New and Advanced Capabilities of Pegasus-WMS

Rafael Ferreira da Silva
Ensemble Manager

• Allow users to submit a collection of workflows (ensembles)
  • Automatically spawn and manage collections of workflows

• Properties
  • Workflows within an ensemble may have different priorities
    • Priorities can also be changed at runtime
  • Ensembles may limit the number of concurrent planned and running workflows

• Prototype feature: support for event triggers that can trigger addition of new workflows to an existing ensemble

• Additional Actions
  • Ensembles can be paused, resumed, removed, re-planned, and re-executed
  • A Debugging mechanism is also provided to investigate failures in workflow runs
  • Actions can be performed both to ensembles and single workflows within ensembles
Running Workflow Ensembles

Ensemble Manager can handle multiple ensembles of workflows independently.

Ensemble Manager

- **Ensemble 1**
  - Priority: 100
  - Priority: 100
  - workflow execution
    - max_running = 2

- **Ensemble 2**
  - Priority: 120
  - Priority: 80
  - workflow execution
    - max_running = 1

- **Ensemble 3**
  - Priority: 50
  - workflow execution
    - max_running = 1

concurrent planning and execution of multiple workflows
Pegasus Container Support

• Users can refer to containers in the Transformation Catalog with their executable preinstalled

• Users can refer to a container they want to use – Pegasus stage their executable to the node
  • Useful if you want to use a site recommended/standard container image.
  • Users are using generic image with executable staging.

• Future Plans
  • Users can specify an image buildfile for their jobs.
  • *Pegasus will build the Docker image as separate jobs in the executable workflow, export them as a tar file and ship them around* (planned for 4.9.X)
Data Management for Containers

• Users can refer to container images as
  • Docker or Singularity Hub URL’s
  • Docker Image exported as a TAR file and available at a server, just like any other input dataset

• We want to avoid hitting Docker/Singularity Hub repeatedly for large workflows
  • Extend `pegasus-transfer` to pull image from Docker Hub and then export it as tar file, that can be shipped around in the workflow.

• Ensure pegasus worker package gets installed at runtime inside the user container
Running Pegasus workflows with Jupyter

http://pegasus.isi.edu
Pegasus est. 2001

Automate, recover, and debug scientific computations.

Get Started

Pegasus Website
https://pegasus.isi.edu

Users Mailing List
pegasus-users@isi.edu

Support
pegasus-support@isi.edu

Pegasus Online Office Hours
https://pegasus.isi.edu/blog/online-pegasus-office-hours/

*Bi-monthly basis on second Friday of the month, where we address user questions and also apprise the community of new developments*