**HTCondor Week 2017 Pegasus Tutorial**

Workflows are a key technology for enabling complex scientific applications. They capture the interdependencies between processing steps in data analysis and simulation pipelines, as well as the mechanisms to execute those steps reliably and efficiently in a distributed computing environment. They also enable scientists to capture complex processes to promote sharing and reuse, and provide provenance information necessary for the verification of scientific results and scientific reproducibility.

In this tutorial, we will focus on how to model scientific analysis as a workflow that can be executed on the Open Science Grid using Pegasus WMS ([http://pegasus.isi.edu](https://urldefense.proofpoint.com/v2/url?u=http-3A__pegasus.isi.edu&d=CwMF-g&c=clK7kQUTWtAVEOVIgvi0NU5BOUHhpN0H8p7CSfnc_gI&r=BjHQfj2qI9kdVxqyrw7Fbw&m=8wA8tcnWk64Z84N3RnQXIUC7i0mCXe_mAQCtHeOPACk&s=oTLfkdQdIcRobMKRvxlaMGyoFVCE99U4Rxl1ehkXR4Q&e=)). Pegasus allows users to design workflows at a high-level of abstraction, which is independent of the resources available to execute them and the location of data and executables. It compiles these abstract workflows to executable workflows that can be deployed onto distributed resources such local campus clusters, computational clouds and grids such as XSEDE and Open Science Grid.  During the compilation process, Pegasus WMS does data discovery, whereby it determines the locations of input data files and executables.  Data transfer tasks are added to the executable workflow that are responsible for staging in the input files to the cluster, and the generated output files back to a user specified location. In addition to the data transfers tasks, data cleanup (cleanup data that is no longer required) and data registration tasks (catalog the output files) are be added to the pipeline.

Through hands-on exercises, we will cover issues of workflow composition, how to design a workflow in a portable way, workflow execution and how to run the workflow efficiently and reliably. An important component of the tutorial will be how to monitor, debug and analyze workflows using Pegasus-provided tools.

The participants will be expected to bring in their own laptops with the following software installed: SSH client, Web Browser, PDF reader. Participants will be provided training accounts to log on to a workflow submit node at ISI, from where they can do the hands on exercises.