Advanced Workflow and Data Management for the Energy Exascale Earth System Modeling Project

Dean Williams (1), Valentine Anantharaj (2), and Renata McCoy (1)

(1) Lawrence Livermore National Laboratory, Livermore CA, United States, (2) Oak Ridge National Laboratory, Oak Ridge TN, United States (anantharajvg@ornl.gov)

The Energy Exascale Earth System Modeling Project is a large multi-laboratory and multi-institution effort sponsored by the Earth System Modeling program in the US Department of Energy (DOE) Office of Biological and Environmental Research. This collaborative project involves over 100 scientists, software developers and management personnel from eight DOE national laboratories, several academic institutions and private industry partners. The main objective of the project is to develop and utilize the ultra-high resolution state-of-the-art earth system model, called the Energy Exascale Earth System Model (E3SM) to address the mission-specific climate change and energy research priorities of the United States and to optimize the use of DOE's next-generation computational facilities.

E3SM is designed and developed to exploit the modern computer hardware architectures and the associated investments in software development. The 10-year roadmap for E3SM spans three generations of leadership class computer architectures and four phases of model development and corresponding major simulation experiments, supported by the implementation of the E3SM Infrastructure that will maintain a disciplined software engineering structure and turnkey workflows for computational experiment design, execution, analysis of output and distribution of results. The first series of production simulation experiments are expected to generate nearly 2PB of simulation output while consuming over 1 billion core-hours of computing resources across the three major DOE Computing Facilities.

Much of the labor-intensive analysis and data management activities have been streamlined and integrated via an automated processflow in an end-to-end workflow. The E3SM Workbench provides the gateway to seamlessly access the backend services and workflow orchestration for (a) model diagnostics, data analysis and visualization; (b) data management via the Earth System Grid Federation (ESGF) and Globus services; and (c) provenance capture and management. We will describe and/or demonstrate the workflow components of the E3SM infrastructure and discuss some of the lessons learned, especially the challenges involved in orchestrating our simulations and workflow across major computing and data facilities.