



Procurement and Contracting Services

Request for Proposals L302501 – High Powered Compute

ADDENDUM #3

**Please mark all proposal submission
Envelopes with the following information**

**Sealed RFP # L302501
Due on March 20th, 2025 no later than 1:00 PM, MST**

RFP L302501 - High Powered Compute Questions and Answers

**The following questions were received prior to the close of the Technical
Question period on January 28th, 2025, at 1:00 PM MST:**

RFP Reference	Question	Answer
5.6 References:	We do not have the exact references required. Would you be willing to accept 3 references regardless of the size, industry, etc.? Is there anything else you would consider here in lieu of what is originally asked?	Yes
5.7 Background:	Would it be possible to know how many racks and servers are in place today?	Respondents are neither required nor expected to consider existing infrastructure when bidding.
5.9.4.2 Rack Design and Footprint	Would you be willing to accept a 1463 or 1552 mm deep and 2340 mm high rack if we could show in overall solution major benefits in performance, density, cooling, energy saving, etc.?	No
5.9.6.1 Liquid Cooling	It is called out that "Proposed solutions MUST NOT utilize immersion cooling." Iceotope is not immersion and we are not direct to chip. We are a true standalone technology that uses Precision Liquid Cooling to remove 100% of the heat generated by the server with absolutely no fans. We do use a Dielectric fluid in our solution. Would you be willing to consider a solution that uses a single phase environmentally safe, NO PFAS, safe to touch, Dielectric fluid?	No
5.10 Compute Requirements:	Additionally it would be helpful to know how much space would be provided for this solution. Can you clarify what the room looks like and how much room we would have?	Refer to Section 5.9 for room information. There are a maximum of seven rack spaces available to accommodate all equipment from 5.10 and 5.11 combined. For compute, three to four racks would be a conservative limit.

5.11 GPU Compute Nodes:	We understand that the solution you all are looking for is a dense solution. It would be helpful to understand how many racks/servers you are hoping for. Can you provide some insight as to how many racks/servers you are currently using and what you are hoping to see from a rack/server count from this RFP?	Dependent on pricing, budget is provided
5.11 GPU Compute Nodes:	Can you also provide insight into how many GPUs per server you are looking for?	Refer to RFP Section 5.11.4.1
5.11 GPU Compute Nodes:	Additionally it would be helpful to know how much space would be provided for this solution. Can you clarify what the room looks like and how much room we would have?	Refer to Section 5.9 for room information. There are a maximum of seven rack spaces available to accommodate all equipment from 5.10 and 5.11 combined. Three to four racks per section is a safe maximum.
5.11.1.4 GPU/Accelerators:	Would a PCIe alternative be acceptable or does it have to be the H200 NVL?	Refer to RFP Section 5.11.1.4. For clarity: NVL is PCIe, SXM is not allowed.
5.6 References	We don't have 3 references from comparable institutions that we can provide but we do have 3 references. Would this be a show stopper or would you be willing to accept 3 references regardless of the vertical?	Yes
5.9.6.2 Liquid Cooling	While we are not Immersion we are just interested to know why you all are not interested in liquid cooling this go around?	Question is out of scope for this RFP.
5.10.4.1	In the RFP you state, "For 21" OCP standard configurations: no less than 288 physical CPU cores per OCP OU." Can you please clarify if you are meaning that there needs to be 288 physical CPU cores per rack or per open unit in the rack? Meaning, if we have a solution with 12 servers, 3 U each for 36U total- are you asking for 10,368 CPU Cores in that rack?	Per Open Rack Unit. The example math provided is correct for proposals responding to section 5.10.
5.10.5.1	IN the RFP you state, "All large memory nodes MUST meet the following requirements: No less	Per rack unit of either standard, 1U server: 96c. 2U server, 192 cores,

	than 96 physical CPU cores per EIA RU/OCP OU" Can you please clarify if this is per server or per rack?	96 cores per rack unit or open rack unit.
5.11 GPU Compute Nodes:	Are you truly wanting separate units dedicated to Compute, GPU, Memory or would you be willing to look at solutions that can achieve density and offer you combined performance?	separate units
5.11.1.4 Node configurations MUST support either 2-way or 4-way NVLink bridge configuration	Is NVLink a deal breaker or would you be willing to evaluate solutions that do not have NVLink	MUST support either 2-way or 4-way bridging
5.8.1.3.3 Shipping costs MUST be included in proposal FOB to zip code 85721.	Can you clarify if you are meaning when the "ownership/liability" transfers or if you are truly meaning that this needs to ship via sea freight?	We mean ownership/liability transfers.
5.6 References:	What is the budget of the project?	Refer to RFP Section 5.10.6 for CPU proposals and 5.11.5 for GPU proposals
5.7 Background:	5.9.2.2 MUST specify the average power usage per rack (kVA) - how is the average to be calculated? - highly utilized HPC clusters average much higher power usage than the same systems if used to run enterprise workloads, for example	Add maximum kVA for all racks as configured in your proposal together and divide by the number of racks to produce an average power usage per rack
5.9.4.2 Rack Design and Footprint	5.10.7.9 Reported benchmark results will be validated during acceptance testing as described in section 5.10.9 - what, if any, variation is acceptable between reported and measured results? - processors with the same SKU from the same vendor historically display variances of up to 5% on HPL benchmark	Meet or exceed

5.9.6.1 Liquid Cooling	5.11.1.4 GPU/Accelerators All GPU/accelerators used in nodes for this section MUST meet the following requirements • Respondents MUST use NVIDIA H200 NVL (PCIe) GPU accelerators • Node configurations MUST support either 2-way or 4-way NVLink bridge configuration - does the University have a preference for the 4-way NVLink bridge configuration?	No
5.10 Compute Requirements:	5.11.1.7 what are we using for the backend storage system?	Respondents are neither required nor expected to consider existing infrastructure when bidding.
5.11 GPU Compute Nodes:	Is there a reason we are not looking at RoCE?	This question is out of scope for this RFP.
5.11 GPU Compute Nodes:	Are you open to looking at alternative storage solutions	No
5.11 GPU Compute Nodes:	5.9.7 delivery path. Can you walk us through the delivery path?	See Appendix B
5.11.1.4 GPU/Accelerators:	what is the total floor space	Refer to Section 5.9 for room information. There are a maximum of seven rack spaces available to accommodate all equipment from 5.10 and 5.11 combined. For compute, three to four racks would be a conservative limit.
	what is the hardest part of writing the RFP?	This question is out of scope for this RFP.
	why do you prefer PCIeexpress?	Serviceability
	Nvidia iv networking, switches have been discontinued how will you handle this?	See Addendum 2 - InfiniBand Requirements
	Can you provide us with a description of the site? For those who were not able to attend the site visit?	Refer to RFP Section 5.9
	can you give us a variation on the water temp	Refer to RFP Section 5.9.6
	What flow rate are we targeting for each row?	Not specified
	Does it have to be true AVX x 12 or can it be an alternative?	Must support AVX512 for compatibility. Native support preferred, dual-issue 256-bit acceptable.

	Who manages the raised floors and tiles located in the server room?	University of Arizona
	Can you give us an idea of the current Infrastructure densities as they relate to CPU, GPU, RAM and storage per server?	Respondents are neither required nor expected to consider existing infrastructure when bidding.
	Is this Tax Exempt?	TBD. Some or all of the initial purchase will be tax exempt, please include taxes on a separate line for our own reference.
	What are your Power limitations.	Refer to RFP Section 5.9.2
	is there a desire for the GPU and CPU Nodes to have the same cores?	Not required, proposals for each are separate
	5.8.15 Does the OS need to be installed prior to acceptance testing	Refer to RFP Section 5.10.9.1.6
	5.8.15 who installs the OS?	For acceptance testing, as stated in RFP Sections 5.10.9.1.6 and 5.11.8.1.6, respondent is responsible for providing acceptance testing plan, including benchmarks. As specified in section 5.10.7.2, benchmarks must be run under Rocky Linux 9. Respondent may install and configure Rocky Linux 9 for acceptance test benchmarking on specified nodes. Respondent will not be involved in production operating environment.
	What is the temperature we are targeting for the room?	68F is the room ambient set point for our chillers.
	Is there a preference for active cooling to ensure a specific offset temperature? Are there temperature set points you can provide that can be used for design parameters?	68F is the room ambient set point for our chillers. Existing MotivAir doors were not able to achieve heat neutrality given the input factors, so they are set to aim for an air output of 74F, which has been sufficient to keep the heat load within the capacity of our room chillers to extract. If your design features a rack-level cooling solution that can be reasonably expected to generate 74F output air given 68F input air, it should work well in our environment. If your solution can achieve heat neutrality, that is an efficiency advantage that will be considered as

		part of the environmental criteria, but not a requirement.
	Benchmarking question - Do we need to run both variants of HPL with and without netlib.	Whatever version you are using to generate the estimate in your proposal is the one you need to use when testing onsite.
	Do any specific GPU tests need to be conducted?	Refer to RFP Section 5.11.6
Pricing	It was stated that the budget was \$2M for GPU nodes, \$2M for CPU nodes, and \$1M for Infrastructure. Is that the hard limit per area, or is it a \$5M budget with flexibility between areas?	Hard limit for proposals. There will be post-purchase expansion, but no committed expenditure amount.
Pricing	Does the cost include PDUs and cables, too?	Refer to RFP Section 5.10.6 and 5.11.5
Pricing	What proportion of standard compute & Large memory nodes will be required for the \$2M budget?	Dependent on pricing, majority standard compute
Pricing	Why try to use HDR when NDR is widely available and well-priced?	We do not require the performance of NDR, HDR is adequate to our needs. Availability and pricing of NDR was not a factor.
Pricing	Should proposals include licensing costs for vendor-provided software, or is it expected that the University will procure software separately?	Must include all required licensing see RFP SECTION 5.10.6.5 and 5.11.5.4
Scope Clarifications	What does RDC stand for? Research Data Center?	Research Data Center, yes.
Scope Clarifications	What are the restrictions, if any, on proposing GPU and CPU nodes within the duplicate submission?	Submissions for CPU and GPU are separate and do not depend on each other or include any duplication.
Scope Clarifications	Can optional features or capabilities outside the listed specifications (e.g., experimental cooling solutions)	Refer to RFP Sections 5.10.12 and 5.11.11 as well as 3.9.8, criteria are in order of importance. No further scoring information will be provided.

	be included, and how will they be evaluated?	
Scope Clarifications	Connections using 100G cables for all 100GbE ports for customer use: will the proposal be out of scope?	Responses MUST comply with fabric requirements in RFP sections 5.10.3.1 and 5.11.3.1 or they will be subject to disqualification.
Scope Clarifications	Confirming if there have been any Addendums post-site walk-through. Are there any pictures from the site walk-through to share with Dell?	No addendums, no photos of site
Floor, Space & Density	Is there a detailed room drawing available showing the actual dimensions?	No
Floor, Space & Density	What are the required maintenance distances between isles, end racks, and walls?	Sufficient to open door fully for maintenance and remove any racked component, roughly two floor tiles for aisles. End of row to wall will run to concrete skirt.
Floor, Space & Density	What is the weightage for criteria 5.9.5 in comparison to 5.10.12 and 3.9.8?	Criteria are in order of importance, weighting will not be provided.
Power Clarifications	How much total power capacity is available? What is the maximum power per rack available? What is the power factor?	Total solution metrics are not provided. Refer to RFP Section 5.9.2.1 for per-rack. Power factor for facility is 0.9.
Power Clarifications	Is the power available for use in this solution enough for all 7 racks to be simultaneously running at 68kVA?	No
Power Clarifications	Are there specific power redundancy requirements beyond UPS and generator backup (e.g., A/B feed redundancy per rack)?	No
Power Clarifications	Are there limitations on the number of power connections per rack? How many racks can the 60A 208V supply? And how many racks can have at 30A?	Refer to RFP Section 5.9.2.4. There are a maximum of seven rack spaces available to accommodate all equipment acquired via 5.10 and 5.11 combined. The RFP neither asserts nor implies a maximum number of racks at any specific power specification.
Power Clarifications	Can vendors propose alternative plug types or power configurations, or must they strictly adhere to the listed specifications?	Examples provided in RFP section 5.9.2 are currently in use and will be easiest to support. Vendors MAY propose alternatives if they believe them to provide notable benefit.

Power Clarifications	Can vendors propose alternative power delivery solutions (e.g., higher-voltage power feeds) to improve rack density or efficiency?	No
Power Clarifications	What voltage/ampereage is available for power distribution from UPS to rack PDUs?	Refer to RFP Section 5.9.2
Power Clarifications	What are the cable lengths provided/provisioned from the DC power whips to PDU connectors?	As stated in RFP Section 5.9.2, power is provided via overhead bus bars which feature modular connection points. PDUs should provide for no more than 2 meters of reach from the top of the rack. Final specifications will be determined based on parameters of winning response.
Power Clarifications	At full start-up or restart, can the data center handle the power harmonic distortion?	Yes
Cooling Infrastructure	Given the mention of both options, is there a preferred approach between rear-door heat exchangers and direct liquid cooling?	No
Cooling Infrastructure	How critical is the requirement for front-to-back airflow? Are exceptions permitted if cooling solutions meet the overall heat dissipation requirements?	Refer to RFP Section 5.9.6.5 and definition of MUST provided in section 5.8.
Cooling Infrastructure	To ensure compatibility with PDUs and water-cooling options, what solutions exist for environmental management and data center monitoring?	There is no expectation or requirement for respondents to consider interoperability with existing PDUs and water-cooling options.
Cooling Infrastructure	What are the facilities' water inlet/outlet temperature ranges and flow rates?	Measured 41-50F over seven years, average 47F as stated in 5.9.6.1. No flow rate monitoring to data center specifically.
Cooling Infrastructure	What is the preferred and maximum heat containment/cooling requirement?	Rack output temperature should target 74F
Cooling Infrastructure	What is the operational cooling technology from inside air handling and outside heat rejection?	Chilled water only in research data center
Cooling Infrastructure	Cooling vendors will require the customer to fill out questionnaires. Please clarify	We cannot do questionnaires for each bidder, all questions must be

	how the submitter and their partners will handle or post additional requirements.	asked and answered publicly via this process.
Waste Heat Management	How much waste heat maximum in total solution can the data center handle in BTU or kW?	Rack output temperature should target 74F, no other information available
Waste Heat Management	What is the maximum allowable waste heat per rack in BTU or kW?	Rack output temperature should target 74F, no other information available
Waste Heat Management	Is there an efficiency metric (e.g., PUE) that the University prioritizes for cooling solutions?	no
Node & Rack Requirements	For Appendix B, what are the dimension restrictions (if any)?	As stated in section 5.9, a confidential document describing the data center, loading dock, and delivery paths can be obtained from the RFP Purchasing Agent Zane Forier, zfforier@arizona.edu .
CPU Node Configurations	What is the desired ratio of standard and large memory nodes for the CPU-only nodes?	Proposals responding to section 5.10 should include only standard nodes in the proposed configuration, but MUST include pricing for large memory nodes. Quantity of large memory nodes for initial purchase will be determined following contract award.
CPU Node Configurations	Do you have a preferred ratio of nodes: CPU vs. MEM vs. GPU?	CPU solution proposals should include only CPU nodes as well as pricing for large memory nodes. GPU solution proposals should include only GPU nodes.
CPU Node Configurations	For section 5.10.5.1, what is the mix of Intel and AMD nodes between the standard CPU and large memory nodes?	There is no mix required. As stated in 5.10.1.3, may use either Intel or AMD. Compute nodes should all be either Intel or AMD. Large memory nodes should all be either Intel or AMD. Compute nodes and large memory nodes do not have to use the same vendor.
CPU Node Configurations	What is the desired ratio of standard nodes to GPU-only nodes?	These are separate bidding sections
CPU Node Configurations	What is the desired ratio of GPU-only nodes to large memory?	No large memory nodes in GPU proposals
CPU Node Configurations	Is there a preferred target for HPL RMax performance for	No

	CPU-only nodes for standard and large memory nodes?	
CPU Node Configurations	Are vendors allowed to propose configurations that deviate from the memory-per-core ratio if overall performance metrics are met?	MUST meet or exceed requirements as stated
CPU Node Configurations	Is there a minimum requirement for single-thread performance, or is the all-core base/boost frequency sufficient?	All-core is sufficient
CPU Node Configurations	For section 5.11.1.3, the 4th bullet point needs clarification if this means a minimum of 64 cores per node if the node has 4 GPUs installed.	That is what it means, no further clarification to provide.
GPU Node Requirements	Can vendors propose accelerators other than NVIDIA H200 NVL GPUs, such as comparable AMD GPUs?	No
GPU Node Requirements	Are GPU-to-CPU ratios flexible based on workload performance data or customer preferences?	Refer to RFP Section 5.11.1.3 and definition of MUST as described in section 5.8.
GPU Node Requirements	Must NVLink bridges be provided if the GPUs can perform similarly without them?	MUST support, not REQUIRED to include with proposal
GPU Node Requirements	Can the GPU nodes have 8-way H200 on a single node?	Yes, as long as all requirements in section 5.11.1 are met.
GPU Node Requirements	What proportion of standard CPU and large memory nodes will be required for the \$2M budget? What specific number of GPU H200NVL quantities does the customer need?	Proposals should include only standard nodes along with pricing for large memory nodes, GPU node count should exhaust \$2m stated budget
GPU Node Requirements	For section 5.11.1.4, if 2-way NVLink is proposed, will there be a later situation, or will we need to add more GPUs? What are the requirements or needs for such expansion?	Either 2-way or 4-way must be supported, moving from 2-way to 4-way is not required
GPU Node Requirements	The H200 SXM GPU performance is higher than that of the current PCIe version of specific servers. Would there be an option to consider 4-way H200s?	No
Rack Specifications	· Can vendors propose custom or modular racks if they	Yes

	meet the specified footprint and weight requirements?	
Rack Specifications	· Is there a preference for standardized vs. vendor-specific racks (e.g., Open19 vs. proprietary designs)?	No
Rack Specifications	· The maximum rack width is 800 mm. Can slightly wider racks (e.g., 820 mm) be proposed if they improve compute density or cooling efficiency?	No
Rack Specifications	· What is the max power per rack available? What is the power factor	Refer to RFP Section 5.9.2.1 for per-rack. Power factor for facility is 0.9.
Rack Specifications	· How many racks can the 60A 208V supply? And how many racks can have the 30A?	Refer to RFP Section 5.9.2.4, no further information to provide.
Rack Specifications	· Is there a maximum heat containment/ cooling requirement?	Rack output temperature should target 74F
Memory	· The specification requires at least one DIMM per memory channel. Can vendors propose systems with higher DIMM densities if they exceed the memory-per-core requirement (e.g., for cost optimization)?	Yes
Memory	· Is there a preference for DDR5 memory, or is DDR4 acceptable if performance metrics are met?	DDR5 preferred for platforms supporting DDR5. DDR4 acceptable for platforms that do not support DDR5.
Memory	· Is the 256 GB of system memory per GPU a hard requirement, or can configurations with slightly lower memory be considered if performance metrics are met?	Refer to RFP Section 5.11.1.5 and definition of MUST as described in section 5.8.
Memory	· Will a mix of Intel and AMD nodes work between the std and large memory nodes?	CPU nodes and large memory nodes may use different CPU OEMs.
Storage	· Does the requirement for a 2TiB NVMe-based disk include redundancy or RAID configurations, or is this purely local storage per node?	Single drive local storage per node
Storage	· Are specific NVMe drive manufacturers or models preferred for reliability or performance?	No

Benchmarking	Are there additional benchmark requirements beyond HPL (e.g., AI/ML-specific benchmarks like MLPerf)?	No
Benchmarking	How should benchmarks be run for all configurations or only the primary solution if multiple configurations are proposed?	Provide benchmark value for standard node for CPU bids (5.10.7), GPU node for GPU bids (5.11.6).
Networking and Fabric Infrastructure	Can vendors propose different network configurations, such as combining or splitting 200 Gbps capabilities in fewer NICs/ports?	Not for ethernet, see addendum 2 for InfiniBand requirements.
Networking and Fabric Infrastructure	Regarding the NIC adaptor, if a maximum of 200Gbps is provided, would the customer consider NDR or NDR200 a better preference? Please detail the requirements.	See Addendum 2 - InfiniBand Requirements
Networking and Fabric Infrastructure	Can the switches be efficiently placed in the racks between CPU and GPU nodes? What restrictions are imposed on TOR and other switch placements?	Refer to RFP Sections 5.10.3.3 and 5.11.3.3. Final rack layouts will be negotiated during best and final offer process.
Networking and Fabric Infrastructure	What are the inter-rack cabling requirements for the entire solution?	No inter-rack cabling
InfiniBand	Is UFM required?	UFM Telemetry at minimum. Proposals may include UFM Enterprise or Cyber-AI if they deem it necessary and appropriate.
InfiniBand	What is the specific InfiniBand topology? Can we assume 1:1 Fat Tree?	InfiniBand is isolated per rack, no larger topology
InfiniBand	Is a non-blocking network be required if additional HDR switches are needed in a leaf-spine architecture?	No spine switches, InfiniBand is isolated per rack, no larger topology
Ethernet	Can vendors propose Ethernet NIC configurations with different port counts or speeds if they perform similarly?	No
Ethernet	Will redundant fabric switches (e.g., for failover) be evaluated as a value-add, or are they mandatory?	Redundant switching not required

Software and Integration	Do we need to provide any software? Unified Fabric Manager (UFM)?	UFM Telemetry at minimum. Proposals may include UFM Enterprise or Cyber-AI if they deem it necessary and appropriate.
Scheduler and Orchestration	What is the preferred software for cluster management?	Respondents are neither required nor expected to consider cluster management software when bidding.
Scheduler and Orchestration	What Scheduler software do you prefer?	Respondents are neither required nor expected to consider scheduler software when bidding.
Scheduler and Orchestration	Is the University open to alternative schedulers beyond those listed (e.g., Kubernetes-based workload orchestration)?	Respondents are neither required nor expected to consider scheduler software when bidding.
Monitoring and Reporting	Are open-source management and monitoring tools (e.g., Prometheus, Grafana) acceptable, or are commercial solutions preferred?	Either is fine to optionally include.
Monitoring and Reporting	Are there specific requirements for monitoring tools? Should they include real-time power usage, thermals, and node status?	No requirements for monitoring tools
Monitoring and Reporting	Must the monitoring tools integrate with existing University systems? If so, what are the APIs or formats?	No requirements for monitoring tools
Other Software	What additional commercial and open-source software is required at check-out? Acceptance testing?	No additional software is required by The University of Arizona. During acceptance testing vendors may install and configure Rocky Linux 9 to complete the benchmarking requirements, including any additional software they require to complete benchmarking.
Other Software	Should the proposal include licenses for vendor-provided software, such as NVIDIA UFM or InfiniBand management tools, or will the University handle procurement?	Yes, see sections 5.10.6.5 and 5.11.5.4
Other Software	Are there specific service-level agreements (SLAs) expected for technical support, such as response times for software issues?	Refer to RFP Section 5.8.1.4, 5.10.10.1.1 and 5.11.9.1.1

Integration and Documentation	What level of integration with the University's existing batch schedulers, data transfer tools, or storage systems is required?	For systems under acceptance testing, there will be no integration of any kind with existing system infrastructure. All construction and configuration of operating environments for production use will be exclusively done by University of Arizona staff. Respondents are not required or expected to plan or execute any integrations of these types.
Integration and Documentation	Is there a preferred documentation format or tooling for system diagrams, labeling, and procedural guides?	No
Benchmark Validation	Will benchmark results from the proposal be re-validated for each node type during acceptance testing or only for a subset?	Only CPU nodes will be benchmarked for CPU proposals. Only GPU nodes will be benchmarked for GPU proposals. See sections 5.10.9 and 5.11.8
Benchmark Validation	What happens if the proposed system's benchmark performance deviates slightly during acceptance testing?	See RFP section 5.10.9.3, 5.10.9.4, 5.11.8.3, 5.11.8.4
Uptime Acceptance Testing	How will failures due to external factors (e.g., facility power outages) be accounted for during the 28-day uptime test?	Outages attributable to The University of Arizona will not count toward calculations in section 5.10.9.2 and 5.11.8.2
Uptime Acceptance Testing	Are there penalties for failing to meet the 99.5% uptime target within the initial 35-day remediation period?	Refer to RFP Sections 5.10.9.4 and 5.11.8.4
Future Scalability and Expansion	Do you have future expansion plans? When? How extensive?	To be determined. This RFP does not represent or guarantee any planned expansion, timeline for such expansions, or prospective scale for such expansions. For our past two major system purchases, system CPU core count capacity expanded approximately 50% over their service life. This does not constitute a promise or guarantee of similar expansion of systems acquired via this RFP.
Future Scalability and Expansion	Are vendors required to design for scalability beyond the initial deployment, and if so, what are the anticipated	all racks for a proposed section should be capable of equal loading, no dependencies between racks

	expansion metrics (e.g., node count, rack count)?	
Future Scalability and Expansion	Will separate future budgets be allocated for expansion, or should this be included in long-term planning?	To be determined. This RFP does not represent or guarantee any planned expansion, timeline for such expansions, or prospective scale for such expansions. For our past two major system purchases, system CPU core count capacity expanded approximately 50% over their service life. This does not constitute a promise or guarantee of similar expansion of systems acquired via this RFP.
Future Scalability and Expansion	· If expansion is required, will the University expect the vendor to guarantee component compatibility for a specific period (e.g., 3-5 years)?	MUST meet hardware maintenance terms, not required to guarantee future availability for purchase
Future Scalability and Expansion	· If 2-way NVLink is designed in the proposal, will there be situation at a later date to add more GPU's ?	MUST support either 2-way or 4-way bridging, moving from 2-way to 4-way is not required
Criteria Weighting	· How does the university weigh cost, performance, flexibility, energy efficiency, etc.?	Weights are not provided, criteria are in order of importance.
Criteria Weighting	Are the weights equal, or are there specific priorities based on the nature of the nodes (e.g., CPU vs. GPU)?	Weights are not provided, criteria are in order of importance.
Criteria Weighting	· How will optional features and capabilities (e.g., enhanced energy efficiency or futureproofing) be scored?	Weights are not provided, criteria are in order of importance.
Environmental Impact	· Will additional credit be given for proposals that exceed energy efficiency expectations, such as PUE or carbon footprint reductions?	No
Regulatory Compliance	Will additional certifications (e.g., ISO, FIPS, ITAR) be required for the proposed solution, or are the listed compliance requirements exhaustive?	No additional compliance requirements
VPAT Requirements	· Are there specific accessibility requirements for any user-facing interfaces or	No

	tools (e.g., those used for management or visualization)?	
VPAT Requirements	Should we expect to provide new VPAT(s) for the proposed system(s)?	Refer to RFP Section 4.3.1
General Questions	If, following clarifications submission & UA response, more technical questions arise, will we be able to ask them?	Yes. However if the committee does not feel that the question warrants an exception to the official schedule, then no answer will be provided.
5.9	Understanding that rack-external interconnects will be ran by the customer, should we plan for these connections to run above the racks or below the floor?	Above the racks, no cabling below the floor
5.9.2 referencing Power The RDC is designed for high density racks with overhead power distribution for 208V single or 3-phase power connections. Backup power is provided by both UPS and generator.	Just want to confirm whether we should include any redundant circuits in the racks or not, considering the power is backed by UPS and generator. I.e. If the rack design requires one full PDU to operate, should we include an additional PDU in a 1+1 configuration to allow operation in the event that one feed goes down? Are all available feeds backed by UPS and generator or only a subset?	No redundant power
Referencing 5.10.3.1 & 5.11.3.1 - InfiniBand Fabric 1x Mellanox QM8700 InfiniBand Switch ○ Airflow direction SHOULD match hot-air exhaust orientation of proposed server systems ○ Proposals SHOULD include a second identical switch if the node count in the rack exceeds 80 nodes	I do not see any provision for uplinks to a spine layer. Should we reserve any number of ports on these switches?	No

5.10 & 5.11	Memory and storage capacity is mentioned throughout using Gibibytes (GiB) and Tebibytes (TiB), whereas most memory and storage products on the market are spec'd in terms of Gigabytes (GB) and Terabytes (TB), leading to a slight discrepancy. GiB = 2^{30} bytes GB = 10^9 bytes TiB = 2^{40} bytes TB = 10^{12} bytes Is it acceptable to replace the GiB or TiB requirements with the closest equivalent rating in GB or TB, or should we go with a higher capacity model that when converted exceeds the rating in GiB or TiB? I.e. A drive marked as 2 TB provides about 1.82 TiB of capacity, so we would have to move up to 3.2 TB or 4 TB in order to exceed 2 TiB.	For RFP sections 5.10.1.6 and 5.11.1.7 $\pm 10\%$ is meant to account for this difference. The same variance may be applied to sections 5.10.1.4 and 5.11.1.5 regarding memory.
5.10 Compute Nodes	Is there a desired ratio of "Standard" compute nodes to "Large Memory" nodes? Is it desirable to have these node types separated in their own racks, or is it acceptable to mix them in racks?	Proposals should include only standard nodes along with pricing for large memory nodes, may be mixed
5.11.1.4 - Referencing Node configurations MUST support either 2-way or 4-way NVLink bridge configuration	Should NVLink bridges be included for the GPUs, or merely supported by the configuration?	Supported

5.11.6.1 - Referencing - Respondents are REQUIRED to provide actual or estimated results for the High-Performance LINPACK benchmark using the GPU accelerators in the node. Rpeak (theoretical) and Rmax (estimated or actual) values MUST both be provided.	Are these benchmark numbers to apply to a single GPU, or collectively across all GPUs in the node?	Provided benchmark numbers should represent one GPU, but all GPUs in each node under acceptance testing MUST be tested individually to demonstrate they all meet or exceed the provided performance values.
Integration and Configuration Specifics	Are there specific standards or guidelines (e.g., labeling conventions, cabling specifications) for the pre-cabling, labeling, and integration tasks?	No
Integration and Configuration Specifics	Should specific firmware, OS, or configurations be pre-installed on the servers before delivery?	No
Integration and Configuration Specifics	Does the University require redundancy testing or failover setup during network configuration?	No
Acceptance Testing and Approval	What benchmarks or key performance indicators will be used for acceptance testing, particularly for network stress testing and redundancy mechanisms?	No vendor acceptance requirements for anything but compute/gpu benchmarks
Acceptance Testing and Approval	Who will be responsible for approving the completed installation and configuration—IT staff, a specific department, or a third-party inspector?	IT staff
Post-Deployment Support	What are the expected SLA (Service Level Agreement) requirements for post-deployment support, including response and resolution times?	Refer to RFP Section 5.8.1.4, 5.10.10.1.1 and 5.11.9.1.1
Post-Deployment Support	Are there specific software tools or platforms preferred for	No

	ticketing and tracking support issues?	
Training and Documentation	Are there preferred formats or templates for network diagrams, user manuals, or troubleshooting guides?	No
Training and Documentation	For the planned workshops and training, are there specific learning objectives or content preferences the University would like covered?	No
Community Engagement Requirements	Beyond the proposed internships and workshops, are there additional community engagement initiatives the University would like bidders to include?	No
Timeline and Site Logistics	Will the University provide access to facilities for conducting site surveys and installation during standard business hours only, or is after-hours access available?	Escorted access only, negotiable working schedule
Timeline and Site Logistics	Is there a preferred timeline for conducting the installation phases (e.g., specific milestones)?	No.
Compliance and Security	Are there specific compliance standards (e.g., PCI, HIPAA, FERPA) or cybersecurity measures that must be followed during integration?	Visitors must be escorted while in the data center, otherwise no.
Compliance and Security	Are background checks required for personnel involved in installation and configuration?	Background checks are not required but all personnel must be escorted by UA staff while in the data center.
Risk Management	Does the University require a formal risk management plan or disaster recovery plan as part of the proposal?	No
Risk Management	Are there any known risks or constraints regarding facility power, cooling, or network capacity that we should consider during planning?	No
Procurement Flexibility	If any equipment or components become unavailable, does the University permit substitutions with equivalent or better specifications?	Yes

Contingency and Change Management	How does the University prefer to handle change requests that might arise during deployment, such as scope modifications or equipment substitutions?	As deemed most appropriate by the University for the type of change or substitution in question.
	Do you have a targeted "Standard Compute Node" to "Large Memory Node" ratio?	Proposals should include only standard nodes along with pricing for large memory nodes
	In the standard compute node, Is there a target mem/core?	Refer to RFP Section 5.10.1.4
	Can you provide a hybrid liquid cooled solution, i.e. GPUs with direct and CPUs with Rear door cooling units?	Yes, can be different for separate proposal sections, do not mix within proposal sections
	Who will do the final connections to facility liquid?	Installer
	Data center Ambient Air Temperature and Humidity?	Average room temperature is ~74F, average relative humidity is ~25%.
	Facility water flow rate (L/min)?	No flow rate tracking limited to data center
	Size of filtration planned for primary side piping [micron/ mesh]?	Water provided by campus-scale service, no specific filtration at data center
The Successful Vendor shall provide and pay for performance and/or payment bonds. Such bonds shall cover the faithful performance (100%) of the Agreement and the payment of all obligations (100%) arising thereunder, in such form as the University may prescribe and with approved sureties.	Request you to kindly change performance and/ or payment bond (10%) of the contract value.	No

Payments by the University shall be subject to the provision of Title 35 of Arizona Revised Statutes relating to time and manner of submission of claims. The University's obligation is payable only and solely from funds appropriated for the purpose of the Agreement. Unless otherwise stated herein, the payment terms for the Agreement are Net 30 days	we understand that 100% payment shall be made to the bidder within 30 days of invoice acceptance. Kindly confirm.	Refer to RFP Section 4.40
General	Request you to kindly define timelines for each phase of the project, bidder is required to take into consideration	Timelines for project implementation vary based on multiple factors. Once a solution has been identified by the committee, timelines will be solidified based on the logistical requirements associated with the winning proposal.
References SHOULD meet the following criteria: • References SHOULD be from comparable United States-based educational or research institutions	Can references be provided for companies with a global experience- Government entities/ Metros/ Smart cities	Yes.
General	Can a bidder bid as a consortium of 2 or more entities.	Yes. However, one vendor must be identified as a primary respondent. We will use the primary respondent for communications and payments.
Limitation of Liability	Limitation of Liability - Kindly confirm that the limitation of liability for the project to the bidder is maximum of 100 % of the contract value.	Vendors limitation of liability will be determined and negotiated at time of contract, but at a minimum will be 100% of the contract value. Increased limitations may be imposed depending on the risk and scope of the final contract.

Agreement Term	Kindly confirm the duration/ Timelines during which the bidder is required to complete the scope of work.	Timelines for project implementation vary based on multiple factors. Once a solution has been identified by the committee, timelines will be solidified based on the logistical requirements associated with the winning proposal.
The parties agree to arbitrate disputes filed in Arizona Superior Court that are subject to mandatory arbitration pursuant to ARS § 12-133.	Request you to kindly change the dispute resolution clause as provided below: "Any dispute or difference whatsoever arising between parties out of or relating to the construction, meaning or operation or effect of this Agreement shall, unless resolved amicably, be settled by Arbitration, conducted in English Language, under the Rules of International Chamber of Commerce. The Arbitration shall be held at New York."	No
5.8.2.1 Respondents MAY provide optional libraries, utilities, applications, etc.	Kindly provide exact scope of these optional libraries, utilities, applications, etc. to be factored by Respondent	If required by respondent for operation of cluster or considered by respondent to be advantageous to The University, such things MAY be provided. Optional, not required, can be ignored.
5.8.2.2 Respondents MAY provide options for any additional hardware or software components that the respondent feels will extend the capacity and/or capabilities of the system to provide enhanced resources and support for UA research computing	Please clarify the scope of additional hardware or software to be considered by Respondent with the purpose.	Unspecified, vendor discretion to provide things you think we should consider in addition to meeting all requirements for proposals. Optional, can be ignored.

5.9.2.1 Respondents MUST specify the maximum power usage for a single compute rack. Maximum power usage per rack MUST NOT exceed 68kVA	Kindly confirm on the max power density available per Rack is 68 KVA as per the current power distribution to the Datacenter Racks. This will support to propose the design the maximum rack load capacity per Rack. Also please provide the total available power for the Datacenter for the designated space for HPC benchmarking and total number of racks required to be hosted.	Refer to Section 5.9.2 for power information. There are a maximum of seven rack spaces available to accommodate all equipment procured via this RFP for both sections 5.10 and 5.11 combined. Three to four racks is a reasonable maximum for proposals to either section.
5.9.2.3 MUST specify the power plug types for rack PDUs. For example, our facility currently provides 460R9W/460C9W for 3-phase 208V 60A and L6-30R for single-phase 208V 30A.	Please confirm the 3-phase or 1-phase power sockets are supported with AC (OR) DC power supply on Rack PDUs.	We do not provide DC power, AC power only
5.10.7.1 Respondents are REQUIRED to provide actual or estimated results for the High-Performance LINPACK benchmark. Rpeak (theoretical) and Rmax (estimated or actual) values MUST both be provided.	Kindly share the detailed LINPACK use cases with expected benchmark parameter wise like Compute Cores, memory, applications to be tested on hardware, throughput for network, IOPS for any SAN network, its performance limits to design & meet the expected level of performance for benchmarking High performance LINPACK.	Refer to RFP Sections 5.10.7 and 5.11.6 for benchmarking requirements, specifically 5.10.7.4 and 5.11.6.4.
5.10.7.2 Benchmarks MUST be run under Rocky Linux 9	We understand Rocky Linux 9 will be operating system used for conducting the LINPACK testing. Please let us know who will have to own the licenses of Operating System and OEM support Cost involved during LINPACK testing.	Rocky Linux 9 is freely available and requires no licenses. High-Performance LINPACK is freely available and requires no licenses. If respondents require any licensed software to complete acceptance test benchmarking, it MUST be provided by the vendor for use by the vendor during acceptance test benchmarking. If licensed software is required to boot a freely available operating system and/or run freely available applications on solution

		provided by respondent, that software pricing MUST be provided as specified in section 5.10.6.5 and 5.11.5.4.
General	Floor Plan of the Layout of existing Datacenter with dimensions and its Type of Racks placement will support for planning our rack spread in the given floor area of DC. Any preference of OEM makes for compute, other components, etc. will certainly limit the respondent to align with OEM for partnership in longer term to support us as and when required during benchmarking.	Refer to RFP Section 5.9 for room information and sections 5.10.2 and 5.11.2 for additional rack infrastructure requirements.
General	If the budget is \$5M and contains the three categories what is the format on how the network and supporting infrastructure is to be priced? The CPU/GPU nodes have their own Appendix C and D for pricing. Appendix C and D also request pricing for switching and supporting infrastructure. Should there be an Appendix E for supporting infrastructure pricing separation?	All components required by the proposal should be priced into the proposal, including fabric switching with required accessories. Respondents should not infer to provide any switching not specified in the sections including but not limited to any "spine" switching for ethernet or InfiniBand or any type of aggregation switching.
General	Can you confirm the Ethernet Fabric switch called out in section 5.11.3.1 should be the Mellanox SN3200C as stated (this is also referenced under section 5.10.3.1)? There does not appear to be a current NVIDIA switch with that model number. Is it possible this specification should read either SN3700C or SN3420, which are currently available Ethernet switches in the NVIDIA 3000 family?	This is an error. For sections 5.10.3.1 and 5.11.3.1, the correct model number is the Mellanox SN3700C.

General	In that same area (communications fabrics), since NVIDIA has acquired Mellanox all switches called out in the RFP are now NVIDIA-branded. Is this substitution in our response acceptable to the University?	Yes, NVIDIA re-branded switches that match the previous Mellanox model numbers are acceptable.
General	Can you please confirm the project budget? During the pre-proposal site visit it was noted that the budget was \$5M, broken down as: - o \$2M or 40% on CPU nodes - o \$2M or 40% on GPU nodes - o \$1M or 20% on switching and other supporting infrastructure	Confirmed

End of addendum, all else remains the same.