Geophysical Research Abstracts
Vol. 17, EGU2015-10622, 2015
EGU General Assembly 2015
© Author(s) 2015. CC Attribution 3.0 License.

Accelerating climate simulation analytics via multilevel aggregation and synthesis

Valentine Anantharaj (1), Krishnaraj Ravindran (2), Raghul Gunasekaran (1), Sudharshan Vazhkudai (1), and Ali Butt (2)

(1) Oak Ridge National Laboratory, National Center for Computational Sciences, Oak Ridge TN, United States (anantharajvg@ornl.gov), (2) Virgnia Tech, Blacksburg VA, United States

A typical set of ultra high resolution (0.25 deg) climate simulation experiments produce over 50,000 files, ranging in sizes from 10^1 MB to 10^2 GB each – for a total volume of nearly 1 PB of data. The execution of the experiments will require over 100 Million CPU hours on the Titan supercomputer at the Oak Ridge Leadership Computing Facility (OLCF). The output from the simulations must then be archived, analyzed, distributed to the project partners in a timely manner. Meeting this challenge would require efficient movement of the data, staging the simulation output to a large and fast file system that provides high volume access to other computational systems used to analyze the data and synthesize results. But data movement is one of the most expensive and time consuming steps in the scientific workflow. It is be expedient to complete the diagnostics and analytics before the files are archived for long term storage. Nevertheless, it is often necessary to fetch the files from archive for further analysis.

We are implementing a solution to query, extract, index and summarize key statistical information from the individual CF-compliant netCDF files that are then stored for ready-access in a database. The contents of the database can be related back to the archived files from which they were extracted. The statistical information can be quickly aggregated to provide meaningful statistical summaries that could then be related to observations and/or other simulation results for synthesis and further inference. The scientific workflow at OLCF, augmented by expedited analytics capabilities, will allow the users of our systems to shorten the time required to derive meaningful and relevant science results. We will illustrate some of the timesaving benefits via a few typical use cases, based on recent large-scale simulation experiments using the Community Earth System Model (CESM) and the DOE Accelerated Climate Model for Energy (ACME).