ECSS Workflows – An Overview and Update

Mats Rynge
Information Sciences Institute /
University of Southern California
rynge@isi.edu

Marlon Pierce
Indiana University
marpierc@iu.edu

Suresh Marru
Indiana University
smarru@iu.edu

ABSTRACT

The focus of this poster is to provide an overview and update on the work of the XSEDE ECSS (Extended Collaborative Support Service) Workflow Community Applications Team. With the effort being less than a year old, it is doubtful that most of the larger XSEDE community knows that the effort exists and what support can be provided. With this poster we hope to introduce ourselves to XSEDE users, open up the workflows discussion, and provide ideas on how users can benefit from workflows and this ECSS effort.

Primarily, the goal of the workflow team is to increase the use of workflows on XSEDE. Our charter is:

"The XSEDE Workflow Community Applications Team's charter is to assist researchers to use scientific workflow technologies on XSEDE to solve challenging scientific problems involving parameter sweeps, multiple applications combined in dependency chains, tightly coupled applications, and similar execution patterns that require multiple applications and multiple XSEDE resources. The workflow team accomplishes its mission through the use of third party workflow software in collaboration with the workflow developers, service providers and XSEDE Extended Collaborative Support Services."

It has been surprisingly difficult to identify users with workflow needs. To find users, the team has done a fair amount of outreach to different groups within XSEDE, such as presenting to Campus Champions, Gateways and ECSS. The team has examined startup and XRAC allocations for potential users. Most users have been found by referrals from existing XSEDE users and staff.

A keystone in our approach is to use existing tools. Even though the team is made up of developers of several workflow systems, the team is workflow agnostic. The goal is not to develop workflow systems, but to help users use the existing ones. Users are guided when it comes to selecting the appropriate workflow system for their problem, and then the team helps with the integration with the user's existing codes, and XSEDE resources.

We believe that in order to make the workflows effort sustainable, it is not enough to help individual users, but what is needed is a to build a community. This is similar to other successful XSEDE efforts such as Gateways and Campus Champions. To build the community, general discussion and information mechanisms, such as mailing lists and XSEDE wiki space, have been set up. At the SC'13 conference, a workflows BOF was hosted to highlight some of the existing workflows on XSEDE, and to bring people in the XSEDE community and the workflow community together. This summer, we will host a seminar series introducing different workflow systems.

The poster will also include highlights of some of the current workflow efforts:

- SoyKB, a genomics pipeline which is developed in collaboration with iPlant. The goal is to provide both a specific solution to the SoyKB requirements, as well as a general reusable pipeline example which can be used by future iPlant and XSEDE users.

- A workflow to collect, sort, and merge large amounts of earthquake simulations data. The workflow in this case is to help scale up an existing solution the users has.

- Integration of the BioKepler portal/workflow engine with XSEDE.

- Support for integration of collaborations with existing workflows, such as LIGO and USATLAS.

- Support for industry/XSEDE project