Geophysical Research Abstracts
Vol. 19, EGU2017-14615-4, 2017
EGU General Assembly 2017

Toward an in-situ analytics and diagnostics framework for earth system models

Valentine Anantharaj (1), Matthew Wolf (1), Philip Rasch (2), Scott Klasky (1), Dean Williams (3), Rob Jacob (4), Po-Lun Ma (2), Kwo-Sen Kuo (5,6)

(1) Oak Ridge National Laboratory, National Center for Computational Sciences, Oak Ridge, United States (anantharajvg@ornl.gov), (2) Pacific Northwest National Laboratory, USA, (3) Lawrence Livermore National Laboratory, USA, (4) Argonne National Laboratory, USA, (5) University of Maryland, USA, (6) NASA Goddard Space Flight Center, USA

The development roadmaps for many earth system models (ESM) aim for a globally cloud-resolving model targeting the pre-exascale and exascale systems of the future. The ESMs will also incorporate more complex physics, chemistry and biology – thereby vastly increasing the fidelity of the information content simulated by the model. We will then be faced with an unprecedented volume of simulation output that would need to be processed and analyzed concurrently in order to derive the valuable scientific results. We are already at this threshold with our current generation of ESMs at higher resolution simulations.

Currently, the nominal I/O throughput in the Community Earth System Model (CESM) via Parallel IO (PIO) library is around 100 MB/s. If we look at the high frequency I/O requirements, it would require an additional 1 GB / simulated hour, translating to roughly 4 mins wallclock / simulated-day => 24.33 wallclock hours / simulated-model-year => 1,752,000 core-hours of charge per simulated-model-year on the Titan supercomputer at the Oak Ridge Leadership Computing Facility. There is also a pending need for 3X more volume of simulation output . Meanwhile, many ESMs use instrument simulators to run forward models to compare model simulations against satellite and ground-based instruments, such as radars and radiometers. The CFMIP Observation Simulator Package (COSP) is used in CESM as well as the Accelerated Climate Model for Energy (ACME), one of the ESMs specifically targeting current and emerging leadership-class computing platforms These simulators can be computationally expensive, accounting for as much as 30% of the computational cost. Hence the data are often written to output files that are then used for offline calculations. Again, the I/O bottleneck becomes a limitation. Detection and attribution studies also use large volume of data for pattern recognition and feature extraction to analyze weather and climate phenomenon such as tropical cyclones, atmospheric rivers, blizzards, etc. It is evident that ESMs need an in-situ framework to decouple the diagnostics and analytics from the prognostics and physics computations of the models so that the diagnostic computations could be performed concurrently without limiting model throughput.

We are designing a science-driven online analytics framework for earth system models. Our approach is to adopt several data workflow technologies, such as the Adaptable IO System (ADIOS), being developed under the U.S. Exascale Computing Project (ECP) and integrate these to allow for extreme performance IO, in situ workflow integration, science-driven analytics and visualization all in a easy to use computational framework. This will allow science teams to write data 100-1000 times faster and seamlessly move from post processing the output for validation and verification purposes to performing these calculations in situ. We can easily and knowledgeably envision a near-term future where earth system models like ACME and CESM will have to address not only the challenges of the volume of data but also need to consider the velocity of the data. The earth system model of the future in the exascale era, as they incorporate more complex physics at higher resolutions, will be able to analyze more simulation content without having to compromise targeted model throughput.