Pegasus / Montage workflow on Amazon Web Services

Mats Rynge – rynge@isi.edu
USC Information Sciences Institute
Goal

1. Use Montage as an example application to introduce Pegasus and HTCondor …
2. Provide you with a good base image set up for workflows / pipelines …
3. Provide ideas and contact information …

… so that you have something to get started with!
Users submit their serial or parallel jobs to Condor, Condor places them into a queue, chooses when and where to run the jobs based upon a policy, carefully monitors their progress, and ultimately informs the user upon completion.

- Job queuing mechanism
- Scheduling policy
- Priority scheme
- Resource monitoring
- Resource management

HTCondor DAGMan example
Pegasus Workflow Management System

- Builds on top of HTCondor and DAGMan.

- Abstract Workflows - Pegasus input workflow description
  - Workflow “high-level language”
  - Only identifies the computation, devoid of resource descriptions, devoid of data locations

- Pegasus is a workflow planner/mapper (“compiler”)
  - Transforms the workflow for performance and reliability
  - Automatically locates physical locations for both workflow components and data
  - Collects runtime provenance
System Overview

![Diagram showing system components]

- **Intermediate Files**
- **Produced Dataset**
- Master
- Worker
- Worker
- Worker
- Worker
- Worker

Published original survey data hosted at IPAC

Caching / rate limiting Squid server at ISI

**EC2**

**S3**

**IPAC**

**USC Information Sciences Institute**
Image

- 20 GB EBS volume
- CentOS 6
- HTCondor 8.0.4
- Pegasus 4.3.0
- Montage 3.3
- If no userdata is given: master node
  If private IP address of master is given: worker node
Exercises

http://goo.gl/7YWAbu
Exercises

1. Start an instance
2. mDAG
3. generate-montage-replica-catalog
4. pegasus-plan
5. Workflow status – command line and dashboard
6. Wait for workflow / pegasus-statistics
7. Outputs – open the mosaic in your web browser
8. Adding more worker nodes
9. Terminate instances
Questions?

rynge@isi.edu

Pegasus - [http://pegasus.isi.edu/](http://pegasus.isi.edu/)
- NSF funded
- Open Source
- Documentation, tutorial, and support available on website