Annexure-II
RFP for HPC cluster

I - General Instructions for Bidders

- The entire solution should be based on certified hardware / software components which are fully validated and supported by the OEM.
- Vendor must quote for all the requirements together viz HPC cluster, storage and Interconnect for the proposed solution. Partial response to the tender will be rejected.
- HPC cluster solution must have dense computing platform based compute nodes.
- All equipment must be compatible with Indian electrical standards and codes. Engineering documentation (in digital form ONLY) on the physical sizes, weights of all major components must be submitted.
- The server and storage management software should provide proactive notification of actual or impending component failure. The OEM should provide the hardware required to host such software.
- The server/chassis management/monitoring software must be from the OEM itself.
- The proposal must include 1) detailed data sheets (in digital form ONLY) for every single proposed component & the necessary technical white papers discussing features, performance and optimized technique 2) pictorial representation of the entire solution indicating required size in U and 3) the power and cooling requirement with supporting documents 4) Complete Bill of Materiel (BOM). The BOM should not have any miscellaneous items.
- The warranty of the total solution shall commence from the date of final acceptance test and sign off on an acceptance report.
- The entire solution should fit within 30 U rack space and its total power requirement should not exceed 15 KW under HPL conditions. The solution exceeding 30U rack space and 15 KW power utilization will not be considered for further evaluation.
- The solution must be implemented within the already existing racks at the IUCAA data centre. The bidder may inspect the IUCAA facility before quoting.

II Abbreviations used:

RDIMM	Registered Dual In-line Memory Module
NSPOF	No Single Point of Failure
IPMI	Intelligent Platform Management Interface
DDR4	Dual Data Rate 4
RHEL	RedHat Enterprise Linux
SLES	Scientific Linux Enterprise Server
HPC	High Performance Computing
EDR	Enhanced Data rate
OEM	Original Equipment Manufacturer
CPU	Central Processing Unit
GUI	Graphic User Interface
CLI	Command Line Interface
HPL	High-Performance Linpack Benchmark
CMS	Cluster Management Software
ECC	Error Correcting Code

LLC	Last Level Cache
RPM	Rotation Per Minute
RFP	Request for Proposal
PFS	Parallel File System
IB	InfiniBand
PO	Purchase Order
HA	High Availability
TF	TeraFlop

Part - 2

Technical Specifications of the proposed Integrated special purpose computing facility for astronomy and astrophysics research.

IUCAA wants to expand the existing 100 TF (Peak performance) CPU based High Performance Computing facility that uses a distributed memory architecture with IB EDR (100 Gbps) Interconnect. The existing all nodes and parallel file system storage (PFS) should be integrated in the proposed system. The proposed system should be scalable using the same architecture and cooling method available at the site.

I - Composition of Cluster nodes

The proposed solution should have two head/master nodes **in HA mode,** two login nodes and 20 compute nodes customized for astronomy and Astrophysics research in dense configuration. Each node should be individually serviceable**.** The bidder is expected to submit the pictorial layout with interconnect and so on of the proposed solution along with relevant product brochures. The bidder is expected to submit the explicit Bill of Materiel (BOM), bids with incomplete BOM will be rejected.

Sr	Technical Specifications	Qty	Unit
1	**Head/Master Nodes:** These nodes with common storage facility would work in HA mode: Software for managing cluster, batch queuing, authentication server etc. will be installed in these nodes. Each node should have: o Dual CPUs–(Intel Ice Lake) Intel Xeon-Gold 6326 (2.9GHz/16-core/185W) or better. o 512 GB (16 X 32GB) memory module ECC DDR4 RDIMM (3200 MHz or better). o Up to 2U rack-mountable form-factor only. o USB 3.0 ports to attach an external CD-DVD reader/writer or with an integrated internal CD-DVD reader/writer. o 4 x 1Gbps RJ45. o 1 x Infiniband EDR or better o 2 X dual port 10Gbps Ethernet HBA with SR transceivers o 2 X 1.2 TB @10K RPM SAS disks; Scalability up to 8 disks within the server. o Appropriate 12 Gbps SAS or 16 Gbps FC to connect to the storage	2	Each
2	**Shared Storage:** o Controller: Dual Controller Storage Array o Cache: Minimum 8GB cache per controller pair o Drives: 8 X 1.8 TB 10K RPM, in Raid 6 with 1 hot spare disk. o Host Interface: Minimum 4 host ports of 12 Gbps SAS or 16 Gbps FC o RAID Support: RAID Level 0, 1, 5 & 6 support o Protocol: The storage should support NFS protocol	1	Each
3	**CPU Compute Nodes:** Each compute node should have o Dual CPUs– (Intel Ice Lake) Intel Xeon-Gold 6326 (2.9GHz/16-core/185W) or better. o Minimum 512 GB (16 X 32GB) memory module ECC DDR4 RDIMM (3200 MHz or better). o 2 X1G Ethernet ports o 1 X Infiniband EDR or better o 1 X 600 GB SAS @ 10K RPM disk.	20	Each
4	**Login/Utility Node:** These nodes offer gateway for users to login / compile codes / submit MPI jobs across the compute nodes/execute mathematical software such as Matlab, Mathematica and IDL installed locally (licenses available at IUCAA). Each node should have o Dual CPUs (Intel Ice Lake) Intel Xeon-Gold 6326 (2.9GHz/16-core/185W) or better. o 512 GB (16 X 32 GB) memory module ECC DDR4 RDIMM (3200 MHz or better). o Up to 2U rack-mountable form-factor only. o USB 3.0 ports to attach an external CD-DVD reader/writer or with an	2	Each

	integrated internal CD-DVD reader/writer. o 4 X 1G Ethernet ports o 1 X Infiniband EDR or better o 2 X dual port 10Gbps Ethernet HBA with SR transceivers and 10 meter FC cable o 2 X 1.2 TB @10K RPM SAS disks; Scalability up to 8 disks within the server.		
5	RHEL Licenses (The total number of licenses should be specified separately in the technical bid. Bundle should contain qty required for the proposed solution)	1	Bundle
6	CMS Licenses (The total number of licenses should be specified separately in the technical bid. Bundle should contain qty required for the proposed solution)	1	Bundle
7	PBS Licenses (The total number of licenses should be specified separately in the technical bid. Bundle should contain qty required for the proposed solution)	1	Bundle
8	Infiniband Switches (The total number of switches should be specified separately in the technical bid. Bundle should contain qty required for the proposed solution)	1	Bundle
9	1G Switches (The total number of switches should be specified separately in the technical bid. Bundle should contain qty required for the proposed solution)	1	Bundle

Additional features expected in the cluster:1

- The total solution should be configured with redundant power supplies. Power supplies can be enclosure based or rack based solution with min. N+1 redundancy.
- Remote management hardware interface at either enclosure or rack level including license.
- IPMI2.0 or equivalent Support with KVM and media over LAN features including licenses, if any.
- The entire compute cluster hardware solution must be based on NSPOF configuration except for the interconnects.

II - Network connectivity:

Primary Interconnect

- The primary interconnect should be InfiniBand EDR or better. The proposed network should be part of the existing InfiniBand network of existing cluster. Bidders can visit IUCAA datacenter to understand existing network.
- Full hardware based Adaptive routing in the interconnect network should be provided. The cluster should be able to choose and alter the route path based on congestion.
- The system should support 100% non-blocking factor with a Fat Tree topology.
- The proposed interconnect technology should be capable of running RDMA operation in hardware and should not load any CPU core for performing RDMA packet process. CPU core are to be used for running application.

- The proposed high-performance network (host adapter as well as switching ASICs) must have hardware support for offloading low-latency MPI collective operations. This may be required to be demonstrated during acceptance tests.

Admin and Console Network

- Admin and Console Network should be on different physical switches and should not be clubbed in the Primary Interconnect Enclosure or Switch.
- The proposed admin and console network should be part of the existing admin and console network.
- All nodes need to be connected by Gigabit network for remote management hardware as well as for system network.
- Managed GigE L2 Switches should be offered with suitable number of cables for Admin and Console Network. Quantity of switches and cables and other required accessories to be provided as per proposed solution requirement.
- Advanced license for IPMI based management and monitoring of all nodes.

III - Cluster management Software

- Commercially available licensed and supported Cluster Management Software (CMS) should be offered for provisioning and managing all the compute nodes in the Cluster.
- CMS should integrate all nodes (number of nodes: 71) in the existing cluster, bidder should consider necessary licenses required for integration. Bidder can visit IUCAA data center to inspect existing cluster.
- CMS is required to manage the complete cluster. It should have the ability to verify and ensure consistency in hardware and system settings across the Cluster's resources from a single console. CMS must support the following:
  - It must be a GUI/Web-based tool accessible from any client.
  - It should provide a single interface for management and control of the complete cluster.
  - It should provide the facility to dynamically add, remove or configure any individual node.
  - It should provide for remote booting/resetting of individual or group of nodes.
  - It should provide for monitoring of vital parameters and provide predictive failure analysis and alarms.
  - The supplied Cluster Management Software Suite must be a commercially licensed product with latest version. All necessary software device drivers are to be provided.
- The tool should have support for GUI/CLI interface.
- Should support provisioning of CentOS and RHEL.
- The cluster management tool should support compute nodes from different OEMs.
- Following features should be supported by the CMS
  - Managing as many different images as needed for different software stacks, different operating systems, or different hardware.
  - Cloning from one to many nodes at a time with a scalable algorithm which is reliable and does not stop the entire cloning process if any nodes are broken.
  - Replicating available images on any number of compute nodes in the cluster.

- Customizing reconfiguration scripts associated with each image to execute specific tasks on compute nodes after cloning.

IV - Operating System:

- The solution should have the latest Version of licensed **RHEL** as an Operating System for the head nodes and equivalent CentOS for the rest of the nodes.

V - Workload Management Software:

- Perpetual & floating license with commercial support for all nodes with warranty upgrade.
- The existing cluster uses Altair PBS works suite as a workload management software
- Bidder should provide required licenses of Altair PBS Works Suite for workload management of all the Proposed Node
- The new nodes should be configured as a different queue and there should be another queue containing all the nodes (existing and proposed)
- OEM is responsible for setting up the policies, queues as per requirement given by IUCAA

VI - Software Stack: All the software quoted must be licensed and perpetual.

Part 3

I - Technical Benchmarks:

- The OEM should carry out below listed benchmark programs (**http://www.iucaa.in/tenders/2018/OCT/computingFacility/Benchmarks.zip**) on 8, 12, 16 nodes of the offered solution and submit the results achieved (with TFLOP count where applicable) in an output file along with the technical bid. While submitting the report, please make sure that all the timings, outputs and makefiles (wherever applicable) are submitted in a digital format. If required IUCAA can help in running benchmark provided remote access to the computing facility is provided.
  1. High Performance Linpack (HPL) (elapsed and CPU times and TFlop count achieved)
  2. GADGET-2: (Galaxies with dark matter and gas interact) benchmark (elapsed and CPU times)

  Note: Benchmark should be run with TURBO OFF and hyper threading disabled. Source codes provided by IUCAA for benchmark applications should not be modified by the vendor.

  In case the benchmarks are run with processor less than or more than 16 cores, the OEM should project the results for that specific processor and extrapolate/interpolate the same for 16 core processor.

The benchmark results should be submitted in following format.

Benchmark	Number of Nodes	Elapsed Time	CPU Time	TFlop
High Performance Linpack	8			
	12			
	16			
GADGET-2	8			NA
	12			NA
	16			NA

Details regarding the benchmarks:

(i) The Server OEM only should run the benchmark in their cluster. OEM is not allowed to run the benchmarks in some other customer premises or with the CPU manufacturer. The bidder/OEM should provide an undertaking certifying that they have done the benchmark themselves.
(ii) Log files of benchmark runs in electronic format must be submitted in the Portable drive before the tender submission date or the log should be uploaded on the bidders shared facility and the link has to be submitted along with the technical bid. If bidder fails to do so, the bid is liable for rejected.
(iii) Details of the Compute node configuration, Interconnects and Software environment used to run the benchmark is to be shared along with the Technical bid.
(iv) If required IUCAA may access the cluster to verify the setup and benchmark results.

II - Warranty:

All proposed cluster components both hardware & software should carry 6 years + 6 months on-site warranty with 24 X 7 operational support with 4 hours of response time, 24 hours of resolution time for any hardware related issue/problem and 72 hours of resolution time for any other issues. The warranty should also cover all the consumable spares including batteries.

III - Support and installation:

The OEM should do the entire installation & integration of the cluster at site.

- During the warranty period, OEM will have to undertake comprehensive maintenance of the entire hardware and its components.
- Quarterly review of HPC cluster health and its report submission.

- Half yearly review of various firmware related to hardware components which are part of HPC solution including third party.
- Proactively upgradation of the firmware of all the HPC component once in a year
- Upgrading firmware if necessary on the basis of criticality apart from the above schedule.
- HPC cluster implementation shall be monitored by a dedicated project manager for smooth implementation.
- Onsite support shall be provided during maintenance window such as DC shutdown, power outage and firmware upgrade.
- Installation/configuration and upgradation of HPC cluster activities should be carried out by direct OEM engineers only.
- In case of an issue, the OEM/Bidder engineer will be responsible for logging a case with OEM, collecting required logs and sharing the required information with OEM.

IV - Acceptance Test Procedure:

1. **Inventory check:** All the hardware and software should be checked against the Purchase Order.
2. **Functional Test:** All the functionalities of the proposed cluster should be tested including the connectivity, PFS storage system, entire suite of workload manager, cluster management tool etc.
3. **Performance Test:** HPL ratings (peak & sustained) for entire cluster configuration should be demonstrated after installation at site. Sustained HPL efficiency of the installed solution should be more than 60% of the offered theoretical peak performance. The application benchmarks should be executed on the installed solution and the demonstrated performance should match the benchmarks submitted with the bid to within 3% or be faster.
4. **Training:** The OEM should give 2 days System Administration training to a group of IUCAA personnel on installed hardware (Compute/storage/interconnect), operating system, installed system software and development tools including API. The training must be arranged at IUCAA.
5. **Documentation:** Documentations should be submitted for the following:

    - Procedure for bringing up and shutting down the fully integrated cluster.
    - Procedure for user Creation/Deletion/Modification
    - Procedure to get user accounting – Storage and Compute
    - Procedure for basic troubleshooting of Compute nodes, Storage & Head nodes (i.e the installed applications on the Head nodes).
    - Pictorial representation of the entire solution
    - Step by step installation guide for the entire HPC implementation/configuration from scratch.
    - Project documentation listing hardware/software with serial numbers, configuration and connectivity.
    - Any other document/manual useful for daily administration