Contents

[1. Objective of NWSC-2 Benchmarking 1](#_Toc415665092)

[2. Ownership and Dissemination of Results 2](#_Toc415665093)

[3. UCAR Right to Request Benchmark Rerun Before Award 2](#_Toc415665094)

[4. NWSC-2 Benchmark Rules 2](#_Toc415665095)

[4.1. General Rules and Reporting 2](#_Toc415665096)

[4.2. Core and Memory Subscription 4](#_Toc415665097)

[4.3. Benchmark System(s) 5](#_Toc415665098)

[4.4. Benchmark Results Projection 5](#_Toc415665099)

[4.5. As-Is and Optimized Benchmark Results 5](#_Toc415665100)

[4.6. Benchmark Code Changes and Optimization 6](#_Toc415665101)

[5. Reporting of Benchmark Results 6](#_Toc415665102)

[5.1. NCAR Benchmark Suite Computational Benchmarks 6](#_Toc415665103)

[5.1.1. HOMME 7](#_Toc415665104)

[5.1.2. HOMME\_COMM 7](#_Toc415665105)

[5.1.3. HPCG 7](#_Toc415665106)

[5.1.4. LES 7](#_Toc415665107)

[5.1.5. MG2 7](#_Toc415665108)

[5.1.6. MPAS-A 8](#_Toc415665109)

[5.1.7. POPperf 8](#_Toc415665110)

[5.1.8. WRF 8](#_Toc415665111)

[5.2. Micro-Benchmarks 9](#_Toc415665112)

[5.2.1. OSU MPI 9](#_Toc415665113)

[5.2.2. SHOC 9](#_Toc415665114)

[5.2.3. STREAM 9](#_Toc415665115)

[5.3. I/O Benchmarks 9](#_Toc415665116)

[5.3.1. General 9](#_Toc415665117)

[5.3.2. Reporting of I/O Benchmark Results 10](#_Toc415665118)

[5.3.3. Optional Runs 10](#_Toc415665119)

[5.3.4. IOR 10](#_Toc415665120)

[5.3.5. MDTEST 11](#_Toc415665121)

[5.3.6. pyReshaper 11](#_Toc415665122)

# Objective of NWSC-2 Benchmarking

The purpose of the NWSC-2 benchmarking is to evaluate the performance of NCAR applications on high-performance computing (HPC) and parallel file system (PFS) systems proposed by the Offeror in response to the UCAR RFP R15-19374 (NWSC-2). The benchmark applications are constructed to evaluate proposed technologies and to represent the computational and data storage demands of scientific simulation and data analysis codes representative of the current and anticipated research to be conducted at the NCAR Wyoming Supercomputing Center.

The NWSC-2 Technical Specifications, section 3.6 “Application Performance Specifications,” describes the NCAR Benchmarking Suite and the I/O and micro-benchmarks associated with the NWSC-2 procurement. That section also provides general guidance to Offerors regarding the use of these benchmarks in both proposal evaluation and acceptance testing. The NWSC-2 Benchmark Instructions (this document) provides specific guidance to Offerors on how the benchmarks are to be run, what information should be returned to UCAR and other reporting requirements.

The Offeror shall run the portions of the benchmarking applications, and report the results thereof, corresponding to the NWSC-2 resources proposed by the Offeror. For example, an Offeror proposing only a PFS resource need not run the computational benchmarks.

# Ownership and Dissemination of Results

Results shall include, but not be limited to, reports, output files, performance and other measurement data and event logs from the benchmark tests, written or otherwise provided by the Offeror to UCAR under this solicitation. The results shall be the sole property of UCAR.

# UCAR Right to Request Benchmark Rerun Before Award

UCAR reserves the right to request that the Offeror rerun any or all NWSC-2 benchmarks before a final award is made. Specifically, UCAR may ask an Offeror to rerun a portion of the benchmark suite should questions arise from the results provided in the Offeror’s response to this RFP. The Offeror may be expected to run selected portions of the NWSC-2 benchmark applications during a Live Test Demonstration, should it be requested by UCAR, to corroborate benchmark results provided in its response to this RFP.

# NWSC-2 Benchmark Rules

This section provides information for obtaining the benchmark codes, their build and execution instructions, and the general rules for running and reporting the results.

## General Rules and Reporting

The benchmarks used for the NWSC-2 procurement are available from the NCAR Benchmarking website: <https://www2.cisl.ucar.edu/hpc_benchmarking>. This site includes links for each of the individual benchmarks and includes source files, build scripts, instructions for building and running the benchmarks, and numerical validation criteria. Offerors should note that, while the NWSC-2 procurement utilizes the NCAR Benchmarking website site, the benchmarking site is independent of the procurement and will be maintained in the future as a source for NCAR-related benchmark codes.

Guidance on which benchmark results and output files are to be returned to UCAR in compressed tar files is provided in §5 and its subsections. Additionally, the Offeror should obtain, from the NWSC-2 RFP website <https://www2.cisl.ucar.edu/NWSC-2>, the benchmark results spreadsheet (**NWSC-2 Benchmark Results.xlsx**). The Offeror shall use the spreadsheet to report timing results of key NWSC-2 HPC computational benchmarks to UCAR with the Offeror’s proposal. The measured performance of the computational benchmarks on NCAR’s IBM iDataPlex **Yellowstone** system is provided therein. With input from the Offeror, the spreadsheet will use the benchmark results to calculate a Yellowstone Sustained Equivalent Performance (**YSEP**) of the proposed HPC production resource. The **YSEP** metric produced by the spreadsheet, and the algorithms used therein, provides a first-order approximation of the NWSC-2 HPC system’s Sustained Performance utilizing both “as-is” and, optionally, “optimized” benchmark results. Offerors should note that only some of the NWSC-2 benchmark timings are entered in the spreadsheet; this does not diminish the importance to UCAR of the remaining benchmarks.

The following table provides a list of worksheets contained in the NWSC-2 benchmark results spreadsheet and a brief description of their use and purpose. Each worksheet is protected (without a password) to prevent inadvertent modification of cells other than those into which results are to be entered. The Offeror may unprotect a worksheet, but is cautioned to not change any cells other than those into which information can be entered while the worksheet is protected.

| Worksheet | Description / Use |
| --- | --- |
| YSEP Benchmark As-Is | For reporting the Offeror’s as-is benchmark timings.The **as-is** timing results for the selected benchmarks should be entered in column C of this worksheet. The number of cores of the benchmark system may also be entered in cell J5. |
| YSEP Benchmark Optimized | For reporting the Offeror’s optimized benchmark timings (optional).The **optimized** timing results for the selected benchmarks should be entered in column C of this worksheet. |
| YSEP Proposed | For calculating the Yellowstone Sustained Equivalent Performance (YSEP) of the proposed system.The number of cores of the Offeror’s proposed system should be entered in cell M5. |
| Benchmark Projection Algorithms | For recording the algorithm(s) used for projecting performance from the Offeror’s benchmark system to the proposed system.The algorithm(s) shall be placed in column C. |
| Benchmark System | For recording the attributes of the Offeror’s benchmark system’s hardware and software.The requested information shall be placed in column B. |

With the exception of those contained within the **Benchmark Projection Algorithms** worksheet, the algorithms provided in the spreadsheet should not be changed by the Offeror. The Offeror is allowed to provide additional, alternate algorithm(s) for determining YSEP (for instance, if the algorithm does not account for a unique attribute of the Offeror’s system or, as stated in Section 4.2, the Offeror needs to account for core or memory undersubscription), provided the benchmark weight factors and configurations are unaltered. If the Offeror provides an alternate algorithm, it shall be included in the returned **NWSC-2 Benchmark Results.xlsx** spreadsheet. Additionally, the technical volume of the Offeror’s proposal shall provide a full description of, and rationale for using, the alternate algorithm.

## Core and Memory Subscription

All benchmarks shall be run in a fully subscribed mode, i.e., utilizing all cores within all participating nodes of a given benchmark run, unless it is not possible for one of the following reasons:

1. Requested core count is not evenly divisible by the number of cores per node on the Offeror’s system. In this case, the benchmark shall be run with all but one node fully subscribed. For example, if a 128-core benchmark run is requested, and the Offeror’s system is based on 24-core nodes, then the benchmark must be run on six nodes, using all 24 Cores on five of the nodes, and eight cores on the remaining node.
2. Memory limitations prevent a benchmark configuration from running in a fully subscribed mode. In this case, the benchmark shall be run using the minimum number of nodes required to allow the benchmark to run, and the cores used may be spread evenly across participating nodes. For example, if a 128-core benchmark configuration is requested and requires 2 gigabytes of memory per core, and the Offeror’s system consists of 24-core nodes with 1.5 gigabytes per core, then it will require eight nodes to run the benchmark. In this case, the benchmark may be run using 16 cores on each of the eight nodes.

If either of the above undersubscribed cases is used by the Offeror, the technical volume of the Offeror’s proposal shall:

1. Provide a detailed explanation of why the benchmark could not run in a fully subscribed mode, and how the undersubscribed run was configured to allow the benchmark to run.
2. Document, explain and make the appropriate additions to the benchmark results spreadsheet to provide an additional, alternate calculation of **YSEP** to account for the unutilized cores, if necessary.

## Benchmark System(s)

The system on which the Offeror runs the NWSC-2 benchmarks shall be as architecturally close as possible to that proposed by the Offeror.

The Offeror shall provide the key characteristics of the benchmark system(s) as requested in the “**Benchmark System**“ worksheet of the benchmark results spreadsheet and provide a description of the system in the technical volume of the Offeror’s proposal. All benchmarks shall be run on the same benchmark system with an identical system configuration used for all benchmarks. If the Offeror runs some benchmarks on a different system, the Offeror shall provide a complete rationale for doing so and explain how it is of benefit to UCAR.

## Benchmark Results Projection

The reporting of synthetic or fully simulated benchmark results is not acceptable. Should the Offeror use a benchmark system that is not identical to the proposed system(s), the technical volume of the Offeror’s proposal shall supply and fully describe the scaling model used to project benchmark timing from the benchmark system(s) to the proposed system(s) from actual measured results on the Offeror’s benchmark system(s). The Offeror shall provide the scaling model as equations in the “**Benchmark Projection Algorithms**” worksheet in the NWSC-2 benchmark results spreadsheet. If the algorithm is sufficiently complex that the worksheet cannot accommodate it, scaling factors should be entered in the worksheet and the scaling model should be supplied, along with the description requested above. The Offeror’s description of the scaling model shall include all measurements and the rationale used to validate the scaling model and shall demonstrate to UCAR the suitability and fidelity of the scaling model.

## As-Is and Optimized Benchmark Results

The Offeror shall supply benchmark results from “as-is” and, optionally, “optimized” configurations as described below.

|  |  |
| --- | --- |
| “as-is” | The Offeror shall return results for all benchmarks using a “base set” of compiler flags and runtime environment settings that allow a benchmark to pass its numerical validation criteria. No application source modifications are allowed. |
| “optimized” | The Offeror may return additional benchmark results with any level of optimization beyond the as-is “base set”, including those resulting from acceptable source code modifications as described in Section 4.6, that still allows a benchmark to pass its numerical validation criteria. |

For benchmarks that include numerical validation criteria, all configurations of a given benchmark must be run with optimizations no more aggressive than those used for the benchmark’s validation.

## Benchmark Code Changes and Optimization

In addition to compiler flags and run-time settings, source code modifications are allowed for submission of “optimized” results, however:

1. The Offeror may not change the floating-point precision of any of the NWSC-2 benchmarks.
2. No assembly-level recoding is permitted.
3. Source code changes are preferably written in the original source language (Fortran, C, C++), or via addition of pragmas and directives (OpenMP, OpenACC, etc.). Rewriting in alternate or proprietary languages (e.g. CUDA, OpenCL) is permissible but discouraged.
4. All source code modifications shall be isolated and enabled or disabled via conditional compilation using pre-processor #if/#endif definitions. For example:

#if (defined NWSC2\_*Offeror*)

 *Offeror-specific code*

#else

 *Original code*

#endif

Where the Offeror should substitute an appropriate moniker for *Offeror*.

1. If extensive changes to a piece of code make the #if/#endif unwieldy, a substitution of a new source file and a renaming of the old is acceptable. For example, rename the original file foo.F to foo.F.orig. The new source shall have the same base name but the extension should reflect the source language used.
2. Modified source code must still pass the individual benchmark’s validation criteria.

All modifications to the NWSC-2 benchmarks shall be documented within the technical volume of the Offeror’s proposal, and modified source files are to be returned to UCAR with the requested benchmark output files in compressed tar files.

# Reporting of Benchmark Results

This section provides the specific instructions for reporting of benchmark results to UCAR in the technical volume of the Offeror’s NWSC-2 proposal. In all cases, compressed tar files should be used to return to UCAR any requested files.

## NCAR Benchmark Suite Computational Benchmarks

The computational benchmarks shall be run by Offerors proposing an NWSC-2 HPC system and as described in the instructions on the NCAR Benchmarks website. The following sections describe what information and files are to be returned to UCAR with the Offeror’s technical proposal.

### HOMME

HOMME should be run on the following number of cores: 1152, 3456, 6912, 9216, 11520. For each HOMME result submitted to UCAR, the HommeTime\_stats timing file should be returned along with standard error and standard output from each run.

The following timer data should be entered into the **NWSC-2 Benchmark Results.xlsx** spreadsheet:

Column 6 of timer "prim\_run" in file "HommeTime\_stats"

Example command: grep "prim\_run" HommeTime\_stats | awk '{print $6}'

### HOMME\_COMM

HOMME\_COMM should be run on the following number of cores: 1152, 3456, 6912, 9216, 11520. For each HOMME\_COMM result submitted to UCAR, the HommeTime\_stats timing file should be returned along with standard error and standard output from each run.

The following timer data should be entered into the **NWSC-2 Benchmark Results.xlsx** spreadsheet:

Column 6 of timer "meth1" in file "HommeTime\_stats"

Example command: grep "meth1" HommeTime\_stats | awk '{print $6}'

### HPCG

The HPCG benchmark should be run on the following node counts: 1, 10, 50, 150, and 250 nodes. The number of MPI ranks and OpenMP threads to place on each node is not prescribed; it should be chosen to maximize performance. In addition to the node counts above, HPCG results for larger configurations that showcase large scale performance are of interest and may be returned as an option. For each HPCG result submitted to UCAR, the **hpcg\_log\_<time stamp>.txt** file, and **HPCG-Benchmark-2.4\_<time stamp>.yaml** file should be returned, along with standard error and standard output for each run.

### LES

The LES benchmark should be run on 512, 1024, and 4096 cores. For each result submitted to UCAR, the **bench.out** file containing timing data should be returned along with standard output and standard error for each run.

The following timer data should be entered into the **NWSC-2 Benchmark Results.xlsx** spreadsheet:

Timer "time-io" in file bench.out

Example command: grep "time-io" bench.out | awk '{print $3}'

### MG2

To submit results for the MG2 benchmark to UCAR, it should be run in several ways:

1. Using a single thread
2. Using one thread for each physical processor core in a shared memory node
3. Using multiple threads for each physical processor core if this provides a performance benefit (e.g. Hyper-Threading, SMT)

For each result submitted to UCAR, the standard output and standard error from the run should be returned.

The following timer data should be entered into the **NWSC-2 Benchmark Results.xlsx** spreadsheet:

Timer "total-time" printed to standard out.

Example output:

[MG2] [NTHR := 2] [itmax:= 10000] total time (sec): 5.802

### MPAS-A

The MPAS-A benchmark consists of two cases: 30 km and 15 km. The 30 km case should be run on the following core counts: 1152, 2304, 3456, 4608, 6912 and 9216. The 15 km case should be run on 6912, 9216 and 13824 cores. For each run of the benchmark, the standard error and standard output should be saved and returned along with the files **log.0000.out** and **log.0000.err** for each run.

The following timer data for both the 30km and 15km cases should be entered into the **NWSC-2 Benchmark Results.xlsx** spreadsheet:

Timer "time integration" from file "log.0000.out"

Example command: grep "time integ" log.0000.out | awk '{print $4}'

### POPperf

The POPperf benchmark should be run using 1084, 2168, 4336 and 7552 cores. For each run of the POPperf benchmark submitted, standard output and standard error should be captured and returned to UCAR.

The following timer data should be entered into the **NWSC-2 Benchmark Results.xlsx** spreadsheet:

Final "STEP" time from run log

Example command: grep STEP <log name> | tail -n1 | awk '{print $9}'

### WRF

The WRF (V3) benchmark should be run using 576, 1152, 2304, 4608, 6912, and 9216 cores. For each run of the benchmark, the standard error and standard output should be saved and returned along with the files **rsl.out.0000**, **rsl.err.0000**, and **wrfout\_d01\_2005-06-04\_09\_00\_00**.

The value for “Tot time” printed by the **timing.pl** script should be entered into the **NWSC-2 Benchmark Results.xlsx** spreadsheet:

“Tot time” is the sum of timers "Timing for main", ignoring first time step. The timing script is included in WRF source.

Example command: ./timing.pl

## Micro-Benchmarks

The micro-benchmarks shall be run by Offerors proposing an NWSC-2 HPC system and as described in the instructions on the NCAR Benchmarks website. The following sections describe what information and files are to be returned to UCAR with the Offeror’s technical proposal.

### OSU MPI

To submit results for the OSU MPI benchmarks to UCAR, each individual test should be run as detailed in the benchmark instructions. The **Makefile** containing the compilation options should be returned to UCAR, along with standard output and standard error from each run.

### SHOC

The SHOC benchmarks (BusSpeedReadback, DeviceMemory, MD, Stencil2D) should be run using a single GPU in a single node. Additionally, the Stencil2D benchmark should be run in parallel utilizing all GPUs in a single node, if multi-GPU nodes are being offered. For each run of the SHOC benchmark, standard output and standard error should be returned, along with the Logs directory and its contents.

### STREAM

To submit results for the STREAM benchmark to UCAR, it should be run in several ways:

1. Using a single thread
2. Using one thread for each physical processor core in a shared memory node
3. Using multiple threads for each physical processor core if this provides a performance benefit (e.g. Hyper-Threading, SMT)

Standard output containing performance timings and validation information should be returned for each submitted result.

## I/O Benchmarks

The I/O benchmarks shall be run by Offerors proposing an NWSC-2 PFS system and as described in the instructions on the NCAR Benchmarks website. The following sections describe what information and files are to be returned to UCAR with the Offeror’s technical proposal.

### General

Modifications to the I/O benchmarks are only permissible to enable correct execution on the target platform. Any modifications must be fully documented and reported back to UCAR. Changes related to optimization and tuning must be practical for production utilization of the file system. For example, optimizations that are tuning hints that can be controlled by users on the system are permissible, but optimizations that would require superuser privilege (but are not part of the steady-state configuration of the file system) are not permissible.

### Reporting of I/O Benchmark Results

The output file for each I/O benchmark run will reside in the main directory. In addition to this output file, Offeror’s must provide an end-to-end description of the environment in which each benchmark was run. This will include:

* Client and server system configurations, including node and processor counts, processor models, memory size and speed, and OS (names and versions)
* Storage used for global file system and storage configurations
* Network fabric used to connect servers, clients, and storage, including network configuration settings
* Client and server configuration settings. These should include:
	+ Client and server sysctl settings
	+ Driver options
	+ Network interface options
	+ File system configuration options
* Compiler name and version, compiler options, and libraries used to build benchmarks

If a single system configuration was used for multiple runs, create a system description and note which runs used that configuration. If benchmarks cannot be run on a system of the size proposed, the Offeror may provide a projection of performance at this size. Please indicate the size at which the benchmark was performed, provide any projection algorithm(s), and explain the calculations used to reach the full system size as well as the rationale for the projection model.

### Optional Runs

Additional runs may be done if the Offeror feels that the default runs do not adequately describe their offering. This includes runs utilizing a hints file and file system specific options to IOR. Scripts for these runs must be provided back to UCAR along with full documentation of any changes or optimizations made to the code. Please create a unique directory for each additional run. This directory should contain the modified scripts, source code for the run, output files and a description of the system the run was made on. A unique name for the output file should be chosen.

### IOR

The intent of these benchmarks is to measure system performance for file read and write operations that access disk. Although optimizations may be possible that enable significant caching or buffering of the transferred data within system memory, and although such optimizations may be beneficial in a production environment, Offerors are not permitted to engage in such optimizations for the benchmark runs. Consequently, the run rules are designed to defeat memory caching effects so as to measure the throughput of the underlying disk subsystem.

**Required Runs**:

Three scripts are provided to run the normal NCAR benchmark runs with IOR. Only the edits called out in the installation and run instructions may be made to these scripts. Please retain the output file names for these required runs.

### MDTEST

The intent of this benchmark is to measure system performance for file metadata operations that access disk. Because it is impractical to fully capture and represent within a benchmark the actual data sizes for I/O operations carried out by the users, the benchmarks utilize smaller working set sizes; however, since disk I/O remains the rate-limiting portion of the I/O workload, the benchmarks must be run in such a way as to perform I/O to disk and not to a DRAM cache. Although optimizations may be possible that enable significant caching or buffering of the transferred data within system memory, and although such optimizations may be beneficial in a production environment, Offerors are not permitted to engage in such optimizations for the benchmark runs.

**Required Runs**:

Four scripts are provided to run the normal NCAR benchmark runs with mdtest. Only the edits called out in the installation and run instructions may be made to these scripts. Please retain the output file names for these required runs. Unlike the first three runs, the fourth run requires that the proposed file system be 80% full. There is an additional IOR script provided to assist in creating the necessary files. If the Offeror chooses to do this in another manner, the script used to create the data must be provided.

### pyReshaper

The intent of this benchmark is to measure system performance of random small block I/O similar to a real data processing application. Although optimizations may be possible that enable significant caching or buffering of the transferred data within system memory, and although such optimizations may be beneficial in a production environment, Offerors are not permitted to engage in such optimizations for the benchmark runs.

**Required Runs:**

1. serial, single node, one task
2. parallel, single node, maxTaskperNode == as many tasks as it takes to saturate off node bandwidth
3. parallel, 4 nodes, maxTaskperNode == the number of cores per node

Note that the generated scripts are configured to run with the LSF resource manager and scheduler. The Offeror will need to modify this script for the appropriate batch system and must provide the modified script with returned results.