**An Introduction to BigJob and BigData – Using Pilots to Manage Computation and Data on XSEDE**

***Instructors:***

Shantenu Jha, RADICAL, Rutgers University

Melissa Romanus, RADICAL, Rutgers University

SAGA-BigJob provides a framework for running both very large-scale parallel simulations and many small high-throughput across a variety of middleware. These applications also may utilize a variety of storage systems when staging data in and out. BigJob has been used for parameter sweeps, many instances of the same task (ensemble), chained tasks, loosely coupled but distinct tasks, as well as tasks with data or compute dependencies.

BigJob have seen their widest usage across the heterogeneous resources that XSEDE provides. Simple installation into user space on any resource that supports Python >2.6 makes the uptake of BigJob virtually seemless. BigJob supports thousand of jobs (millions of SUs) and has been at the heart of two recent and successful ECSS projects. It has been used by a wide range of application types -- ranging from Computational Chemistry applications (uncoupled ensembles) to loosely coupled applications.

BigJob is a reference implementation of the P\* Model of Pilot-Jobs. Pilot-Jobs support the decoupling of workload submission from resource assignment. BigJob and its data-management layer (BigData) address the fundamental challenges of co-placement and scheduling of data and compute in heterogeneous and distributed environments with interoperability and extensibility as first-order concerns.

We present a half-day introductory tutorial to using BigJob to effectively manage submission of computations and movement of data on XSEDE. Attendees will need to bring their laptops and wireless connectivity will be required.

**Content to be covered during presentation:**

[**http://saga-project.github.com/BigJob/sphinxdoc/index.html**](http://saga-project.github.com/BigJob/sphinxdoc/index.html)

Agenda

* Introduction and Overview to BigJob [1 hour]
* Hands-On Session [2 hours]
  + Installing BigJob
  + Writing your First BigJob Script
  + Running BigJob on XSEDE resources
  + Working with Data Staging
  + Advanced Workflows using BigJob