

FAQs – Innovative Technology Insertion of Advanced HPC Systems

Last updated September 29, 2026

Are software-based solutions within the scope of this effort? While the opportunity appears focused on infrastructure and computing platform capabilities, would solutions that improve system performance, operational efficiency, resource utilization, reliability, automation, or the overall user experience also be considered? Specifically, are technologies that help optimize the use of existing resources, reduce administrative burden, minimize downtime, or enhance consistency across enterprise environments within scope?

No. Software-based solutions are not within scope of this effort which is based on the acquisition of HPC systems and follow-on system administration and maintenance support options. Software-based solutions that improve existing system performance, operational efficiency, resource utilization, reliability, automation, or the overall user experience will not be considered. Technologies that help optimize the use of existing resources, reduce administrative burden, minimize downtime, or enhance consistency across enterprise environments are not considered to be within scope since this effort is based on the acquisition of new HPC systems.

Could any of the acceptance testing be completed at the MFG Factory prior to shipment?

Yes. Portions of the acceptance testing may be completed in the factory prior to shipment.

Is there a general criteria on acceptance testing that can be shared?

It is expected that the government and offerors would collaborate on the final acceptance testing criteria.

For most TI acceptance processes, there are typically two phases – Capability Test (CT) and Effectiveness Level Testing (ELT).

During CT, each production system would be subject to a set of functional and capability tests. The functional set usually includes basic system hardware and software assessments to determine whether the system meets the proposed configuration. The capability set is generally a suite of application and synthetic (I/O, networking) benchmarks intended to represent the workloads anticipated on the system. CT is typically expected to be completed within 5 business days of starting.

Upon completing CT, the system enters ELT, which is generally a 60-day window in which 30 consecutive days of system metrics must be met. These metrics are usually system utilization ($\geq 75\%$), system availability ($\geq 97\%$), system interrupts ($\leq$ minimum # allowed), and node interrupts ($\leq$ minimum # allowed). If a system is unable to satisfy all of the metrics over the 30 day period, the window to complete the metrics will slide until 30 consecutive days are met. For

example, if a system exceeds its system interrupt count on day 29 of ELT, and the first system interrupt occurred on day two, ELT would need to continue for two additional days to surpass the initial system interrupt, and then one more day to reach 30 consecutive days.

What quantitative performance objectives and acceptance benchmarks apply to each DSRC system and scenario?

The government intends to collaborate with the selected offeror on performance objectives and acceptance benchmarks.

Please identify the relevant measures for CPU HPC, AI/ML, storage, and fabric performance, as well as any pass/fail thresholds for the required multi-week burn-in period.

Generally speaking, during Effectiveness Level Testing (ELT, i.e. multi-week burn-in period), the system must meet the metrics outlined in the previous response. Offerors may utilize the prescribed benchmark suite - again, developed in collaboration with the selected offeror - to both meet the required metrics while observing performance levels for issues.

What workload mix and job profiles are expected at AFRL, Navy, and MHPCC?

In general, HPCMP systems will support a wide variety of modeling and simulation workloads. For more information on these workloads, please see: <https://hpc.mil/solution-areas/technology-areas>. Consideration should be taken that some modeling and simulation applications are able to leverage accelerators configured to support double precision calculations.

Regarding AI/ML, a sample listing of the domains would include autonomous systems and robotics; predictive maintenance; situational awareness and decision support; cybersecurity and threat detection; training and simulation; intelligence, surveillance, and reconnaissance (ISR); logistics and supply chain optimization; target recognition and precision engagement; human-machine teaming; and weather modeling and forecasting. Within these domains, specific AI/ML workloads might include synthetic data generation, sparse inverse problems, open-ended design optimization, geometric deep learning, science-constrained deep learning, digital twins, uncertainty estimation, agentic systems, GPU-accelerated multi-physics codes, ethical AI and governance.

Please identify the expected balance of CPU simulation, AI/ML training, AI inference, visualization, data analytics, and interactive use, including typical job scale, GPU count/job, memory/job, concurrency, and user demand.

The offeror should consider the wide variety of workloads referenced in the previous question's response as well as the intended architectural capabilities for each system as stated in APPENDIX A: Potential Solution Scenarios.

What GPU form factor is intended for the GPU-forward use cases?

For the GPU-forward use cases, the form factor would be expected to be the most performant version, typically the OAM or SXM versus a PCIe version, and to have sufficient networking capabilities for multi-GPU data sharing.

Are there mandatory processor, accelerator, or architecture constraints?

Only that they be TAA-compliant and able to support current and future high-performance computing and AI/ML workloads.

Is the desired commonality primarily across DSRC sites, across workload types, or both?

The desired commonality is to achieve the specified “Desired End State” providing a “cost-effective, efficient, and reliable enterprise approach covering multiple HPC systems operating at multiple sites within various classification environments”. The offeror should consider the resulting enterprise acquisition, administration, and maintenance as specified “Requirements” section.

Should CPU-heavy, mixed CPU/GPU, and GPU-forward systems share common host architecture, management tools, fabric operations, and spare parts where feasible, or may systems be optimized independently by site?

The Government desires as much commonality as feasible. Systems may be optimized independently by site, but it is up to the offeror to determine the best approach in optimizing a proposal that supports one or more DSRCs.

Would the Government favor a modular rack-scale approach in which validated CPU, GPU, storage, fabric, utility, and cooling building blocks can be added independently as site demand grows?

It is up to the offeror to determine and propose an approach to support the requirements.

For the MHPCC solution, the Government favors scalable, modular architectures that allow independent expansion to align with future workload growth and phased funding profiles.

Is a common enterprise management and observability model desired across DSRCs, while preserving fully separate operational and security boundaries for each classification domain?

Yes, this is a desirable goal, but it is up to the offeror to determine and propose an approach to support the requirements.

Storage, Data, and Fabric

Is the high-performance parallel scratch file system in offeror scope, or will each DSRC provide existing storage/file-system services?

High-performance scratch storage and associated file-system services should be considered within the offeror's scope. Solutions may use parallel file systems or other high-performance scale-out or distributed storage architectures appropriate to the proposed system and workload.

If in scope, please provide target usable capacity, aggregate throughput, IOPS/metadata needs, expected file profile, checkpoint requirements, and concurrent client scale.

Details of the file system implementations will be determined during scope collaboration with the selected offeror. Offerors should provide optimized solutions to match the expected workload based on the architectures being considered as listed in Appendix A.

Is cross-site data federation, replication, or a shared namespace desired, permitted, or prohibited?

The HPCMP is in the process of implementing an enterprise data fabric based on Hammerspace. The data fabric will be a shared namespace across HPCMP, connected to multiple HPCs at a given classification level. The initial Hammerspace installs will be in the collateral secret and unclassified realms. Any proposed system(s) in those classifications should be able to integrate the Hammerspace solution, making it available to the entire compute pool.

If any cross-site data service is allowed, please identify the approved DSRC pairs and classification boundaries.

See answer above.

What production parity is required for each Test and Development System?

See answer below.

Must the TDS replicate CPU/GPU families, fabric generation, file-system stack, scheduler/container environment, security configuration, and classification level, or is a more limited functional test environment acceptable?

In order to properly support testing of production system updates, it's generally expected that the TDS will match in all of the above-mentioned areas. Exceptions can be made in some instances. Final determinations should be made during scope collaboration with the selected offeror.

What security deliverables must the system provider provide for RMF/STIG, encryption/key management, vulnerability management, and classified-media lifecycle handling?

Leading up to final system acceptance, the offerors must support the required Assess Only process for all contracted systems. HPCMP Security approval for the Assess Only package (including STIGs, hardware/software checklists, etc) must be received prior to the system being connected to a DSRC's network.

Requirements to support encryption/key management will need to be determined during scope collaboration.

Assuming the selected offeror provides administration and maintenance support, vulnerability management and classified-media lifecycle handling would be detailed in the administration & maintenance requirements and formalized during scope collaboration.

Can the response use a teaming model in which an offeror provides systems, factory integration, Tier 2/3 support, and training while a cleared services partner provides enterprise administration and on-site services?

Yes, a support model comprised of multiple contractor teams is acceptable. All personnel directly involved in the factory integration of the solution must be U.S. citizens.

Depending on the type of Tier 2/3 support, those personnel may also need to be cleared. If Tier 2/3 support involves user data from the system, then a clearance would be required.

Details of the relationships between the offeror teams and the government, including project management, points of contact for various types of support and escalation procedures, should be provided during scope collaboration with the selected offeror.

How should the Navy two-hour response requirement be interpreted?

The Navy two-hour response requirement is intended as a two-hour acknowledgment of an issue.

Does it mean acknowledgment, remote diagnosis, on-site dispatch, service restoration, or full resolution? Are local spare parts required?

Contracted TI offerors typically keep a spare parts cache for each system at a given DSRC.

Section A describes stove-piped support across the DSRCs and states that the Government may obtain system administration and application support from parties other than the system provider. Would the Government consider a solution brief that offers only an enterprise support coordination and governance layer (cross-site service management, SLA/KPI framework, acceptance and transition planning, and vendor coordination among the OEM, administrator, and third parties), without proposing hardware or cleared on-site administration?

Yes, the Government will consider solutions that offer only enterprise support without proposing hardware or cleared on-site administration. Such solutions should describe how they would support hardware and on-site administration acquired through other sources.

If such a brief is acceptable, should it identify the specific OEM and administrator it would coordinate with, or may teaming with those parties be arranged after selection?

The brief should indicate how teaming would be accomplished.

May a solution brief be submitted by a team, with a prime and subcontractors or partners? If so, should each team member submit separately or should one brief cover the team?

Yes, a brief may be submitted by a team with a prime and subcontractors or partners identified. One brief will be sufficient.

Will small business or socioeconomic status (for example, WOSB, WBE, MBE, or DBE certification) be considered in selection or in later teaming and subcontracting, given the full and open basis of the CSO?

Small business participation will be considered after potential solutions are identified for selection and further actions.

Does the September 21 FAQ's exclusion of software-based solutions permit third-party orchestration software as a component of a complete new HPC-system offering with administration and maintenance support?

The exclusion was intended to remove from consideration any software-based solutions proposed for existing HPC systems. Any software-based solutions that are proposed with new HPC systems will be considered as part of the software stack delivered and/or as part of follow-on administration and maintenance support.

May an administration and maintenance provider submit independently of a new-system acquisition offer, or must its services be included in a system provider's submission?

Yes, an administration and maintenance provider may submit a solution independent of a new system acquisition. The solution should describe how it would support a new system(s) acquired from another source.

Please confirm the controlling submission deadline and time zone. The announcement lists October 23, 2026, at 4:00 pm CT but labels that date Thursday October 23 is Friday.

The day has been corrected. The submission deadline is Friday, October 23, at 4 pm CST/5 pm EST.