Pegasus WMS: Enabling Large Scale Workflows on National Cyberinfrastructure

Karan Vahi[1]
vahi@isi.edu

Ewa Deelman[1]
deelman@isi.edu

Gideon Juve[1]
gideon@isi.edu

Mats Rynge[1]
rynge@isi.edu

Rajiv Mayani[1]
mayani@isi.edu

Rafael Ferreira da Silva[1]
rafsilva@isi.edu

[1] Information Sciences Institute, University of Southern California

ABSTRACT

Pegasus WMS is a Workflow Management System that can manage large-scale scientific workflows across desktops, campus clusters, grids and clouds. This poster will introduce the capabilities of managing these workflows on diverse national cyber-infrastructures like XSEDE, OSG (Open Science Grid), and FutureGrid in an efficient, reliable and automated fashion.

The different national cyber-infrastructures that have been developed over the past decade offer different styles of high-performance computing. Leadership class systems, such as many of the resources available through XSEDE, are optimized for highly parallel, tightly coupled applications. They provide scalable, shared filesystems like Lustre. On the other hand, collaborative systems like OSG cater to high throughput loosely coupled applications. Typically, the sites don't provide a shared filesystem and encourage a model where jobs access data from community storage. Finally, cloud-based resources such as FutureGrid, and Amazon can be customized to user's needs.

Pegasus WMS provides a means for representing the workflow of an application in an abstract form that is independent of the resources available to run it and the location of data and executables. It compiles these abstract workflows into an executable form by querying information catalogs. The executable workflows are deployed on local or remote distributed resources using the Condor DAGMan workflow engine.

Pegasus WMS optimizes workflow execution and data movement by leveraging existing Grid and Cloud technologies via a flexible pluggable interface. While executing jobs on XSEDE, Pegasus relies on the shared filesystem to place input data for the workflows, while on OSG and cloud environments it may send input data directly to the worker nodes using Condor File I/O or Amazon S3 object storage. The data movement capabilities of Pegasus allow users to do cross-site runs, where parts of the workflow can be executed on different resources, with Pegasus automatically taking care of the intermediate data movement. Pegasus also provides advanced features such as reusing existing data, automatic cleanup of generated data, task clustering and recursive hierarchal workflows. It also captures all the provenance of the workflow lifecycle from the planning stage, through execution, to the final output data, helping scientists to accurately measure the performance of their workflows and reconstruct the history of data products. Pegasus provides debugging and monitoring tools that allow users to easily track failures in their workflows by analyzing system logs.

Many XSEDE systems are tuned to run large tightly coupled MPI applications. However, these systems are also beneficial for large fine-grained workflows containing hundreds of thousands of serial jobs because of their computation and storage capabilities. In the past users have provisioned nodes from XSEDE resources using pilot job tools such as GlideinWMS. However, recent state-of-the-art systems such as Kraken have architectures that make the deployment of pilot jobs infeasible due to network limitations. To address this issue Pegasus ships with an MPI-based task management tool called *pegasus-mpi-cluster* that can be used to run partitions of large, loosely coupled workflows on such petascale resources. *pegasus-mpi-cluster* allows us execute the tasks in these partitions by considering core/memory requirements of the task, and tracking available cores/memory on the compute nodes. This approach is used by a majority of our users to run on large XSEDE systems such as Stampede and Kraken.

Pegasus WMS has been in development since 2001 and is used in production in a variety of scientific domains, including: astronomy, seismology, bioinformatics, and physics. Several scientific communities have used Pegasus to run large-scale workflows on XSEDE, namely SCEC Cybershake workflows, IPAC Galactic Plane workflows, LIGO IHOPE pipelines, and iPlant genome sequencing pipelines.

Acknowledgments: This work was supported by the National Science Foundation under grant #OCI-1148515.