FAQs – Transformative AI/ML High-Performance Computing

Last updated October 7, 2025

Can the government confirm that the one page allocated to spell out acronyms, abbreviations and symbols is outside the 10-page limit for solution briefs?

This page is within the 10-page limit.

Can we still submit our solution to the Transformative AI/ML High-Performance Computing CSO if we do not currently have a SECRET clearance?

Yes. For this project, a facility clearance is preferred, but not mandatory.

Must all submissions include both a fully scaled hardware system and software stack?

The Government will consider all offers that fulfill the requirements in the CSO.

While the CSO emphasizes high-performance hardware, we're researching software technologies to control and secure data entering AI/ML systems at classified levels — including a standardized protocol to manage context and selectively ingest data. Would ERDC consider proposals focused on the data and access-control layer running atop the HPC hardware, or is the scope limited to physical infrastructure? Also, would ERDC be interested in R&D efforts that focus on developing a standardized, secure data-ingestion layer to reduce this bottleneck and improve system utilization?

The Government is interested in a full system proposal only including hardware, software, support, installation, training and consulting

Is there a preference for a Kubernetes orchestrator, are you tied to existing licensing schemes? NVAIE, NGC? What are the preferred testing and lab software tools? (ex: Jupyterlab? OCP AI?)

The Government anticipates Kubernetes orchestration will be part of the proposed solutions. The solutions need to support a wide range of AI/ML testing and benchmark applications. The Government is looking for a broad regimen of software tools.

Are there any per-rack power density limitations?

60KW per rack.

Per the Background section, several use cases are defined, are any of these use cases identified as the test-use-case for the initial system?

The Government is looking for a broad regimen of software tools based on background focus areas, and published DoD AI/ML priorities.

Will there be any required connections to public cloud / remote resources not described in the documentation? What networks are planned to provide these connections?

External resources can be whitelisted within CONUS and accessed via supported utilities (git, curl, https).

Per the Background section, is there any prioritization of these use cases for initial system use?

The Government is looking for a broad regimen of software tools based on Background focus areas.

Are there any use cases not listed which has surfaced since launch of this RFI?

No.

Will ARL pursue any ATO or IL security standards during this prototype analysis?

ARL operates under the DISA security profile and ATO.

Is there an interconnect fabric preferred (InfiniBand NDR/HDR, Ethernet 200/400/800G?)

ARL anticipates NDR IB fabric support.

Will the prototype be deployed via DREN?

Prototype will be accessed via DREN.

Do you require multi-tenancy / partitioning of GPU resources?

Multi-tenancy will be restricted at the NODE level initially.

Will the prototype submitters need to meet defined security standards?

ARL operates under the DISA security profile.

Will the prototype be accepted with air-cooling vs liquid-cooling?

ARL/DSRC prefers liquid-cooling solution based on existing infrastructure.

Is the government open to a multi-tenant, cloud-based alternative HPC environment at the secret IL6 level with similar capabilities that can scale down or up to meet demand?

No.

If the solution must be on-premises, what are the primary constraints that necessitate an on-premises system be located at ARL?

Security in depth.

If cloud-based resources are proposed, to what extent would they need to integrate with existing ARL or other DSRC facilities and resources (e.g. Storage), and at what locations?

The Government is looking for an on-premises solution.

Do the ARL facilities to be used already have physical security accreditation for SECRET operation (or higher if/where required)? If the ARL facilities do not have the appropriate accreditation, should those processes to attain necessary accreditation be included in the proposed scope of work?

No. An accreditation is in progress.

What is the anticipated capacity required and workload details for edge computing architectures and for other "distributed" environments?

Future edge integration can be considered a positive in solution proposal.

Where would these environments be located?

Future edge environments would be located across DREN.

Does the SOW include procurement and 5-year operation of infrastructure with a specific capacity at these locations, or only demonstration of capabilities such as portability of workloads to operate in those environments in the future?

No.

What, if any, GFE network capabilities will be provided for any required testing and operation of AI/ML distributed and edge computing infrastructure? To what extent would these locations need to operate independently with intermittent, contested, or no connectivity to each other and/or to the ARL DSRC and cloud environments?

DREN access will be provided. No connectivity to each other.

Can the Government provide additional details on evaluation of the solution with respect to "adaptability to dynamic environments?

The Government is looking for a broad regimen of software tools based on background focus areas.

What, if any, specific software solutions are planned, and would the Government be providing said software as GFE?

None.

What existing networks, such as DREN/SDREN and operational security levels (UNCLASS, SECRET, TOP SECRET), would be connected to the required new infrastructure at each of the locations?

DREN/SDREN (UNCLASS/SECRET)

For pricing purposes, what portion of the acceptance testing is expected to be performed by the contractor? Will the Government provide additional details about the software to be tested and government personnel supporting acceptance testing?

90% of system administration and workflow configuration will be performed by the selected vendor during acceptance testing. Government personnel and contractors will be developing and running workflows on the system to establish and debug the working environment.

Can the Government provide a breakdown showing the approximate percentage of the various anticipated AI/ML workloads which may require different GPU characteristics? (For example: ML model training vs inference, GenAI, and language vs image or video analysis.)

The Government cannot provide a percentage breakdown for AI/ML workloads, however, published DoD AI/ML priorities should provide an approximation.

What is the storage capacity needed per GPU of AI infrastructure?

10 TB per GPU.

What AI workloads/use cases does the Government anticipate deploying in this environment (eg. Autonomous Systems, Computer Vision, Combat Simulation, etc.)?

Published DoD AI/ML priorities should provide an approximation.

What portion of the AI infrastructure will be for training and what portion will be for inferencing?

40% will be for training; 60% will be for inferencing.

Does the Government require multi-node clustering for the GPUs?

Yes.

What kind of growth capacity does the Government expect to have over the next 5 years and does the Government expect pricing to include that capacity growth?

The Government expects possibly 50% growth. The Government does not expect pricing to include that capacity growth.

What are the storage requirements (eg. total storage, throughput, etc.)?

Total storage >= 10PB, 10GB/s per GPU.

Is there a requirement for high performance storage (eg. Lustre, NFS high performance, Weka, etc.)?

Total storage >= 10PB, 10GB/s per GPU.

Is there a requirement for high performance storage (eg. Lustre, NFS high performance, Weka, etc.)?

Yes.

What MLOps software or NVIDIA Software does the Government use today?

JAIC tools or other IL5/6 tools.

Does the Government plan to use the AI Infrastructure to train or fine tune any custom or open source models?

Yes.

Does the Government use containers (NGC, docker-which one) or run on the host directly?

It is currently authorized for singularity/podman.

Does the Government require any specific operating system?

It must meet DISA STIG compliance.

Does the Government require any scheduler (e.g. Slurm, or Kubernetes)?

Slurm is preferred for resource allocation; Kubernetes for orchestration within allocated resources.

What are the required amount of CPU and GPU, CPU Memory, GPU Memory?

Estimating ~600+ GPUs, 80GB HBM.

What is the anticipated amount of data per month to be transferred to and from this ERDC solution?

Multiple TB/month.

For Tier 2 and 3 support, will the Government provide hoteling space for onsite contractor activities?

No.

For acceptance testing, will the Government provide hoteling space for onsite contractor activities?

No.

Will the Government provide GFE for configuring solutions on customer networks?

Yes—for privileged access.

Will the Government provide 7 business days after receipt of answers for the offerors to incorporate changes?

No.

Would the Government consider any alternative pricing models (ie: consumption based) versus firm fixed price model?

No.

What are the Government's expectations for knowledge transfer to Government personnel (e.g., documentation, shadowing, train-the-trainer programs)?

The Government is expecting a rigorous training plan for system administrators, and customer service developers—with ongoing training and assistance available for customers.

Does the Government require the development of customized training/workshops specific to this environment, or are standard training materials acceptable?

Initially standard training should be acceptable.

To what extent is the contractor responsible for developing and delivering training for users and administrators?

The Government has a strong system administration and customer service cadre that will need intense training initially, then a continuing education for users.

Could the Government clarify which specific consulting activities are expected? For example, does the scope include requirements analysis, system architecture design, system integration, customization, training, documentation, or ongoing support?

System configuration and operation, customization, training, workflow development, problem investigation/resolution.

Is the vendor expected to provide only initial system delivery and implementation, or does the Government anticipate a need for ongoing advisory or strategic consulting services after deployment?

Ongoing advisory and consulting services.

What is the power capacity (MW) available for the Production System, and for the Test and Development System (TDS)?

2MW for production system, 0.2MW for TDS.

What is the requested capacity (TB) of high-speed storage?

Capacity <= 10PB.

Is the total system power draw expected to be 1.4 MW total (TDS + Main) or 1.2 MW total?

2MW for main, 0.2MW for TDS.

Do you have a significant need for x86 to be included on your GPU servers for your workloads?

Yes. Most of the Government's current code base is X86 validated, so it wants to merge those codes with new AI/ML capabilities.

What is the largest scale training you are looking to perform?

100B parameters.

Regarding integration with the ARL DSRC networking infrastructure, what are the preferred protocols or APIs for connecting to the Center Wide File System (CWFS) and Enterprise Data Fabric (EDF) via the 10/40 Gb/s private storage network, and are there any throughput or latency requirements for data transfer nodes?

NFS. No—there are no throughput or latency requirements.

What Workshare is the end user planning on using for this system?

Typically, the Government would run SLURM for resource allocation, then Kubernetes withing the allocation.

How will the computational load be divided across nodes/cores/GPUs?

Typically, users would be allocated a group of nodes via a resource allocation, then cores/GPUs would be orchestrated within that allocation.

Which job scheduler/resource manager are they planning to use?

Typically, the Government would run SLURM for resource allocation, then Kubernetes withing the allocation.

What workload mix the system will see? Batch jobs vs. interactive

50/50.

CPU-heavy vs. GPU-heavy

20/80.

AI/ML training vs. simulation/CAE

AI/ML/SIM: 80/20.

Short jobs vs. long-running jobs

50/50.

Whether it's single-tenant or multi-tenant use — i.e., do they have multiple research groups or teams that need fair-share scheduling?

Yes.

How will the system availability will be measured and how the fee will be prorated?

System availability will be measured with actual node time available per month/scheduled node time available per month. The fee will match the aggregate for the system.

Is there a specific type of software stack requirement?

The Government is looking for recommendations to meet published DoD AI/ML priorities.

What is the impact to your stated schedule if there is a Government shutdown?

None.

How much lead time will the awardee have for fielding the TDS system from notification, funding and delivery?

6 months.

Are there an additional list of required certifications, etc? (outside of those in the "constraints and considerations" section)

No.

Will there be additional customization needed after sharing recommended configurations?

Probably.

What mission outcomes will storage most influence for this prototype and potential follow on production?

The Government is looking for a fully supported AI/ML solution that includes HW/SW/High speed networking/storage/support/training and consulting. Storage can have a significant impact on overall functionality.

Which evaluation factors (technical risk, schedule, mission impact) will be weighted for the storage subsystem?

All evaluation factors will be applied to storage subsystem.

Do you anticipate a prototype-to-production path, and should designs demonstrate readiness for scale-out and sustainment beyond the prototype phase?

Production systems will be in the classified realm, while prototype will be used for preproduction testing.

What end to end performance KPIs will you measure at system level: sustained read/write bandwidth, random IOPS, latency, and metadata ops per second under mixed small and large file workloads?

The Government is looking for a fully supported AI/ML solution that includes HW/SW/High speed networking/storage/support/training and consulting. Storage can have a significant impact on overall usability.

How should storage demonstrate GPU saturation for AI/ML training and fine-tuning jobs at target cluster sizes, including high concurrency and overlapping IO/compute phases?

The Government is looking for a fully supported AI/ML solution that includes HW/SW/High speed networking/storage/support/training and consulting. Storage can have a significant impact on overall usability.

Are there representative datasets, job traces, or benchmark harnesses (e.g., data loading for LLMs, image pipelines, mixed POSIX+S3) that offer a realistic test of metadata intensity and small file behavior?

Representative datasets and benchmarks are under development currently.

Should the solution support non-disruptive horizontal expansion with deterministic performance scaling as capacity and compute grow?

Yes.

ARL recently awarded a substantial HPC storage contract for your HPC environments. When is the implementation expected to commence?

Not applicable.

The storage requirements specified in this request are they compatible with the existing storage infrastructure? If so, which connection type (Ethernet or InfiniBand) is preferred, and what are the associated bandwidth requirements?

Not applicable.

Our next-generation AI computing platform is set to be released soon, capable of generating 600KW of power per rack. Are you prepared to accommodate such upgrades over the coming years?

No—probably topping out ~ 100KW per rack.

Will the successful bidder of this contract automatically win the RFP?

See Section 5 (Selection) of the CSO announcement.

Do you have any plans to provide a larger supply of chilled water in the future?

Expansion is built into the current building infrastructure.

If you intend to establish connections to other DoD HPCMP sites, how do you envision establishing those connections?

100Gb/s DREN.

Could the vender provide additional cooling for the compute racks? (Bryan correct this question if I'm off)

Expansion is built into the current building infrastructure.

Has the U.S. Army decided on a data center facility infrastructure reliability mission? Example: Uptime Tier III, Uptime Tier IV.

The Government has a goal of reaching Tier 3, but due to costs probably will not get there.

Are we to provide CFD models?

No.

Please confirm that TI-26 Base system 1 Room of 2,730 sq. ft. is currently vacant and adjacent to an existing operating data center.

Confirmed.

In addition to 480V, are we to also consider/provide 415 V power to potential AI racks (i.e. NVIDIA, Lenovo, etc.)?

No.

Is there an existing raised floor system in the "vacant" ±2,730 sq. ft. room?

Yes.

Does the U. S. Army structural outline performance specification reflect the potential 3,000-4,000 lb. per rack loads of AI?

Yes.

What is the space (ft.) between the existing suspended ceiling and the underside of deck?

The height from raised floor to ceiling trusses is 10'.

Are the existing UPSs configured to support AI/HPC in rush loads?

Using IEM RotoBLOC UPS.

Is there a preferred method of power distribution to the racks? For example, overhead busway system or Row RPP?

Power distribution is PDUs at each row for lower voltage (220v) and dedicated whips for each rack back to the PPU for 480v. There is no over overhead bus for power.

Given that the TDS network is a fraction of the Ti26, would you like to explore a hybrid cloud solution via your DREN connection for flexible testing and innovation?

No.

Is the proposed 64 sq. ft. computer room space included as part of the data center floorplan that was attached (approximately ±2,400 sq. ft.)?

No.

Are there two (2) compute racks proposed for the ±64 sq. ft. area? If yes, what are the approximate loads (kW) per cabinet?

~60kw.

If AI, can we consider providing 415V power instead of 480V?

No.

Are the UPS systems designed to support AI loads?

Yes.

Would the Government consider granting a one-week extension?

No.

Please define which CUI datasets may be used during ATC/NAIC testing, sanitization rules, and data retention/return requirements.

The system will remain on unclassified DREN network until it is fully functional and government workflows are developed and tested, this could take months. The production system will then be moved to a classified SDREN network intact where classified datasets can then be loaded and processed. The TDS system will remain on unclassified DREN network indefinitely.

In addition to providing connectivity to the CWFS and EDF, is any additional storage required?

The Government is looking for a fully supported AI/ML solution that includes HW/SW/High speed networking/storage/support/training and consulting. Storage can have a significant impact on overall functionality.

Please confirm expected curriculum topics, duration per session, hands-on vs lecture mix, and evaluation method for the bi-annual training (up to 20 participants).

Bi-annual training should be focused on supported workflow tools and utilities to maximize usage of the delivered system. Training sessions can be virtual while highlighting system features and tools in support of published DoD AI/ML priorities.

If we could guarantee wattage levels in a cloud-based offering, would that be considered?

No.

What are the functional benchmarks? Are there any performance requirements associated with them?

IOR-SEQUENTIAL Test Description and Guidelines

IOR-SEQUENTIAL is a parallel file system performance test designed to determine the aggregate data throughput in gigabytes per second (GB/s). This test must be executed using IOR version 4.0.0 (available for download at https://github.com/hpc/ior/releases/tag/4.0.0) and must use **eight** MPI processes per node. For TI-26, offerors are required to run IOR using MPI-IO mode with one file per process. Offerors are not permitted to use optimizations related to client-side caching (e.g., file read ahead). The file system must implement, at a minimum, POSIX-compliant inode attributes. The use of "-noatime" mount options and similar mechanisms to disable access time updates are forbidden.

Using IOR, two measured executions must be performed:

(1) A "duplex" sequential read of 8GB per node (one 1GB file per process with 8 processes per node) while a longer sequential write of 64GB per node (one 8GB file per process with 8 processes per node) is in progress, and

(2) A "duplex" sequential write of 8GB per node (one 1GB file per process with 8 processes per node) while a longer sequential read of 64GB per node (one 8GB file per process with 8 processes per node) is in progress.

The two "duplex" rates must be summed to produce a single "full duplex" rate for that node count and must be converted into units of GB/s:

IOR result (in MiB/s) = "duplex" max read result (in MiB/s)
+ "duplex" max write result (in MiB/s)

IOR result (in GB/s) = IOR result (in MiB/s) * (1024**2) / (10**9)

Offerors must provide guaranteed rates for IOR-SEQUENTIAL in GB/s as calculated above. The IOR-SEQUENTIAL benchmark can be successfully executed using 16, 32, 64, and 128 nodes. Please consult the **Performance Work Statement (PWS)** for specific requirements associated with each order.

For a given node count, all jobs will be executed in the same directory with eight MPI processes per node. The "duplex" read and write executions each require two IOR jobs executing at the same time: a long-running IOR job that provides background I/O traffic, and a short-running IOR job (of the opposite kind) that measures performance.

For runs measuring guaranteed performance, the number of iterations required for the foreground IOR-SEQUENTIAL job can be determined using the following formula:

iter = 768/<nodes>

Using this formula, iter = 48, 24, 12, and 6 for 16, 32, 64, and 128 nodes, respectively. The offeror must specify the number of iterations for the long-running background IOR job so that the long-running IOR job spans the entire execution of the short-running job.

After the long-running I/O traffic job is confirmed to be executing, the short-running measured job of the opposite kind may be launched. When the short-running measured read job completes, the long-running write job should be allowed to complete normally because the files it is writing will be needed by the subsequent long-running read job.

For each file system, five IOR jobs are required:
1. a set-up job to create files needed for the subsequent measured read job,
2. a long-running IOR write running concurrently with
3. a short-running IOR read (the duplex read), and
4. a long-running IOR read running concurrently with
5. a short-running IOR write (the duplex write).

An example of the required job sequence for a 64-node case is illustrated below. It is possible to combine background and foreground jobs in the same batch script, in which case only three batch jobs are needed.

Examples of combined jobs can be found in the "ABTP/ded/ior/seq/" directory.

Setup:
```
aprun -n 512 -N 8 ./ior -A 1 -a MPIIO -vv -w -F -o data -k \
   -b 1024M -t 4096K -i 12 -m
```

Duplex read test:

```
(sequential write background job)
aprun -n 512 -N 8 ./ior -A 1 -a MPIIO -vv -w -F -o data_bg -k \
   -b 8192M -t 4096K -i 12 -m &
sleep 20

(sequential read foreground job)
aprun -n 512 -N 8 ./ior -A 1 -a MPIIO -vv -r -F -o data -k \
   -b 1024M -t 4096K -i 12 -m
wait
```

Duplex write test:

```
(sequential read background job)
aprun -n 512 -N 8 ./ior -A 1 -a MPIIO -vv -r -F -o data_bg -k \
   -b 8192M -t 4096K -i 12 -m &
sleep 20

(sequential write foreground job)
aprun -n 512 -N 8 ./ior -A 1 -a MPIIO -vv -w -F -o data -k \
   -b 1024M -t 4096K -i 12 -m
wait
```

The output for each measured execution must be preserved.

IOR-RANDOM Test Description and Guidelines

IOR-RANDOM is a parallel file system performance test designed to determine the aggregate data throughput in gigabytes per second (GB/s). This test must be executed using IOR version

4.0.0 (available for download at https://github.com/hpc/ior/releases/tag/4.0.0) and must use **one** MPI process per node. For TI-26, offerors are required to run using MPI-IO mode with one file per process. Offerors are not permitted to use optimizations related to client-side caching (e.g., file read ahead). The file system must implement, at a minimum, POSIX-compliant inode attributes. The use of "-noatime" mount options and similar mechanisms to disable access time updates are forbidden.

Using IOR-RANDOM, two measured executions must be performed:

(1) a "duplex" random read of a 0.5GB binary file while a longer sequential write operation (i.e., 64GB) is in progress and

(2) a "duplex" random write of a 0.5GB binary file while a longer sequential read operation (i.e., 64GB) is in progress.

The two "duplex" rates must be summed to produce a single "full duplex" rate for that node count and must be converted into units of GB/s:

IOR result (in MiB/s) = "duplex" max read result (in MiB/s)
+ "duplex" max write result (in MiB/s)

IOR result (in GB/s) = IOR result (in MiB/s) * (1024**2) / (10**9)

Offerors must provide guaranteed rates for IOR-RANDOM in GB/s as calculated above. The IOR-RANDOM benchmark can be successfully executed using 16, 32, 64 and 128 nodes. Please consult the **Performance Work Statement (PWS)** for specific requirements associated with each order.

For a given node count, all jobs will be executed in the same directory with **one** MPI process per node. The "duplex" read and write executions each require two IOR jobs executing at the same time: a long-running IOR job which provides background sequential I/O traffic and a short-running IOR job (of the opposite kind) which measures random performance.

The number of iterations for the foreground IOR-RANDOM jobs are fixed at 8 for the 16 nodes runs, 6 for the 32 node runs, 4 for the 64 node runs, and 3 for the 128 node runs. The offeror must specify the number of iterations for the long-running background IOR job so that the long-running IOR job spans the entire execution of the short-running job.

After the long-running I/O traffic job has been confirmed as executing, the short-running measured job of the opposite kind may be launched. When the short-running measured read job completes, the long-running write job should be allowed to complete normally because the files written will be needed by the subsequent long-running read job.

For each node count, five IOR jobs are required:
1. a set-up job to create files needed for the subsequent measured read job,
2. a long-running IOR sequential write running concurrently with
3. a short-running IOR random read (the duplex read), and
4. a long-running IOR sequential read running concurrently with
5. a short-running IOR random write (the duplex write).

An example of the required job sequence for the 64-node case for IOR-RANDOM is illustrated below. Background and foreground jobs may be combined into the same batch script, in which case only three batch jobs are needed. Examples of combined jobs can be found in the "ABTP/ded/ior/ran/scripts" directory.

IOR-RANDOM

Set-up:

```
aprun -n 64 -N 1 ./ior -A 1 -a MPIIO -vv -w -F -o data –z \
   -k -b 512M -t 4K -i 4 -m
```

Duplex read test:

```
(sequential write background job)
aprun -n 64 -N 1 ./ior -A 1 -a MPIIO -vv -w -F -o data_bg \
   -k -b 65536M -t 4K -i 32 -m &
sleep 20

(random read foreground job)
aprun -n 64 -N 1 ./ior -A 1 -a MPIIO -vv -r -F -o data -z \
   -k -b 512M -t 4K -i 4 -m
wait
```

Duplex write test:

```
(sequential read background job)
aprun -n 64 -N 1 ./ior -A 1 -a MPIIO -vv -r -F -o data_bg \
   -k -b 65536M -t 4K -i 32 -m &
sleep 20

(random write foreground job)
aprun -n 64 -N 1 ./ior -A 1 -a MPIIO -vv -w -F -o data -z \
   -k -b 512M -t 4K -i 4 -m
wait
```

The output for each measured execution must be preserved.

MDTEST Test Description and Guidelines

MDTEST is an MPI-based application for testing the metadata performance of file systems. It can be run on any POSIX-compliant file system, including parallel file systems such as Lustre. When run on a parallel file system, the file system must be fully installed and configured with all services running including the management server (MGS), metadata server (MDS), and object storage servers (OSS). Also, there must be a population of clients to mount the parallel files system on all nodes in the compute pool.

MDTEST source code and its development has been combined with the IOR project (https://github.com/hpc/ior). The compilation of IOR also includes the compilation of MDTEST. Therefore, MDTEST source code is no longer distributed with the ABTP. Instead, IOR version 4.0.0 is required for this benchmark and can be downloaded at https://github.com/hpc/ior/releases/tag/4.0.0.

To meet the benchmark testing requirements, each run of MDTEST requires two jobs to be executed simultaneously: one background job to provide I/O traffic and one foreground job to measure the file system performance, reported as I/O operations per second (IOPS). Both jobs are executed on the same number of nodes (non-overlapping) with **four** MPI processes per node. The background job should be launched first, and the foreground job should be launched after the background job begins execution. When the foreground job is completed, the background job can be terminated. The test package includes scripts to orchestrate the execution of the two jobs with the number of iterations set to keep the background job running during the full execution of the foreground job. Vendors may modify the number of background iterations but must ensure that the background job completely encapsulates the foreground job. The test package accuracy test is used to ensure the proper order of execution was followed.

The TI-26 MDTEST benchmark uses weak scaling (i.e., each process/node performs the same amount of file operations regardless of the number of nodes used). Each foreground process will create, stat, and remove 6,000 files for 10 iterations at the leaf nodes of a directory structure with 3 nested levels and a branching factor of 10. The MDTEST command line for the foreground processes is as follows.

Foreground: mdtest –d <filesys>/duplex –F –L –b10 –z3 –n6000 –i10

Each background process will create, stat, and remove 60 files for 100,000 iterations at the leaf nodes of a directory structure with only 1 level and a branching factor of 10. The MDTEST command line for the background processes is as follows.

Background: mdtest –d <filesys>/background –F –L –b10 –z1 –n60 –i100000

In each command, <filesys> represents the path to the file system under test.

MDTEST reports metadata IOPS rate maximums, minimums, means, and standard deviations for file creation, stat, read, and removal, as well as, for tree creation and removal. Offerors

must report the **mean** rates for file creates, file stats, file removals, and their sum for each required node count as seen in the example output below.

Value to report = duplex mean file creates + mean file stats
+ mean file removals

The file system must implement, at a minimum, POSIX-compliant inode attributes. To ensure accurate performance data, the use of “noatime” mount options and similar mechanisms to disable access time updates are forbidden. Performance may be improved on Lustre file systems by setting the number of metadata targets (MDTs) and the MDT index used for the test directories. However, the foreground and background test directories must use the same set of MDTs, so as not to reduce contention.

The MDTEST benchmark can be successfully executed using a variety to be determined nodes based on size of machine.

The output (*.o* file) for each measured execution must be preserved, as well as the job directory and its contents.

FUN3D-HVW-GPU Test Description and Guidelines

FUN3D is a computational fluid dynamics (CFD) code developed by NASA. It performs flow analysis for all the major regimes on the Mach scale. While it does not contain a grid generation capability, it accepts 3D grids from VGRID, SolidMesh/AFLR3, Pointwise, and GridEx, and 2D grids from AFLR2 and BAMG [1].

The FUN3D application source code is export restricted and will not be supplied with the TI-26 distribution. "It can only be given to a 'U.S. Person,' which is a citizen of the United States, a lawful permanent resident alien of the U.S., or someone in the U.S. as a protected political asylee or under amnesty" [1]. More information about FUN3D and its licensing requirements can be found at https://fun3d.larc.nasa.gov/chapter-1.html#request_fun3d. TI-26 offerors may request the software directly from NASA by emailing Ms. Maxine Saunders at maxine.a.saunders@nasa.gov or larc-sra@mail.nasa.gov. Please consult Appendix A of this document for specific information related to obtaining FUN3D for TI-26 benchmarking activities.

The TI-26 ABTP distribution includes three test cases for FUN3D that target the accuracy and performance of the application on the two ends of the Mach scale: incompressible flows and hypersonic flows. The third test case, named "HVW-GPU," stands for "hypersonic viscous waverider." HVW-GPU is a generic, viscous-optimized hypersonic waverider airframe geometry.

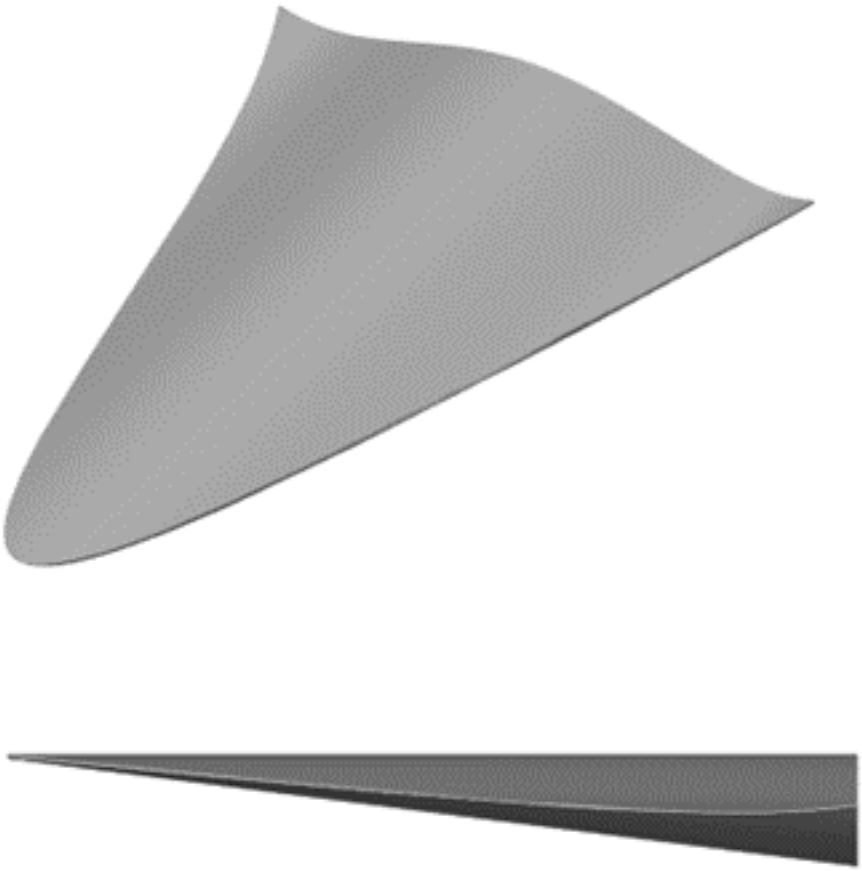

Figure 3: Waverider Geometry in Isometric View (top) and Side View (bottom)

The mesh used for HVW-GPU has roughly 7 million grid points.

For this test case, the HLLE++ Riemann solver and van Albada limiter are used for the inviscid fluxes. The freestream conditions used for the problem are those at an altitude of 100,000 ft, which is the median altitude for an airbreathing hypersonic vehicle at Mach 8 [2]. With a freestream Reynolds number of over $3x10^6$, the flow is assumed to be fully turbulent, and a compressible form of the Spalart-Allmaras (SA-Catris) turbulence model is used. The solver uses FUN3D's generic gas equation set with a five species, one temperature non-equilibrium air model. The initial composition of air is set to the standard two species composition of 76.7% diatomic nitrogen and 23.3% diatomic oxygen. The surface of the waverider is set to a no-slip, non-catalytic wall with a constant wall temperature of 300 Kelvin.

FUN3D version 14.2-ffaff71 will be used in TI-26. **This test case is required to be run on accelerators (e.g., GPUs, APUs)**. The provided HVW-GPU benchmark can be successfully executed using 4, 8, and 12 GPUs. Please consult the **Performance Work Statement (PWS)** for specific requirements associated with each order.

An accuracy check will be executed as the final step in the benchmark run. The FUN3D benchmarks must run to completion and must meet all accuracy criteria. Offerors must provide guaranteed total time-to-solution, calculated as elapsed execution time, for each required process count. The proper timing quantity is calculated in the batch scripts and reported as "walltime".

Sample batch script templates and an automated tool for generating batch scripts can be found in "ABTP/ded/fun3d/hvw-gpu/scripts". The generated batch output files (fun3d_hvw-gpu_<procs>.o<jobid>) will contain everything written by the batch script to stdout and stderr.

Each complete execution requires submission of the following three items:

1. Total walltime with a PASSES accuracy result
2. The batch output file in its original, unedited form, which also contains a copy of the batch script
3. All output directories except "restart" and "outputs/visualization"

The provided example scripts use PBS, LSF, or SLURM commands. If these tools do not match the configuration of the systems being benchmarked, offerors may edit the scripts, using the provided scripts as a guide to the execution process.

[1] Anderson, W.K., et al., "FUN3D Manual: 14.1," NASA TM-20240006306, June 2024.

[2] James, L., Hunt and Eiswirth, E., "NASA''s Dual-Fuel Airbreathing Hypersonic Vehicle Study," Nov. 1996.

Appendix A: Application Licensing Information

FUN3D LICENSING INFORMATION

The FUN3D source code is export restricted and will not be supplied with the TI-26 distribution. It can only be given to a 'U.S. Person,' which is a citizen of the United States, a lawful permanent resident alien of the U.S., or someone in the U.S. as a protected political asylee or under amnesty. More information can be found at https://fun3d.larc.nasa.gov/chapter-1.html#request_fun3d.

Offerors must request the FUN3D software through the NASA Software Catalog via https://software.nasa.gov/software/LAR-20188-1. Please have the following information ready as NASA will need it to process the request:

- Company Information:
  - Company name
  - Company address **[No P.O. Boxes are allowed]**
  - Is this a US owned company?
  - Does this company have a parent company or is "Doing Business As" for another company and if so what country?

- Software Recipient/User Information **[This is the person that will receive the source code.]**
  - Full legal name
  - Work address
  - Email address **[use company email]**
  - Phone number
  - Reason for software

- Signatory Authority Information **[This is someone who can legally bind the company, for example, President/CEO, Procurement or Contracting Officer, Legal Counsel, Senior Management.]**
  - Full legal name
  - Title
  - Company address
  - Phone
  - Email address **[use company email]**

- Contract Information
  - If responding using a specific Government contracting vehicle (e.g., BOA, ITAS), please use that contracting information (contract number, contracting agency, COR, expiration date)

FUN3D version 14.2-ffaff71 is required for TI-26 benchmarking activities.

The government reserves the right to include other benchmarks throughout the process.