21st International Conference on Computing in High Energy and Nuclear Physics **CHEP2015** Okinawa Japan: April 13 – 17, 2015

Contribution ID: **418**   Type: **poster presentation**

Exploring Two Approaches for an End-to-End Scientific Analysis Workflow

The scientific discovery process can be advanced by the integration of
independently-developed programs run on disparate computing facilities
into coherent workflows usable by scientists who are not experts in
computing. For such advancement, we need a system which scientists can
use to formulate analysis workflows, to integrate new components to
these workflows, and to execute different components on resources that
are best suited to run those components. In addition, we need to monitor
the status of the workflow as components get scheduled and executed, and
to access the intermediate and final output for visual exploration and
analysis. Finally, it is important for scientists to be able to share
their workflows with collaborators.

We are involved with a project to develop such an analysis framework for
the Large Synoptic Survey Telescope (LSST) Dark Energy Science Collaboration (DESC).
Following upon the development of several detailed use cases for LSST DESC,
we have been working on two approaches for the framework; the first one is
based on the use and extension of Galaxy, a web-based portal for biomedical
research, and the second one is based on a programming language, Python.
There are benefits to each approach as we discovered while implementing one example use case.
Both approaches allow scientists to run complicated workflows that involve the use of a
variety of computational resources (including grid resources,
supercomputing resources at NERSC, and local compute nodes) for the
execution of workflows on simulations of LSST images.
Adding a new application in the Python-based workflow description is straight forward,
however,
adding new applications through the Galaxy interface requires expert knowledge of
the Galaxy system and interaction with Galaxy infrastructure.

In this paper, we present a brief description of the two approaches,
describe the kinds of extensions to the Galaxy system we have found necessary
in order to support the wide variety of scientific analysis in the
cosmology community, and discuss how similar efforts might be of benefit
to the HEP community.

Author: SEHRISH, Saba (urn:Google)

Co-authors: KOWALKOWSKI, Jim (Fermilab); Dr PATERNO, Marc (Fermilab)

Presenter: Dr PATERNO, Marc (Fermilab)

Track Classification: Track4: Middleware, software development and tools, experiment frameworks,
tools for distributed computing