand/Ethernet
8	Network 2	10Gb/s onboard NIC with RJ45 100Gb/s Ethernet NIC
9	I/O ports	At least: 2x/4x USB 2.0/3.0, and,1xVGA/HDMI
10	Cooling	Appropriate cooling must be provided with in the chassis to cool all the parts of the system. There must not be any failures and performance issues due to heating.
11	Remote Management	Server should have dedicated 1Gbps remote management port. It should support remote power on/off, virtual media, virtual console, virtual KVM, hardware monitoring.

12	Power Supply	Appropriate energy efficient redundant (N+N) hot swappable power supply to support the configuration and the GPUs.
13	Form Factor	2U/4U/6U/8U/10U or better
14	OS Support	RHEL 8/RHEL9, Ubuntu 20.04 LTS Server, Ubuntu server 22.04 LTS
15	Warranty	5 Years Onsite Comprehensive warranty, NBD support (Parts & Labor)
16	Note	<p>a). All Cables such as power cables for C-19 socket, GPU power cables, rack mounting kits or any other accessories should be provided.</p> <p>b) Supplied server must show quoted OEM make, model and Serial number of the server in system bios.</p> <p>c) The Offered System must not be End of Life, End of support, End of Sale.</p>

Specification of the PFS (Parallel File System) Storage (Qty:01)

Sr. no	Parameters	Minimum technical Specifications
1.	Storage features	The solution should deliver at least 4 nines (99.99) uptime reliability.
		The solution should be software/hardware defined storage and run on multiple industry standard platforms to provide multiple vendor choice and avoid vendor lock-in.
		Optionally The solution should consist of one native filesystem for simplicity, performance, and reliability - and not be built on top of another filesystem like ext4, XFS, or ZFS which adds complexity, slows performance and degrades reliability.
		The solution should be scale-out storage that is fully integrated may requiring separate metadata servers, administrative servers, or protocol gateways
		The solution should provide a single namespace that can scale to EB's
		The solution shall provide a minimum 500 TB of usable capacity using NVMe SSDs only.
		System must support 4K block sizes, matching NVMe SSD media block sizing for optimum performance and efficiency.
		The solution shall support both InfiniBand and Ethernet networking
		The solution shall support Kubernetes Container Storage Integration (CSI)
		The solution must be POSIX compliant and support NFS, SMB, S3 protocols and GPU Direct storage (GDS)
		All system features and data services must be included in the base license with no additional upcharges.
2.	Storage Management	Optionally the solution shall provide a full-featured GUI in addition to a command line interface. All tuning parameters must be accessible through the GUI
		The management dashboard should have health and performance reporting, visualization, and overall system management functions are accessible using the command line interface (CLI) or the intuitive graphical user interface (GUI) management console. The system resources required for management and health monitoring in PFS may be considered additional in the solution to meet the performance requirements of the RFP.

		The solution shall allow for expansion of a filesystem without downtime
		The solution shall support user authentication via LDAP and Active Directory
		The solution should have role-based access control (RBAC) to restrict access to authorized end users
3.	Storage Performance	All data and metadata are distributed evenly across the storage cluster delivering performance, without any storage hot spots or I/O bottlenecks as asked in benchmarks in Annexures
		The solution shall provide a minimum bandwidth of at least 400 Gb/second of reads (100%) and 400 Gb/sec writes (100%).
		The solution shall provide a performance of 4K IOPS of at least one Million IOPs (reads)
		The solution shall support both high transactional and bandwidth-based performance on a single filesystem. All performance specifications shall refer to a single filesystem that has not been further configured, adjusted, or tuned between runs of different benchmarks or I/O loads.
4.	Data Protection Features	Optionally the system shall support encryption at rest and in flight
		The solution shall support snapshots and clones of an entire filesystem without performance degradation
		Snapshots and clones occur instantaneously and are differential after the first instance without any performance degradation.
		Support to take snapshots and clones on a per file system basis.
		Ability to recover the entire filesystem or individual file from snapshots.
		The solution shall support the concurrent failures without loss of access to data
		The solution shall support end-to-end data protection
		The solution must use checksum to protect against bit flip or a drive fault to protect against silent data corruption
		The solution must have intelligent rebuild mechanism to rebuild only the data that is actively stored on the failed server or SSD
		All storage nodes, if any, must participate in the recovery process during rebuild for faster recovery.
5.	Warranty	5 Years Comprehensive On Site Warranty on next business day support basis.

Network Switches (Qty: 01)

S.No	Item	Description of Requirement
1.	Storage and compute node Communication	<ul style="list-style-type: none"> - Switches should have 64-ports NDR, InfiniBand Switch. - All ports on front, Redundant PSUs, and redundant fans. <p>Performance: - 400Gb/s per port</p> <p>Switch radix: - 64 400Gb/s non-blocking ports with aggregate data throughput up to 51.2Tb/s</p> <p>All cables and adapter cards required for connecting all storage nodes quoted in above should be included/bundled.</p> <p>5 Years Comprehensive On Site Warranty on next business day support basis.</p>

Server Racks (Qty: 02)

1. The racks should be quoted with smart PDUs and it should be from Netrack, PDUs should show the power being utilized on a panel embedded on it. The rack should have ample space and power sockets to accommodate three such servers asked in specification in a single rack. And there should be at least ten C-19 sockets on each PDU. The depth of the rack should be 1200mm.

Implementation Work

- a) All the components shall be supplied, installed, integrated & commissioned along with benchmarking specified.
- b) The PFS storage and servers, switches to be integrated with the cluster.
- c) Scheduling of workloads in cluster should be demonstrated.
- d) All activities to be executed for integration and persistent mounting of storage on base operating system and containers on Compute and Master Server.
- e) Workloads should be able to use PFS storage from any of the compute nodes.
- f) Flexibility of scheduling on any node should be demonstrated.
- g) Movement of workloads from one compute node to another node to be demonstrated in case of failure of a compute node.
- h) Interconnect cables should be provided by the bidder for the complete solution.
- i) Entire solution must be quoted with five years of 24x7 support with Next Business Day parts delivery. All software & firmware upgrades and updates for storage array, parallel file system and monitoring & management components must be freely available to customer without any additional cost during warranty period
- j) IIIT-Delhi will decide whether to implement Kubernetes cluster or any other job scheduler for E.g. Slurm/Open PBS scheduler at the time of implementation

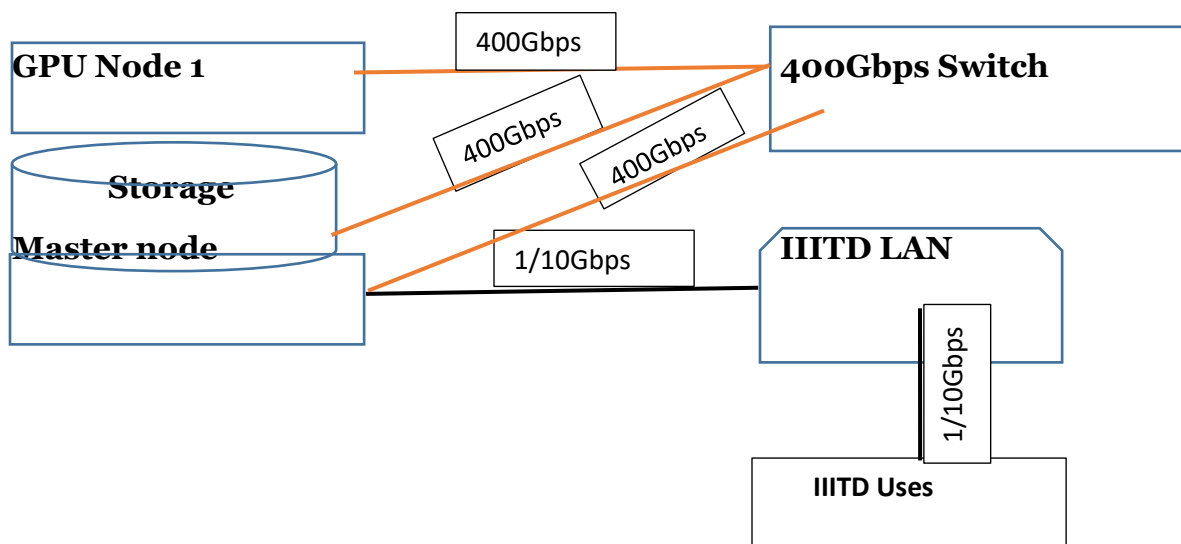
Training:

- a) Installation, Integration and hands-on training shall be carried out for all compute, networking and storage equipment.
- b) Admin: Two days on administrators including scheduler implementation and management, system configuration details and troubleshooting
- c) End User training: Two days training to users on job submission, basic model training & Inferencing, Optimization parameters, Best practices of using GPU Controllers / Accelerators & multi GPU Controllers / Accelerators. Remote access and usage of Systems through remote.

All software updates Software (both major and minor releases), Firmware's, Subscriptions during the warranty period Bidder need to provide without any additional costs.

All the licenses quoted shall be perpetual to use

Connectivity diagram



SLA

1. Hardware replacement should be done on the next business day.
2. The overall warranty for all active equipment should be 5 years.
3. The troubleshooting support should respond within 2hrs and should be resolved within 4hrs(except hardware replacement).
4. The complete installation must be completed within one month after the delivery

Minimum Eligibility Requirement:

- 1 Bidder should be OEM/Authorized Partner/service provider of the OEM, here and after referred as bidder. In case the Bidder is an Authorized Partner or Service Provider, a valid Agency-ship/Dealership Certificate (**MAF specific to this Tender**) to quote on behalf of OEM should also be enclosed along with the technical Bid. **A document in support of this must be enclosed. [Bidder should attach as Annexure 1]**
- 2 OEM & Authorized Partner should have Sales and support office in India. **A self-certified document in support of this must be enclosed. [Bidder should attach as Annexure 2]**
- 3 OEM or Authorized Partner should have a service and support office in Delhi NCR. **A self-certified document in support of this must be enclosed. [Bidder should attach as Annexure 3]**
- 4 The warranty provided by the Bidder should have a back-to-back arrangement with the OEM. **The declaration should be part of a Letter of Authorization and signed by a competent authority at the OEM. [Bidder should attach as Annexure 4]**
- 5 The Bidder should be ISO 9001 certified. **A copy of the ISO Certificate should be enclosed. [Bidder should attach as Annexure 5]**
- 6 The Bidder should have a support center with a minimum of 3 relevant support engineers. **A self-certified document in support of this must be enclosed. [Bidder should attach as Annexure 6]**
- 7 The bidder should be able to provide 24x7 NOC & Tele support of their own if required by IIITD at agreed terms. **A self-certified document in support of this must be enclosed. [Bidder should attach as Annexure 7]**
- 8 The Bidder shall provide the Registration number of the firm along with the valid GST number with the PAN Number allotted by the competent authorities. **A self-certified document in support of this must be enclosed. [Bidder should attach as Annexure 8]**
- 9 The Bidder must not be blacklisted by the Central Government, State Government, or Government of Corporations in India. **A certificate or undertaking to this effect must be submitted. [Bidder should attach as Annexure 9]**
- 10 If the Bidder is an authorized partner or service provider of an OEM, **an undertaking from the OEM is required** (please enclose) stating that they would facilitate the Bidder regularly with technology/product updates and extends support for the warranty as well. **[Bidder should attach as Annexure 10]**
- 11 The Bidder must be responsible for supply, deploy and support the infrastructure. **[Bidder should attach as Annexure 11]**
- 12 If bidder does not meet its SLA, IIITD will put the fine of Rs. 5,000/- per day for the first seven days and @ Rs.8, 000/- per day from 8th day onwards will be levied. **[Bidder should attach as Annexure 12]**

SLA	
Support	Response in 4 hrs.
Hardware replacement	Next business day

- 13 Bidders can seek clarifications, raise technical, financial queries, etc., related to the tender at tenders@iiitd.ac.in. Inquiries may be submitted within a seven-day period following the publication of the bid.

The following information must accompany the financial Bid:

1.	Name, address, and telephone number of the firm/company	
2	Name of the contact person and contact details (mobile/telephone number etc.)	
3	Name of the Bank and full address	
4	Bank Account Number	
5	PAN & GSTIN (Attach self-certified copy)	
6	Valid self-certified copy of authorization from Bidder	
7	Copy of Partnership Deed/ Certificate of registration of the company or any other document evidencing the registration of the Bidder	
8	Number of Years of Experience	
9	Details of DD towards Tender Fee: EMD:	
10	Provide the previous PO's of the same items work executed during the last three years (attested copies of the Orders to be enclosed)	
11	List of service centers, nearest location of the support centre.	
12	Turnover of the Bidder in the financial years: 2021-22 2022-23 2023-24 Please attach CA certified copy of the turnover.	

I /We hereby certify that the information furnished above is full and correct to the best of my/our knowledge.

(Signature of the Authorized Signatory)

Name:

Office Seal.

Date:

Place:

TERMS AND CONDITIONS

1. The Financial Bid should be valid for a period of not less than 60 days from the date of opening of Bid.
2. Upon placing of the Purchase Order (PO), the successful Bidder is required to submit a performance bank guarantee (PBG) equivalent to 5% of the PO value within 15 days from the date of PO, failing which the Bidder shall be notified as blacklisted. The PBG will be valid for a period of 60 days beyond the stipulated date for cessation of the contract, which is co-terminus with the warranty period. No interest is payable on the PBG.
3. PBG should be in favour of “IIIT-Delhi Collections” payable at New Delhi. PBG will be realized by IIIT-D in case of termination of the contract for unsatisfactory performance and/or non-performance of the contract.
4. The Product to be supplied within a period of 150 days from the date of the Purchase Order by the Institute. If the vendor fails to supply the item as quoted in the Technical & Financial bid, the EMD amount will be forfeited, and the Bidder shall be notified as blacklisted or as deemed appropriate by the Institute. In case of late delivery the penalty will be considered as a fine of Rs. 5,000/- Rs per day for the first seven days and @ Rs. 8,000/- per day from the 8th day onwards will be levied for a maximum of 15 days. Following this period, the order will be canceled, and the earnest money deposit (EMD) and performance bank guarantee (PBG) amount will be forfeited.
5. The Bidder should have their own test and repair facility with certified engineers.
6. 100% payment will be released only on satisfactory installation/services as per the scope of work as certified by the officer in charge of the Institute and after producing the GST invoice. Bidders who do not agree to the above payment terms are requested not to submit their Bid.
7. Payment will be made only after completion of the installation and commissioning. No advance and part payment will be released.
8. Sealed bid can be sent either by post or by messenger. The responsibility of delivery of the bid lies entirely with the bidder
9. In the event of a dispute, Director IIIT-Delhi shall be the sole arbitrator, and his decision shall be final and binding on both parties.
10. IIIT-Delhi does not bind itself to accept the lowest or any other offer and reserves the right to accept or reject any or all the offers either in full or in part without assigning any reason.
11. In case the Bidder is not able to execute the work as per terms, EMD/PBG shall be forfeited.
12. Bidders should provide details of its support, escalation matrix, certification to this effect from itself. If Bidder fails to meet the SLAs, a fine of Rs. 5,000/- Rs per day for the first seven days and @ Rs. 5,000/- per day from the 8th day onwards will be levied.

13. The bidder/tenderer shall submit an undertaking on its letter head, duly signed and stamped, that none of the staff, faculty members, relatives, etc. of the Indraprastha Institute of Information Technology-Delhi are related directly or indirectly to any employees, Directors, or Key Managerial Personnel, etc. of the bidder/tenderer. In the event of the IIIT-D coming to know or pointed about the same, the bidder/tenderer undertakes to deposit a sum of Rs.1,00,000/- (Rs. One Lakh only) as a penalty with the Institute. Such bidders/tenderers shall be liable to be blacklisted and announced on the website of IIIT-D.
14. Supply, installation and commissioning will be at IIIT-Delhi campus.
15. The selected bidder has to sign an agreement with IIITD on a stamp paper after issue of Purchase/Work order. The agreement is attached Annexure 'Y'

QCBS evaluation method

Quality and Cost Based Selection Criterion

Technical Evaluation - Total 40 Points					
S. No	Particulars	Points System	Max Points	Points Awarded (Not to be filled by the bidder)	Supporting Documents Attached
1	No. of deviations cited in the Technical Specifications Sheet (GPU Rack Server) submitted by the bidder/OEM.	0 10 points 1-3 7 points 4-6 3 points >6 0 point	10		
2	No. of deviations cited in the Technical Specifications Sheet (PFS Storage, Network Switches) submitted by the bidder/OEM.	0 10 points 1-2 7 points 3 or more 0 points	10		
3	Performance benchmarks as stated in Annexure-II	<p>Passing all the 2 benchmarking criteria and</p> <ul style="list-style-type: none"> ● the best* performer (may be multiple) 5 points ● second best performer (may be multiple) 4 points; ● third best performer (may be multiple) 3 points 	5		

		<p>Passing 1 out of the 2 benchmarking criteria and</p> <ul style="list-style-type: none"> ● the best* performer (may be multiple) 2 points ● second best performer (may be multiple) 1 points; 			
		<p>Passing none of the 2 benchmarking criteria 0 points.</p> <p>All of the above benchmarking results should be supported by detailed log files, which will be used for final evaluation. Failure to do so will lead to disqualification.</p>			
4	<p>Performance benchmarks as stated under the Application Testing setting described in Annexure-I.</p>	<p>Passing all the 11 benchmarking criteria and</p> <ul style="list-style-type: none"> ● the best* performer (may be multiple) 15 points ● second best performer (may be multiple) 12 points; ● third best performer (may be multiple) 10 points ● fourth best and beyond (may be multiple) 8 points 	15		

Passing **at least 8 out of 11** benchmarking criteria **and**

- the best* performer (may be multiple) | 7 points;
- the second best performer (may be multiple) | 6 points;
- the third best and beyond (may be multiple) | 5 points

Passing **at least 5 out of 11** benchmarking criteria **and**

- the best* performer (may be multiple) | 4 points;
- the second best performer (may be multiple) | 3 points;
- the third best performer and beyond (may be multiple) | 2 points

Passing fewer than 5 out of 11 benchmarking criteria | 0 points. Any of the above benchmarking results **that do not meet the quality target** as specified in **Annexure-I** will be assigned **0**

		<p>points. All of the above benchmarking results should be supported by detailed log files, which will be used for final evaluation. Failure to do so will lead to disqualification.</p>		
Technical Evaluation : Total Points Awarded (A)		40		

* best performer is computed based on average relative improvement.

For the training time benchmark:

- average_training_time_score** = $\text{average}_{j=1}^8 \left(\frac{\Delta t_{\text{task}_j}}{\text{spec}_t_{\text{task}_j}} \right)$, where $\Delta t_{\text{task}_j} = (\text{spec}_t_{\text{task}_j} - \text{achieved}_t_{\text{task}_j})$, where $\text{spec}_t_{\text{task}_j}$ is the training time specified in Annexure-I for the MLPerf Training Benchmarks.

For the inference benchmark:

- average_inference_time_score** = $\text{average}_{j=1}^3 \left(\frac{\Delta \text{tp}_{\text{task}_j}}{\text{spec}_{\text{tp}_{\text{task}_j}}} \right)$, where $\Delta \text{tp}_{\text{task}_j} = (\text{achieved}_{\text{tp}_{\text{task}_j}} - \text{spec}_{\text{tp}_{\text{task}_j}})$, where $\text{spec}_{\text{tp}_{\text{task}_j}}$ is the inference time specified in Annexure-I for the MLPerf Inference Benchmarks.

Final performance score used for identifying best performer and following ranks

- average_score** = $0.7 * \text{average_training_time_score} + 0.3 * \text{average_inference_time_score}$

Presentation Evaluation - Total 10 Points					
S. No	Particulars	Points System (Excellent - 2 Good - 1 Poor - 0)	Max Points	Points Awarded (Not to be filled by the bidder)	Supporting Documents Attached
1	Clarity of the Technical Proposal and overall approach to implementation.		2		
2	Type and quality of User Interface including ease of operation, value added features, and innovative implementation plan to enable scalability and cost efficiency.		2		
3	Project Timelines and lead time for the supply of hardware.		2		
4	Plan for extending software & training support.		2		
5	Proposal for connecting with existing data center infrastructure (already at IIT-Delhi)		2		
Presentation Evaluation : Total Points Awarded (B)			10		
Overall Points Awarded (A + B) - Total 50 points					
Note: The bidder must score a minimum of 28 points (70%) in Technical Evaluation (A) and a minimum of 5 points (50%) in Presentation Evaluation (B) to qualify. The financial bids, for the bidders who do not qualify will not be opened and therefore, will not be considered as a part of the financial bid evaluation.					

Quality score (QS) = (Technical Evaluation + Presentation Evaluation)/50*75

Financial score (FS) = (FP/F) * 25 , where FP is the lowest financial bid and F is the financial bid under consideration.

Final score = 75*QS + 25*FS

75% weightage for Quality score and 25% weightage for financial score.

Annexure - I: Performance Benchmarks

The performance benchmarking will be done using the well accepted MLCommons benchmarking scheme. More details are available at <https://mlcommons.org/>

Benchmarks required for QCBS:

1. MLPerf Training v3.1: There are 8 different benchmarks that need to be reported as shown in Table 1.
<https://mlcommons.org/benchmarks/training/>
2. MLPerf Inference v3.1: There are 4 different benchmarks that need to be reported as shown in Table 2.
<https://mlcommons.org/benchmarks/inference-datacenter/>
3. The aforementioned 12 benchmarks (from 1. and 2. above) need to be reported under the Application Testing setting described below.

Application Testing (System with Compute, Networking and Storage nodes)

Complete system performance in terms of latency must be within 5% of MLPerf Training and Inference benchmarks **when the data is stored on the storage nodes**. For example, for the ‘ResNet-50 v.15’ model training, the training time should be at most 15.75 minutes ($(15.75-15)/15=0.75/15 = 0.05$, i.e., 5%).

Table 1. MLPerf Training v3.1: Training time (in minutes) requirements for the benchmark tasks descriptions listed below

BERT-Large	ResNet-50 v1.5	RNN-T	3D-UNet	Stable Diffusion-V2
Training Time (min) 24 x H100 GPUs				
-	-	-	5.5	-
Training Time (min) 16 x H100 GPUs				
4	8	12	-	-
Training Time (min) 8 x H100 GPUs				
-	15	17	14	47

Task Description

Benchmark	Dataset	Quality Target	Model	Ver
Image classification	ImageNet	75.90% classification	ResNet-50 v1.5	v3.1
Image segmentation (medical)	KiTS19	0.908 Mean DICE score	3D U-Net	v3.1
Speech recognition	LibriSpeech	0.058 Word Error Rate	RNN-T	v3.1
NLP	Wikipedia 2020/01/01	0.72 Mask-LM accuracy	BERT-large	v3.1
Image Generation	LAION-400M- filtered	FID<=90 and CLIP>=0.15	Stable Diffusion-V2	v3.1

MLPerf Inference

<https://mlcommons.org/benchmarks/inference-datacenter/>

Table 2. MLPerf Inference v3.1: Inference time requirements for the benchmark tasks descriptions listed below

Inference Time (samples/s) 8 x H100 GPUs			
BERT- Large	ResNet-50 v1.5	RNN-T	3D- UNet
65-70k	650-700k	160-175k	50

Task Description

Task	Model	Dataset	QSL Size	Quality	Server latency	Ver
Image classification	Resnet 50 v1.5	ImageNet (224×224)	1024	99% of FP32 (76.46%)	15 ms	v3.1
Speech-to-text	RNNT	Librispeech dev-clean (samples < 15 seconds)	2513	99% of FP32 (1 – WER, where WER=7.452253714852645%)	1000 ms	v3.1
Language processing	BERT-large	SQuAD v1.1 (max_seq_len=384)	10833	99% of FP32 and 99.9% of FP32 (f1_score=90.874%)	130 ms	v3.1

Annexure - II: Storage Performance Benchmarks

MLPerf Storage

<https://mlcommons.org/benchmarks/storage/>

Model	#Simulated Accelerators	Data	Throughput samples/s, MB/s
3D UNET (146MB/sample)	20	3.4TB	54, 7545
BERT-large (2.5KB/sample)	24	2.4TB	1170, 2.8

Task Description

Task	Model	Nominal Dataset	Ver
Medical image segmentation	3D UNET	KITS 2019 (602x512x512)	v0.5
Language processing	BERT-large	Wikipedia (2.5KB/sample)	v0.5

PROFORMA FOR FINANCIAL BID

S. No.	Details	Qty.	All Inclusive Cost (Please quote in INR only)
1	Supply and installation of AI Cluster(As per scope of work)	As per scope of work	
	Discount, if any		
	Total Amount Rs.		

Total Final Cost (in figures) with Installation at IIIT Delhi campus Okhla Phase III, New Delhi.

The Bidder may obtain price in Forex(USD) however the quote in financial Bid should be INR only.

****If any documents are required for availing custom duty exemption, the IIITD will provide the same. Please quote price accordingly.***

Please note the price should be quoted for each of the items and should be inclusive of all taxes/charges and installation at IIIT-Delhi.

The discount, if any, should be mentioned herein and nowhere else.

Total Cost (all-inclusive) of quantity mentioned above (in words) at IIIT-Delhi campus:

We accept that the rate quoted above shall remain valid for a period of 60 days from the last date of the tender document, i.e., 60 days from 27th September 2024. It is certified that the rates quoted above are not more than the rates charged from any Central / State Govt. Deptt. / Institution / GeM.

(Signature and Seal of the Bidder)

Annexure Y

ON NON JUDICIAL STAMP PAPER OF RS 100/-

AN AGREEMENT made on _____ day of _____ two thousand
Twenty Four

BETWEEN

(Hereinafter called the contractor, which expression shall include its proprietor, partners, heirs, executors, administrators, legal representatives, successors and assignees) WITH REGISTERED ADDRESS.....of the one part

AND

The REGISTRAR Indraprastha Institute of Information Technology Delhi, Okhla Industrial Area Phase III, New Delhi - 110020 (hereinafter called the IIITD, which expression shall include its successors and assignees) of the other part.

Whereas the IIITD had invited Bids <Please write tender name> ; vide its Bid Document No. _____, which shall be deemed to be a part of this agreement;
FOR THE SCOPE OF WORK/SUPPLY

Whereas the contractor submitted its TENDER Bid dated _____, a copy of the price bid, submitted by the contractor, is annexed hereto as Annexure;

Whereas the IIITD has accepted the Bid submitted by the contractor, on the terms and conditions mentioned in the IIITD's said Bid Document and conveyed its acceptance to the contractor; vide its letter No. _____ dated _____, AND
ANY OTHER CORRESPONDENCEwhich shall be deemed to be a part of this agreement;

Whereas the contractor is agreeable to the terms and conditions mentioned in the IIITD's said Bid document;

Whereas the contractor undertakes to comply with all relevant laws like Contract Labour (Regulation and Abolition) Act, 1970; Employees' State Insurance and Miscellaneous Provisions Act, 1952; Employees' State Insurance Act, 1948; Minimum Wages Act, 1948; Payment of Bonus Act, 1972; Payment of Wages Act, 1936; Income Tax Act; GST Act etc. and to indemnify the IIITD from the contractor's acts of omission or commission, as regards the compliance with the relevant laws;

Whereas the contractor declares that he/she/it shall own all responsibility for any act of omission or commission, as regards the compliance with the relevant laws;

AND WHEREAS the IIITD is agreeable to make necessary payment to the Contractor, at the rates mentioned in the Annexure annexed hereto and as per the terms and conditions mentioned in the IIITD's said Bid Document;

In WITNESS whereof Shri (name), (designation), the authorized representative of the contractor, for and on behalf of the contractor, has hereunto set his hand and for and on behalf of the IIITD has hereunto set his hand.

(Signature of the authorized representative of the Contractor)

Name and designation of the contractor's representative

In the presence of

1

2

(Signature of witnesses with full name and full address)

Registrar

for and on behalf of the IIITD

In the presence of

1

2

(Signature of witnesses with full name and full address)