Benchmark Rules

NWSC-3: NCAR’s Next-Generation High-Performance Computing and Storage System

*Released 2 April 2020
Amendment #1, 27 April 2020
University Center for Atmospheric Research
UCAR RFP000074*

Table of Contents

[1. Objective of NWSC-3 Benchmarking 3](#_Toc38634031)

[2. Ownership and Dissemination of Results 3](#_Toc38634032)

[3. UCAR Right to Request Benchmark Rerun Prior to Award 3](#_Toc38634033)

[4. NWSC-3 Benchmark Instructions and Rules 3](#_Toc38634034)

[4.1 General Rules and Reporting 3](#_Toc38634035)

[4.2 Core and Memory Subscription 5](#_Toc38634036)

[4.3 Benchmark System(s) 6](#_Toc38634037)

[4.4 As-is and Optimized Benchmark Results 6](#_Toc38634038)

[4.5 Benchmark Results Projection 7](#_Toc38634039)

[4.6 Benchmark Code Modifications and Optimization 7](#_Toc38634040)

[5. Reporting of NWSC-3 Benchmark Results 8](#_Toc38634041)

[5.1 NCAR Benchmark Suite: Homogeneous Node Benchmarks 9](#_Toc38634042)

[5.1.1 CLUBB 10](#_Toc38634043)

[5.1.2 DART\_WRF 10](#_Toc38634044)

[5.1.3 MG2 10](#_Toc38634045)

[5.1.4 WACCM 10](#_Toc38634046)

[5.1.5 WRF 11](#_Toc38634047)

[5.2 NCAR Benchmark Suite: Heterogeneous Node Benchmarks 11](#_Toc38634048)

[5.2.1 GOES 11](#_Toc38634049)

[5.2.2 MPAS-A 12](#_Toc38634050)

[5.3 NCAR Benchmark Suite: Micro-benchmarks 12](#_Toc38634051)

[5.3.1 OSU MPI 13](#_Toc38634052)

[5.3.2 STREAM 13](#_Toc38634053)

[6. References 13](#_Toc38634054)

# Objective of NWSC-3 Benchmarking

The purpose of NWSC-3 benchmarking is to evaluate the performance of NCAR applications on the High-Performance Computing (HPC) system proposed by the Offeror in response to the NWSC-3 RFP (UCAR RFP000074). The benchmark applications are constructed to evaluate proposed technologies and to represent the computational demands of the scientific simulation, data analysis, and Machine Learning (ML) codes that are representative of current and anticipated research to be conducted at the NCAR Wyoming Supercomputing Center (NWSC). This document provides specific guidance to Offerors regarding how the NWSC-3 benchmarks are to be run, what information should be returned to UCAR, and other reporting requirements.

# Ownership and Dissemination of Results

NWSC-3 benchmark results include but are not 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 Prior to Award

UCAR reserves the right to request that the Offeror rerun any or all NWSC-3 benchmarks before a final award is made. Specifically, UCAR may ask an Offeror to rerun all or a portion of the benchmark suite should questions arise from the results provided in the Offeror’s response to this RFP. The Offeror may also be expected to run selected portions of the NWSC-3 benchmark applications during a Live Test Demonstration (LTD), should it be requested by UCAR, to corroborate actual or projected benchmark results provided in its response to this RFP.

# NWSC-3 Benchmark Instructions and Rules

This section provides information for obtaining NWSC-3 benchmark codes, their build and execution instructions, and the general rules for running and reporting the results.

## 4.1 General Rules and Reporting

The benchmarks used for the NWSC-3 procurement are available from the NCAR HPC Benchmarks website [1]. The site provides information on the benchmarks, including how to access them, the instructions for building and running them, and their numerical validation criteria. (Offerors should note that, while the NCAR HPC Benchmarks website [1] is used for the NWSC-3 procurement, it is independent of the procurement and will be maintained in the future as a source for NCAR-related benchmark codes).

Guidance regarding which benchmark results and output files are to be returned to UCAR in compressed tar files is provided in §5 and its subsections below. Additionally, the Offeror should obtain the benchmark results Excel spreadsheet (**Attachment 2A Benchmark Results Spreadsheet**) from the NWSC-3 RFP website [2] to record a subset of the benchmark performance results. The Offeror must use the spreadsheet to report the performance results of a subset of the NWSC-3 HPC benchmarks to UCAR with the Offeror’s proposal. The measured performance of those select computational benchmarks on NCAR’s Cheyenne system [3] is provided in the spreadsheet and, when input with the Offeror’s benchmark results, the spreadsheet will calculate the Cheyenne Sustained Equivalent Performance (CSEP) metric.

The CSEP metric produced by the spreadsheet provides a first-order approximation of the NWSC-3 HPC system’s Sustained Performance utilizing “as-is” and, optionally, “optimized” benchmark results. Offerors should note that, while only some of the NWSC-3 benchmark timings are to be entered in the spreadsheet, this does not diminish the importance to UCAR of the remaining benchmarks and their results.

Table 1 provides a list of worksheets contained in the Benchmark Results spreadsheet and a brief description of their use and purpose. The Offeror is cautioned to not change any cells other than those into which information is to be entered (i.e., “enter value”). The formulas provided in the spreadsheet may not be changed by the Offeror.

**Table 1.** Description of Worksheets in the Benchmark Results Spreadsheet

| Worksheet | Description/Use |
| --- | --- |
| CSEP Benchmark As-Is | Used for reporting the Offeror’s **as-is** benchmark timings. The **as-is** timing results for the selected benchmarks should be entered in column B of this worksheet.The number of Homogeneous Nodes, number of cores per Homogeneous Node, number of Heterogeneous Nodes, and number of accelerator coprocessors per Heterogeneous Node for the Offeror’s benchmark or proposed system may be entered in cells M8, N8, O8, and P8, respectively. |
| CSEP Benchmark Optimized | (Optional) Used for reporting the Offeror’s **optimized** benchmark timings. The **optimized** timing results for the selected benchmarks should be entered in column B of this worksheet.The number of Homogeneous Nodes, number of cores per Homogeneous Node, number of Heterogeneous Nodes, and number of accelerator coprocessors per Heterogeneous Node for the Offeror’s benchmark or proposed system may be entered in cells M8, N8, O8, and P8, respectively. |
| CSEP Proposed System | Used for calculating the CSEP of the proposed system. The Offeror may use this worksheet to enter **as-is**, **optimized**, or **projected** benchmark timing results for the proposed NWSC-3 HPC system.The number of Homogeneous Nodes, number of cores per Homogeneous Node, number of Heterogeneous Nodes, and number of accelerator coprocessors per Heterogeneous Node for the Offeror’s proposed system should be entered in cells M8, N8, O8, and P8, respectively. |
| Cheyenne | Contains the benchmark performance metrics from NCAR’s Cheyenne system [3]. |
| 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.If multiple benchmark systems are used, the Offeror should replicate the Benchmark System worksheet and provide the requested information for each benchmark system. |

## 4.2 Core and Memory Subscription

All Homogeneous Node 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 576-core benchmark run is requested, and the Offeror’s system is based on 92-core nodes, then the benchmark must be run on seven nodes, using all 92 cores on six of the nodes, and 24 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 576-core benchmark configuration is requested and requires 2 GB (gigabytes) of memory per core, and the Offeror’s system consists of 92-core nodes each with 128 GB of memory (~1.39 GB per core), then it will require nine nodes to run the benchmark. In this case, the benchmark may be run using 64 cores on each of the nine nodes.

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

1. Provide an 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 CSEP to account for the unutilized cores, if necessary.

## 4.3 Benchmark System(s)

The system(s) on which the Offeror runs the NWSC-3 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 rationale for doing so and explain how it is of benefit to UCAR. The Offeror should replicate the “**Benchmark System**” worksheet and provide the requested information for each system used by the Offeror to run the NWSC-3 benchmarks.

## 4.4 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 §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.

## 4.5 Benchmark Results Projection

The reporting of only 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) and provide projected benchmark timing results in the Benchmark Results spreadsheet to calculate the proposed NWSC-3 system’s CSEP metric, the Technical Volume of the Offeror’s proposal shall 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). This scaling model description should be sufficiently detailed to be both understandable and convincing.

## 4.6 Benchmark Code Modifications 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-3 benchmarks.
2. No assembly-level recoding is permitted.
3. Source code changes are preferably written in the original source language, 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 NWSC3\_*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-3 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 NWSC-3 Benchmark Results

Offerors proposing an NWSC-3 HPC system shall run the benchmarks as described in the instructions provided in the Globus ‘NCAR HPC Benchmarks’ collection. Access to that Globus collection is described on the NCAR HPC Benchmarks website [1]. The following sections set forth what information and files are to be returned to UCAR with the Technical Volume of the Offeror’s proposal.

The NWSC-3 Benchmarks are comprised of five codes for the Homogeneous Nodes, two codes for the Heterogeneous Nodes[[1]](#footnote-1), and two micro-benchmarks, as summarized in Table 2.

**Table 2.** Brief description of the NWSC-3 benchmarks

|  |  |  |
| --- | --- | --- |
| Benchmark | Type | Description |
| CLUBB | Homogeneous | The Cloud Layers Unified by Binormals (CLUBB) benchmark represents the calculations performed in the parameterization of clouds and turbulence in the CESM-2 model, which typically consumes approximately one-sixth of the total computational cost of running CESM-2. It is representative of the type of physics calculations performed by numerous NCAR applications. |
| DART\_WRF | Homogeneous | This benchmark represents the state-update calculation from the Data Assimilation Research Testbed (DART), which typically consumes approximately two-thirds of the total execution time of DART-WRF. |
| MG2 | Homogeneous | This benchmark represents the Morrison Gettleman version 2 (MG2) microphysics code that is used in the CESM-2 model, which typically consumes approximately one-twelfth of the total computational cost of running CESM-2. It is representative of the type of physics calculations performed by numerous NCAR applications. |
| WACCM | Homogeneous | This benchmark represents the chemical solver from the Whole Atmosphere Climate Community Model (WACCM), it contains a large number of multiply-add instructions with very little, to no, cache reuse and thus is a measure of the total number of linear system solves that can be calculated per second on a single node. |
| WRF | Homogeneous | This benchmark is based on version 4.1 of the Weather Research and Forecasting Model (WRF) and is used in NWSC-3, in part, to evaluate a system’s scalability in the range of typical production-scale codes at NCAR. |
| GOES | Heterogeneous | The GOES benchmark times the training of a deep convolutional neural network to predict the probability of lightning using imagery from the GOES-16 infrared geostationary satellite. |
| MPAS-A | Heterogeneous | This benchmark is an idealized atmospheric case of the Model for Prediction Across Scales (MPAS). While MPAS-A runs on both CPUs and accelerator coprocessors, it is used in NWSC-3 as an accelerator coprocessor benchmark. |
| OSU MPI | micro | This benchmark is a subset the Ohio State University (OSU) MPI benchmark and is used in NWSC-3 to measure the performance of basic MPI communications: point-to-point bandwidth and latency, one-sided put bandwidth and latency, collective allreduce, and collective broadcast. |
| STREAM | micro | This benchmark measures the sustainable bandwidth of the memory subsystem. |

The NCAR HPC Benchmarks website [1] provides the source code (compressed tar files) and instructions (PDFs) for building and running each of the NWSC-3 benchmarks. This section provides the specific instructions for reporting benchmark results to UCAR in the Technical Volume of the Offeror’s proposal. In all cases, compressed tar files should be used to return any requested files to UCAR.

## 5.1 NCAR Benchmark Suite: Homogeneous Node Benchmarks

Five benchmarks are used to measure the computational capacity of the Homogeneous Nodes for the computational component of the NWSC-3 HPC system. The Offeror shall run these benchmarks in as-is and, optionally, optimized configurations and return the results and requested benchmark output files with the Offeror’s proposal. The Offeror shall also include the requested timer data in the appropriate worksheets of the NWSC-3 Benchmark Results spreadsheet. If the Offeror’s benchmark system is technologically different from the proposed NWSC-3 system, the Offeror should also return benchmark performance projections for the proposed NWSC-3 system and a description of the projection methodology as requested in §4.5.

### 5.1.1 CLUBB

CLUBB should be run on a single node, using all available cores, and using one MPI rank for each of the available cores. For each CLUBB result submitted to UCAR, the **standard error** and **standard output** should be saved to a file and returned.

[A1] The number of processor cores on the benchmark system’s node and the ~~The~~ following timer data should be entered into the **Benchmark Results** spreadsheet:

Average columns per sec value from standard out

Example command: grep "Average columns" clubb.stdout | awk '{print $6}'

###

### 5.1.2 DART\_WRF

DART\_WRF should be run on a single node, using all available cores, and using one MPI rank for each of the available cores. For each DART\_WRF result submitted to UCAR, the **standard error** and **standard output** should be saved to a file and returned.

[A1] The number of processor cores on the benchmark system’s node and the ~~The~~ following timer data should be entered into the **Benchmark Results** spreadsheet:

Average state updates value from standard out

Example command: grep "Average state" dart\_wrf.stdout | awk '{print $6}'

###

### 5.1.3 MG2

MG2 should be run on a single node, using all available cores, and using one MPI rank for each of the available cores. For each MG2 result submitted to UCAR, the **standard error** and **standard output** should be saved to a file and returned.

[A1] The number of processor cores on the benchmark system’s node and the ~~The~~ following timer data should be entered into the **Benchmark Results** spreadsheet:

Average columns per sec value from standard out

Example command: grep "Average columns" mg2.stdout | awk '{print $6}'

### 5.1.4 WACCM

WACCM should be run on a single node, using all available cores, and using one MPI rank for each of the available cores. For each WACCM result submitted to UCAR, the **standard error** and **standard output** should be saved to a file and returned.

[A1] The number of processor cores on the benchmark system’s node and the ~~The~~ following timer data should be entered into the **Benchmark Results** spreadsheet:

Average System solves per sec value from standard out

Example command: grep "Average System" waccm.stdout | awk '{print $6}'

### 5.1.5 WRF

The WRF benchmark should be run using 576, 1728, 3456, 6912, and 11520 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** and the output of the validation and timing script. The output files **wrfout\_d01\_2019-05-05\_22:00:00** and **wrfout\_d01\_2019-05-05\_22:06:00** should be returned for one of the above core counts, but need not be returned separately for all core-counts.

The following timer data, for runs at each of the above core counts, should be entered into the **Benchmark Results** spreadsheet:

The total step time from the output of the validation and timing script.

Example command: grep "Total Time" validation.out | awk '{print $3}'

###

## 5.2 NCAR Benchmark Suite: Heterogeneous Node Benchmarks

Two Heterogeneous Node benchmarks are used in NWSC-3 to measure the computational capacity of the Heterogeneous Node portion of the computational component of the NWSC-3 HPC system. The Offeror shall run these benchmarks in as-is and, optionally, optimized configurations and return the results and requested benchmark output files with the Offeror’s proposal. The Offeror shall also include the requested timer data in the appropriate worksheets of the NWSC-3 Benchmark Results spreadsheet. If the Offeror’s benchmark system is technologically different from the proposed NWSC-3 system, the Offeror should also return benchmark performance projections for the proposed NWSC-3 system and a description of the projection methodology as requested in §4.5.

### 5.2.1 GOES

The GOES benchmark should be run on a single, accelerator coprocessor-equipped node. Results should be returned using a single coprocessor, and using all coprocessors in the node. For example, on a node with eight coprocessors, results should be returned for 1 and 8 coprocessors. If, by using a subset of the node’s coprocessors, superior performance results are observed, those results should also be returned. For each run of the GOES benchmark, **standard error** and **standard output** should be captured and returned to UCAR, along with the contents of the **output** directory.

The following timer data for the best-performing number of coprocessors should be entered into the **Benchmark Results** spreadsheet:

Timing information for the [A1] “Epoch” ~~“Elapsed”~~ timer in standard out

Example commands: grep "INFO:root:Elapsed:" goes.out | awk '{print $2}'

[A1] In addition, the number of accelerator coprocessors used in obtaining the best performance should be entered into the **Benchmark Results** spreadsheet.

### 5.2.2 MPAS-A

For MPAS-A, the 15km case should be run on accelerator coprocessor-equipped nodes, utilizing the minimum number of nodes necessary to run with 24, 48, and 96 coprocessors. The 30km case should be run using two coprocessors on a single node, and using all available coprocessors on a single node. For each run of the benchmark, the **standard error** and **standard output** should be saved and returned along with the file **log.atmosphere.0000.out** for each run.

The following timer data for the 15km case, for each of the above coprocessor counts, should be entered into the **Benchmark Results** spreadsheet:

[A1] The “total” column for t~~T~~imer "time integration" from file "log.atmosphere.0000.out"

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

[A1] Please note that, along with Amendment #1 to the NWSC-3 RFP, the 30km MPAS-A namelist.atmosphere file has been updated to increase the number of timesteps in the benchmark run to 200. Offerors should download the updated **MPAS\_2020-04-27\_data.tar.gz** file from the NCAR HPC Benchmarks Globus repository.

[A1] The following timer data for the 30km case, for the two coprocessor and all coprocessor single-node runs, should be entered into the **Benchmark Results** spreadsheet:

The “total” column for timer "time integration" from file "log.atmosphere.0000.out"

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

[A1] In addition, the number of accelerator coprocessors on the benchmark system node used for the 30km single-node, all-coprocessor case should be entered into the **Benchmark Results** spreadsheet.

## 5.3 NCAR Benchmark Suite: Micro-benchmarks

Offerors proposing an NWSC-3 HPC system shall run the micro-benchmarks as described in the instructions provided on the NCAR HPC Benchmarks website [1]. The following sections describe what information and files are to be returned to UCAR with the Technical Volume of the Offeror’s proposal. Offerors should note that the micro-benchmark results are not reported in the **Benchmark Results** spreadsheet, but their results are important to UCAR in understanding the target system’s inter-node communications and memory bandwidth characteristics; thus, the output files should be returned as requested.

### 5.3.1 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 error** and **standard output** from each run.

### 5.3.2 STREAM

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

1. Powers of two threads (i.e. 1, 2, 4, 8, ...) up to the number of physical cores on a single socket
2. One thread per physical core in a single compute node
3. One thread per computational unit (hyperthread, etc.) in a single compute node
4. Any other configuration that showcases the unique capabilities of a vendor’s platform; for example, if the solution includes a large on-die, high-bandwidth memory, include a run which uses only that memory space

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

# 6. References

[1] NCAR HPC Benchmarks Website: [*https://www2.cisl.ucar.edu/hpc\_benchmarking*](https://www2.cisl.ucar.edu/hpc_benchmarking)

[2] NWSC-3 RFP Website: [*https://www2.cisl.ucar.edu/nwsc-3*](https://www2.cisl.ucar.edu/nwsc-3)

[3] NCAR’s Cheyenne System: [*https://www2.cisl.ucar.edu/resources/computational-systems/cheyenne/cheyenne*](https://www2.cisl.ucar.edu/resources/computational-systems/cheyenne/cheyenne)

1. See Article 1, Definitions, of the Terms and Conditions document, UCAR RFP000074 Attachment 4 Terms and Conditions, for a complete description of Homogeneous and Heterogeneous Nodes. [↑](#footnote-ref-1)