

# **ACCESS Pegasus**

# A hosted scientific workflow system, part of the ACCESS Support approach

Mats Rynge, USC/ISI



OpenCI Award: OCI 21-38296

#### **About Me**

- Computer scientist at USC Information Sciences Institute
- Workflow developer
- OSG Facilitator
- Formerly ECSS
- ... and now ACCESS Support







**Scientific Workflows ACCESS Support** Logging In **Setting Up Resources Designing Workflow Provisioning Resources Monitoring Workflow and Resources** 





### **Scientific Workflows**

- An abstraction to express ensemble of complex computational operations
  - Eg: retrieving data from remote storage services, executing applications, and transferring data products to designated storage sites
- A workflow is represented as a directed acyclic graph (DAG)
  - Nodes: tasks or jobs to be executed
  - Edges: depend between the tasks
- Have a monolithic application/experiment?
  - Find the inherent DAG structure in your application to convert into a workflow







## **Key Pegasus Concepts**

#### Pegasus WMS == Pegasus planner (mapper) + DAGMan workflow engine + HTCondor scheduler/broker

- Pegasus maps workflows to infrastructure
- DAGMan manages dependencies and reliability
- HTCondor is used as a broker to interface with different schedulers

#### Workflows are DAGs

- Nodes: jobs, edges: dependencies
- No while loops, no conditional branches
- Jobs are standalone executables
- Planning occurs ahead of execution

#### Planning converts an abstract workflow into a concrete, executable workflow

Planner is like a compiler









Innovation

Data Flow for **LIGO Pegasus Workflows** in OSG

#### Advanced LIGO

Laser Interferometer Gravitational Wave Observatory

60,000 Compute Tasks Input Data: 5000 files (10GB total) Output Data: 60,000 files (60GB total) Processed Data: 725 GB

> Executed on LIGO Data Grid, EGI, **Open Science Grid and XSEDE**





**PI: Dong Xu** 

Trupti Joshi, Saad Kahn, Yang Liu, Juexin Wang, Badu Valliyodan, Jiaojiao Wang







### **NSF ACCESS Program Structure**









#### **Researcher Support Services - MATCH**

#### https://support.access-ci.org/

Advancing

Innovation

۲

Shelley Knuth, Alana Romanella, Dylan Perkins (University of Colorado Boulder) Ewa Deelman, Mats Rynge (U. Southern California) John Goodhue, Julie Ma, Andrew Pasquale (MGHPCC) James Griffioen, Vikram Gazula, Tony Elam, Joel Adams (Univ. Kentucky) David Hudak, Alan Chalker (Ohio Supercomputer Center)

TIER MATCH-Premier Enable innovative research through **Å** LONG-TERM EMBEDDED SPECIALISTS equitable and scalable support TIER MATCH-Plus Four tiers of support 3 SHORT-TERM SUPPORT PARTNERSHIPS Tools, growing knowledge base TIER **Curated Knowledge Base** Match-making with experts 2 **COMMUNITY EXPERTS** Student engagement Engagement from community Easy to Use Tools TIER **ACCESS ONDEMAND & PEGASUS CSSN** incentives



### **ACCESS OnDemand Pilot**



ACCESS OnDemand (AOD) Goals:

Advancing

- 1. Improve productivity for researchers by integrating Open OnDemand with ACCESS services (e.g, authentication, support portal, and metrics)
- 2. Decrease effort for resource providers to install, configure, and update OnDemand through managed releases with accompanying integration roadmaps

Pilot Project

- Initial AOD pilot targeting current OnDemand deployments on ACCESS resources (e.g., Anvil, Bridges 2, Delta and Expanse)
- 2. Working group is forming! Please contact Dave Hudak (<u>dhudak@osc.edu</u>) for details.





# **Opportunities for Engagement**

- ACCESS CSSN Community Participation Program (CCEP)
  - Earn travel support rewards by contributing to the ACCESS Knowledge Base, joining committees, becoming a mentor
- Advisory committees forming now!
  - CCEP Committees
  - MATCH-Plus/MATCH-Premier Steering Committees
- Join our mailing list and apply for the CCEP!

https://support.access-ci.org/cssn

- Pegasus, OnDemand, MATCH-Plus and MATCH-Premier pilots – opportunities for researchers, resource providers, consultants, mentors and students!
- Listening tours we need your feedback!







#### https://support.access-ci.org/pegasus





### Prepare Logging In

CILogin with your ACCESS ID and institutional login

<u>https://access.pegasus.isi.edu</u>

All registered ACCESS users with an active allocation automatically have access







#### Prepare Setting Up Resources

One time setup

Use <u>Open Ondemand instances</u> at resource providers to install ssh keys, and determine local allocation id









### Step 1 Designing Workflow

Pegasus API in Jupyter Notebook

Fully hosted environment, based on Open Ondemand







Advancing

Innovation

#### Step 2 Provision Resources

Use the HTCondor Annex tool to dynamically bring in compute nodes from one or more resource providers





#### **HTCondor Annex / Pilot Jobs**

- A pilot can run multiple user jobs it stays active until no more user jobs are available or until end of life has been reached, whichever comes first.
- A pilot is partitionable job slots will dynamically be created based on the resource requirements in the user jobs. This means you can fit multiple user jobs on a compute node at the same time.
- A pilot will only run jobs for the user who started it.





#### *Step 3* **Monitoring Workflow and Resources**

Workflows can be monitored from within the Jupyter notebook, or via command line

HTCondor Annex can be monitored on the command line









**Documentation:** 

https://support.access-ci.org/pegasus

Open a ticket:

https://support.access-ci.org/open-a-ticket

#### **Questions?**



