Communications regarding NWSC-3 Benchmarks

Revision History:
Version 1.2: 31 Mar 2020
Version 1.1: 04 Sep 2019

Table of Contents

1 Overview 1
2 Conventions 1
2.1 Example brief description of question 2
3 Questions & Answers regarding NWSC-3 Benchmarks 2
3.1 Cold-start initial conditions for the WRF benchmark 2
3.2 Time steps in WRF benchmark 2
3.3 More detailed instructions for running WRF benchmark 3
3.4 Weights for time steps in WRF benchmark 3
3.5 Is WRF 4.1 a hard requirement? 3
3.6 Open source and directive-based methods for WRF 3
3.7 Proprietary GPU based WRF 4
3.8 Source code and run script modifications 4
3.9 Weight of radiation 5
3.10 CPU-only scaling results regarding GOES benchmark 7
3.11 Where to find MPAS-A code 7
3.12 Questions regarding the STREAM benchmark instructions 7

1 Overview

This document contains prospective Offeror questions related to the benchmarking application suite used for NWSC-3, and UCAR's responses to those questions.

2 Conventions

Each question and its corresponding response is formatted as shown below, providing a unique question identifier and a brief title for the question, the question itself, and UCAR's response to the question.

Example:

2.1 Example brief description of question

Question: The text of the Respondent's question will appear here. It may be stated verbatim or modified slightly to remove any irrelevant attributes of the question or any indication of the Offeror's identity.

UCAR's response to the question immediately follows.

3 Questions & Answers regarding NWSC-3 Benchmarks

The following questions were received by UCAR between 31 July 2019 and 31 March 2020.

3.1 Cold-start initial conditions for the WRF benchmark

Question: Previous WRF benchmarks, including the WRFv3 benchmark http://www2.mmm.ucar.edu/wrf/WG2/benchv3 and benchmarks released by NCAR as part of earlier RFPs, provided initial conditions in the form of restarts from runs that had been spun up over a period of hours. Please provide insight regarding why there are now cold-start initial conditions for the WRF benchmark in NWSC-3. Can unspun-up moisture physics from a cold start adequately represent the computational profile of an actual WRF run?

Yes. For the NWSC-3 benchmarks, we have made an effort to reduce data set sizes and the complexity of building, running, and validating the benchmarks. The primary purpose of the WRF benchmark is not to reproduce exact timings for production WRF simulations on future hardware. Rather, the primary purpose is the intercomparison of proposed systems and we feel this benchmark achieves that adequately despite trade-offs made to reduce complexity. In the context of the larger NWSC-3 suite, we are mainly interested in the scalability of WRF over small to medium scales.

3.2 Time steps in WRF benchmark

Question: Since the new benchmark is a more complete simulation, should there be more time steps so that this can best represent the steady-state profile?

No. Again we are satisfied that the shorter run time will still provide useful data, and should ease the burden of making multiple tuning runs to maximize the performance of the benchmark.

3.3 More detailed instructions for running WRF benchmark

Question: Can you provide more detailed instructions for how the benchmarks are to be run and what materials are to be returned, including what you're looking for in terms of performance and scalability and whether non-representative components for the timings (e.g. the timing for the first time step (which includes initialization and I/O) will be evaluated?

Additional information will be provided in an RFP attachment when it is released. The first WRF time step with initialization and I/O is not timed as a portion of the benchmark result. The rsl.out.0000 files that contain this timing information will be required to be returned as well.

3.4 Weights for time steps in WRF benchmark

Question: Currently, the WRF benchmark validation script reports two metrics:

1. *Timing for radiation step (in seconds)*
2. *Average Time for non-radiation (in seconds)*

Which weights will metric 1 and metric 2 have in overall scoring of new WRF benchmark?

The weight for the radiation step will be ¼ and the weight for the average time for non-radiation steps will be ¾.

3.5 Is WRF 4.1 a hard requirement?

Question: Is WRF 4.1 a hard requirement in the benchmark, or is there some flexibility in the WRF release?

- *NV-WRFg is based on 3.7.1 (OpenACC)*
- *TQI development we named WRFg is based on 3.8.1 (CUDA + OpenACC)*

Yes, the WRF 4.1 benchmark is a hard requirement. If vendors choose to return results for an equivalent simulation based on a different version of WRF, we will review it as additional information. However, the WRF 4.1 must be run to satisfy the benchmarking requirements.

3.6 Open source and directive-based methods for WRF

Question: Are there restrictions listed in the following two points for the WRF code used in the benchmark?

- *Is open source required? Or is proprietary code as objects/executables permitted?*

Open source is preferred. Any proprietary code provided as objects/executables would only be considered if the ability for UCAR to compile identical objects and executables was provided to UCAR following the award of a contract.

- *Can the code contain platform-specific language such as CUDA?*

Directive-based methods (OpenACC/OpenMP) are preferred. Any CUDA code provided should maintain the ability to compile the standard non-CUDA WRF from the same

source by including added CUDA code in appropriate #ifdef sections, or separate files swappable at build time. For more information, see the benchmarking rules included in the RFP attachment, when it is released.

3.7 Proprietary GPU based WRF

Question: For NWSC3 production, under what conditions could a proprietary GPU-based WRF be allowed?

UCAR is not only an end-user of the WRF model but also the developer of the WRF code base. For staff resource reasons, UCAR cannot assume the burden of supporting a CUDA-based WRF, even if the required software were made available to build and run it. At the same time, UCAR cannot prevent users from obtaining and running this code on its system, if they have a third-party support agreement from the vendor.

3.8 Source code and run script modifications

Question: Which modifications in the source code or run script are acceptable for benchmark submission? For example, are compiler SIMD directives (e.g., $dir simd or $omp simd) acceptable?

4.5, as described here:

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 below, 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-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.

5. 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.
6. 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.

3.9 Weight of radiation

Question: By analysis of several NWP workloads, we found that new workload has extremely high weight of radiation which seems to be uncommon for hi-res weather forecast. I provide comparison of NWSC-3 WRF configuration with configuration from HRRR at NCEP (same 3km resolution) and with industry standard CONUS workloads. The table below shows differences in physical parameterization for 4 workloads:

	NWSC-3: North America 3km (2019)	*CONUS-2.5km (2008)*	*CONUS-12km (2008)*	*HRRR 3km NCEP*
mp_physics	*8 (Thompson)*	*4 (WSM5)*	*4 (WSM5)*	*28 (Thompson?)*
cu_physics	*0 (none)*	*1 (Kain-Fritsch)*	*1 (Kain-Fritsch)*	*0 (none)*
ra_lw_physics	*4 (RRTMG)*	*1 (RRTM)*	*1 (RRTM)*	*4 (RRTMG)*

ra_sw_physics	*4 (RRTMG)*	*1 (RRTM)*	*1 (RRTM)*	*4 (RRTMG)*
bl_pbl_phyiscs	*2 (Mellor-Yamada-Janjic)*	*1 (YSU)*	*1 (YSU)*	*5 (MYNN)*
sf_sfclay_physics	*2 (Eta similarity)*	*1 (MM5 similarity)*	*1 (MM5 similarity)*	*5*
sf_surface_physics	*2 (Noah)*	*2 (Noah)*	*1 (5-layer thermal diffusion)*	*3*
radt (in minutes)	*3*	*10*	*30*	*15*
time_step (in seconds)	*18*	*15*	*72*	*20*
radt / time_step	*4**	*40*	*25*	*45*

** determined separately in performance metric by weighting radiation and non-radiation timestep.*

This motivates to raise the following question: Could you shed some light why radiation frequency has been significantly (order of magnitude!) increased, even for high-resolution forecast? For instance, same resolution HRRR calls radiation at every 45 timesteps. Further information regarding radiation time step is highly appreciated.

A good rule of thumb for NWP is to call the radiation calculation about every 10 model time steps. This frequency is a compromise. As with other physics schemes, ideally the radiation would be called at each time step. However, radiation is so expensive that the computed tendency is kept constant for a given number of time steps. To get a good physical representation of the evolution of the clouds (specifically how those clouds interact with the radiation scheme), and to not be exorbitantly expensive, calling the radiation about every 10 time steps is what the WRF group recommends.

For this benchmark, the run-time configuration sets the radt = 3 minutes and the time_step = 18 seconds. This yields a ratio of radt / time_step = 10 (i.e., run the radiation every 10 model time steps).

This WRF benchmark is set up to run a very short number of time steps. Excluding the initial step (which is not measured since it includes I/O and other one-time start-up processing), there are 18 non-radiation time steps and a single radiation time step. The script associated with the WRF benchmark reports both the single radiation time and the average non-radiation time, as the scaling from each will behave differently.

For an operational group (NCEP running the HRRR model), additional external constraints are imposed on the numerical model. It is entirely understandable that due

to time pressure NCEP extracts every bit of performance possible from their operational setting.

3.10 CPU-only scaling results regarding GOES benchmark

Question: What is the possibility for providing CPU-only scaling results for the future RFP on the GOES benchmark. While a single CPU-only server would provide lower performance compared to use of a CPU server with GPUs on this benchmark, we'd like the option of providing measurements looking at GOES performance across use of multiple nodes. TensorFlow can provide efficient scale out with CPUs (we've run TensorFlow with customers utilizing very high node counts at high scaling efficiencies). Can you accommodate the possibility of providing CPU-only scale-out measurements for this benchmark.

If a vendor can demonstrate better performance per dollar by running GOES on CPU nodes then we would certainly like to see those numbers. We will accept multi-node CPU results for GOES as an option. However, GOES results on GPU are still mandatory.

3.11 Where to find MPAS-A code

Question: When trying to build and run NCAR benchmark MPAS code, we did not find the benchmark data set described in the MPAS guide. Where I did not find any directory named: benchmark_15k, or benchmark_30k?

The benchmarking files are available through the Globus NCAR HPC Benchmarks collection, in the file MPAS_2019-08-09_data.tar.gz.

After unzipping the file, the benchmarking directories are labeled as:

- MPAS-A_benchmark_15km_L56
- MPAS-A_benchmark_30km_L56
- MPAS-A_benchmark_120km_L56
- validation_files_120km

3.12 Questions regarding the STREAM benchmark instructions

Question 3.12.1: The instructions refer to "One thread per physical core in a single compute node." Does this mean a single run using all physical cores on a node?

Yes.

Question 3.12.2: The instructions refer to "One thread per computational unit (hyperthread, etc.) in a single compute node." Does this mean a single run using all physical and logical cores on a node?

Yes.