Community Earth System Model Gateway

Pengxuan Zheng
Purdue University
pzheng@purdue.edu

Raminder Singh
Indiana University Bloomington
ramifnu@iu.edu

Christopher S. Thompson
Purdue University
thompscs@purdue.edu

1. INTRODUCTION

The Community Earth System Model (CESM) is a fully-coupled advanced global model for simulating the climate system on the Earth [1]. It is one of the most widely used climate simulation tools worldwide. CESM model consists of five separate components simulating the Earth's atmosphere, ocean, land, land-ice, and sea-ice systems. A central coupler component is responsible for inter-component data transformations and transfers. The CESM model is data and computationally intensive, requiring high performance computing (HPC) resources and significant data storage for large scale experiments. Because of the complexity of the model software, it is challenging and time-consuming for researchers to port, install, configure, and validate the full system as well as to archive the results. It presents a significant challenge for educators to train students in using the model, and for the broader research community to utilize the model in their research. To tackle the above challenges, we developed a science gateway that allows users to run CESM model simulations on the Purdue Carter cluster. Post process and visualize the results can be done via a web browser [2]. The gateway sets up a CESM model on the Purdue Carter cluster and uses a shared community account to submit jobs via Globus. It shields scientists from the many tasks involved in setting up and running CESM on the Purdue Carter cluster, e.g. compiling CESM on a specific computing system, validating the model installation, and acquiring individual accounts and time allocation on the Purdue Carter cluster. The gateway has been successfully used to support classes and research activities [3].

2. Features

The CESM portal is a web-based user interface for accessing CESM services. It allows users to create, configure and submit CESM cases to the Purdue Carter cluster, access the input and output data and publish results to the ESG (Earth System Grid). Users are also able to submit the data for post-processing and visualize the results directly in the browser. The system architecture of the CESM portal is shown in Figure 1. There are three layers. At the bottom are the computation and storage resources. The model currently runs on the Purdue Carter cluster. The output is stored in a scratch disk space after simulation and may be transferred into an iRODS data grid later [4]. There is a standard ESG data node running on a virtual machine at Purdue which has access to the model data in either the scratch disk or iRODS via a FUSE interface. It directly communicates with and publishes model results to the ESG gateway. The middle layer consists of a set of web service interfaces named CESM-WS. It provides standard interfaces for setting up and running CESM simulations, postprocessing output, transferring data, and publishing model data/metadata to ESG. CESM-WS is integrated with a token manager which supports authentication against the CESM gateway account database.

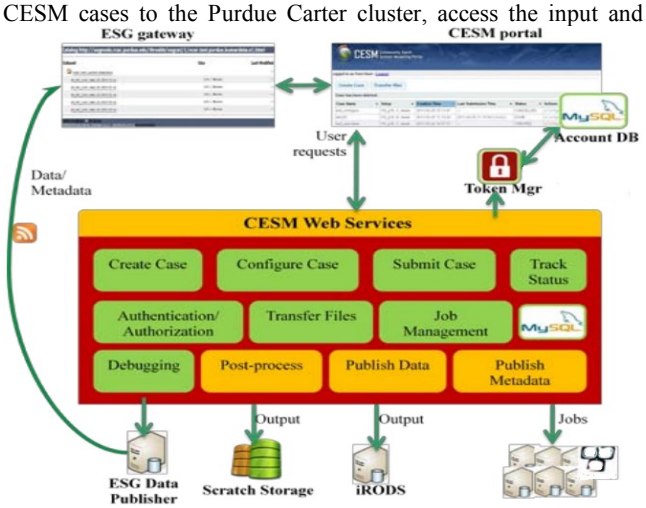

Figure 1. System Architecture.

3. Integration to Airavata

Apache Airavata is an open source science gateway framework [5] used to manage the complex execution for scientific applications on distributed computing resources. Integration is done using APIs for job submission, monitoring, canceling and data retrieval; listing & downloading intermediate results. For the CESM gateway, we are using Airavata to enable model execution on SDSC's Trestles and potentially other XSEDE resources. One of the goals is to remove dependency on Globus and explore other middleware options like GSISSH and UNICORE. CESM is using Airavata's Java APIs to provide wrapper for the "create case", "configure case" and "submit case" submission and monitoring steps. In the future, we will examine contributing CESM's advanced features like the data publisher developed to the Airavata open source community.

4. REFERENCES

[1] Community Earth System Model (CESM): http://www2.cesm.ucar.edu/.

[2] L. Zhao, C. Song, C. Thompson, H. Zhang, M. Lakshminarayanan, C. DeLuca, S. Murphy, K. Saint, D. Middleton, N. Wilhelmi, E. Nienhouse, and M. Burek. "Developing an integrated end-to-end teragrid climate modeling environment", TeraGrid 2011 conference, Salt Lake City, UT, July 2011.

[3] L. Zhao, W. Lee, C. X. Song, M. Huber and A. Goldner. "Bringing High Performance Climate Modeling into the Classroom", TeraGrid 2010 conference, Pittsburgh, PA, August 2010.

[4] iRODS: Data Grids, Digital Libraries, Persistent Archives, and Real-time Data Systems. http://www.irods.org/

[5] Apache Airavata: http://airavata.apache.org/.