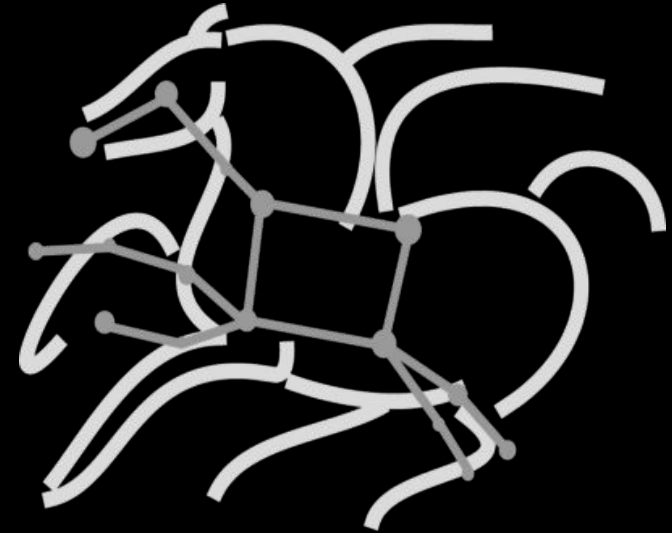

Pegasus Scientific Workflows on the Open Science Grid



Automate, recover, and debug scientific computations.

<https://pegasus.isi.edu>

Mats Rynge rynge@isi.edu

Outline

1. Quick introduction to Open Science Grid and OSG Connect
2. Use the “tutorial” command to get a sample workflow
3. Discuss site catalog and dax generator
4. Submit and monitor the workflow
5. Discuss the generated jobs
6. Containerized workflow

This is an interactive session. Please interrupt at anytime to ask questions.

Why Pegasus?

Automates complex, multi-stage processing pipelines

Enables parallel, distributed **computations**

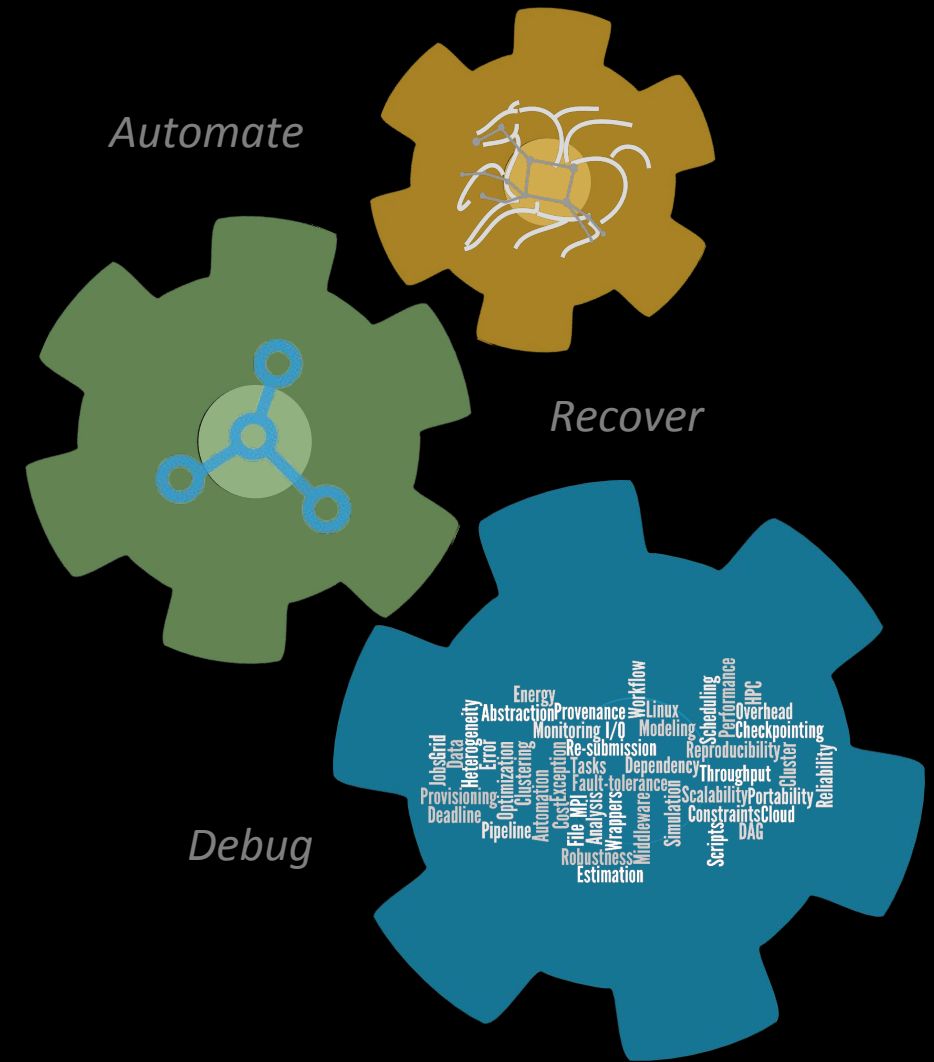
Automatically executes data transfers

Reusable, aids **reproducibility**

Records how data was produced (**provenance**)

Handles **failures** with to provide reliability

Keeps track of data and **files**



Open Science Grid

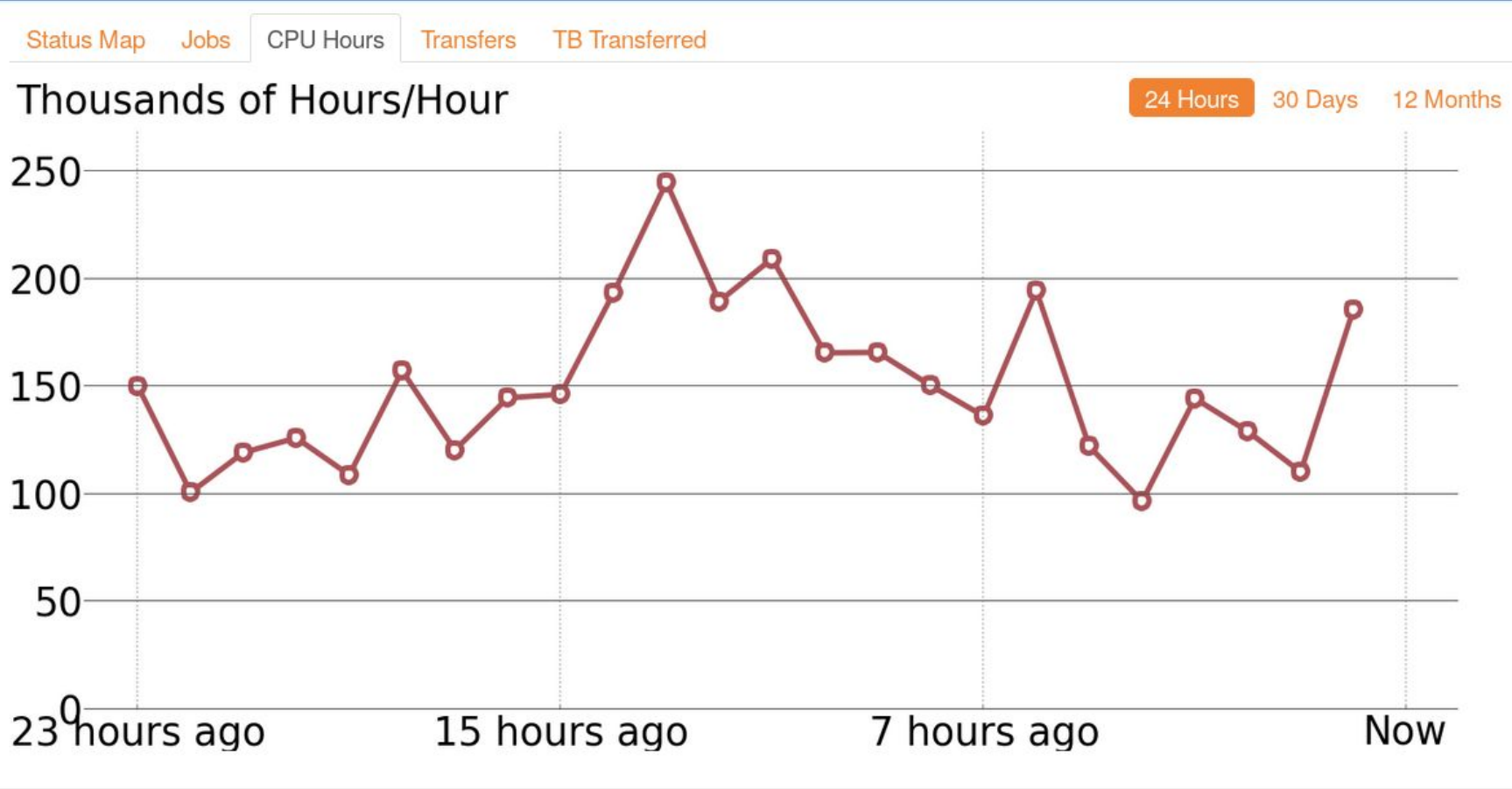
A **framework** for large scale distributed resource sharing addressing the technology, policy, and social requirements of sharing computing resources.

OSG is a **consortium** of software, service and resource providers and researchers, from universities, national laboratories and computing centers across the U.S., who together build and operate the OSG project. The project is funded by the NSF and DOE, and provides staff for managing various aspects of the OSG.

Integrates computing and storage resources from over 120 sites in the U.S.



Open Science Grid

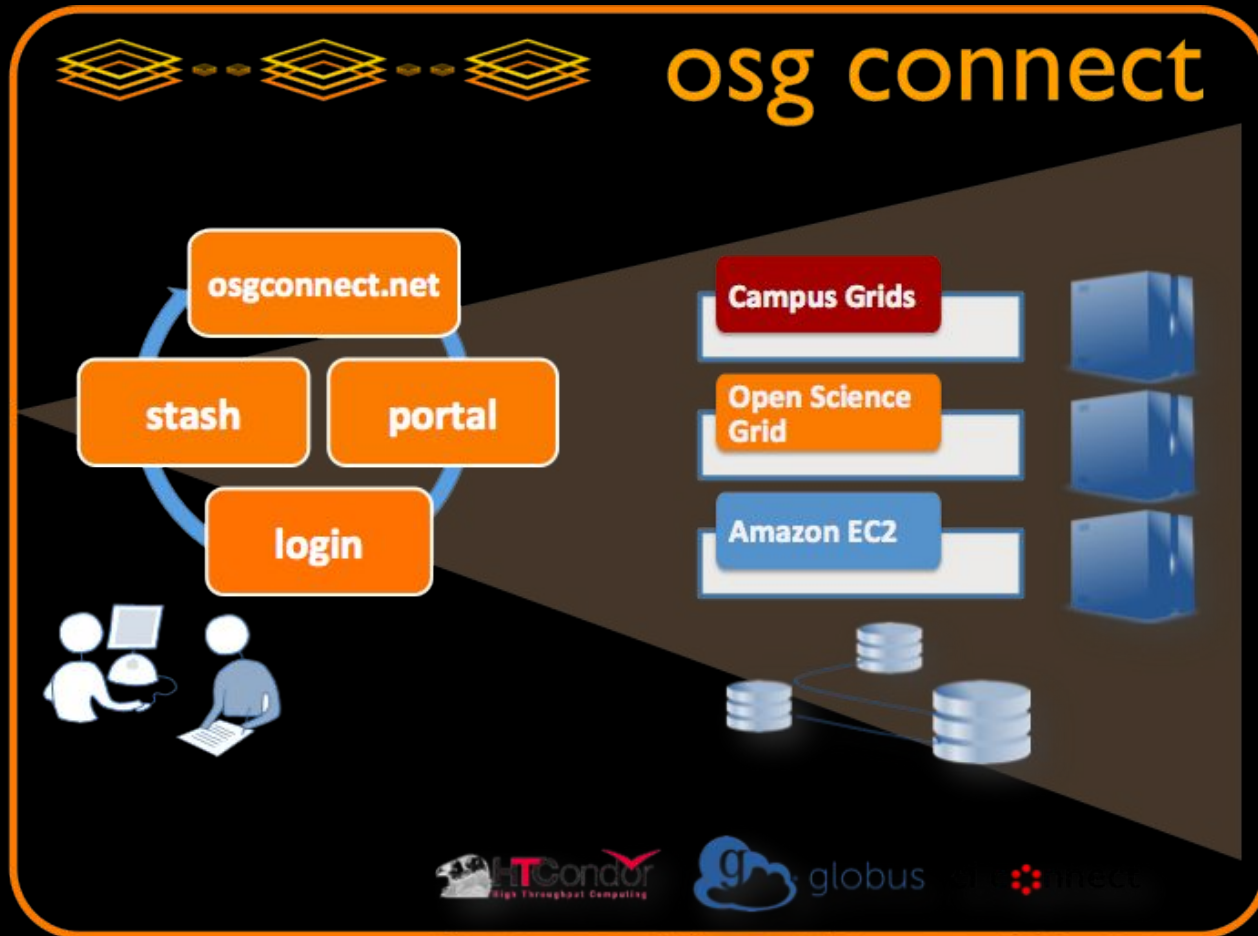


In the last 24 Hours	
278,000	Jobs
3,614,000	CPU Hours
13,859,000	Transfers
499	TB Transfers
In the last 30 Days	
8,586,000	Jobs
127,120,000	CPU Hours
0	Transfers
0	TB Transfers
In the last 12 Months	
121,418,000	Jobs
1,623,768,000	CPU Hours
681,704,000	Transfers
68,777	TB Transfers

OSG delivered across 118 sites

~ 3.5 million CPU hours delivered per day

OSG Connect Service



OSG Connect Provides:

- ★ Login host
- ★ Job scheduler
- ★ Software
- ★ Storage

<http://osgconnect.net/>

OSG User Support

Helpdesk:

<https://support.opensciencegrid.org>

- Knowledge Base
 - User guides/tutorials
 - HTC Recipes
- Forums
- “How do I...?” articles
- Interactive online chat

Support email:

user-support@opensciencegrid.org

The screenshot shows the OSG help desk interface. At the top, there's a navigation bar with 'Home', 'Solutions', and 'Forums'. The main heading is 'How can we help you today?' with a search bar and a 'SEARCH' button. To the right, there are links for 'New support ticket' and 'Check ticket status'. The main content area is divided into two columns. The left column is titled 'Knowledge base' and includes an 'Overview' section with links like 'Getting Started (4)' and 'OSG XSEDE Users (1)'. Below this is the 'OSG Connect User Guide' section with links for 'Getting Started with OSG Connect (4)' and 'Choosing Resources for jobs (4)'. The right column is titled 'Community forums' and includes sections for 'OSG Connect' with links like 'Announcements (3)' and 'Re-trying failed jobs - PeriodicRelease', and a 'How do I ... ? (2)' section with links like 'What are the available software on OSG' and 'How do I know if my jobs/applications can be run ...'.

opensciencegrid help desk

Welcome
Login Sign up

Home Solutions Forums

How can we help you today?

Enter your search term here... SEARCH

+ New support ticket
Check ticket status

Knowledge base

Overview

Getting Started (4) OSG XSEDE Users (1)

A welcome from the User Support Team
Communicate with us via Twitter
Is high throughput computing for you?
Acknowledging the Open Science Grid

OSG XSEDE User Guide

OSG Connect User Guide

Getting Started with OSG Connect (4) Choosing Resources for jobs (4)

Registration and Login for OSG Connect
OSG Connect Quickstart
Start a Project with OSG Connect
The "tutorial" command

Steer your jobs with HTCondor job requirements
Large Memory Jobs
GPU Jobs
Multicore Jobs

Running applications on OSG Connect (5) Transferring Data to OSG Connect (2)

Accessing Software using Distributed Environment Mo...
Software modules catalog
Requesting a software installation
Software transfer via HTCondor or HTTP
Compiling Applications for OSG Connect

Using scp to transfer files
Data transfer with Globus

Community forums

Showing recent updates
Start a new topic

OSG Connect

Announcements (3)

Re-trying failed jobs - PeriodicRelease
Posted by Bala Desinghu, about 1 month ago ,
Last Reply by Rob Gardner about 1 month ago

High Throughput Computing Examples
Posted by Bala Desinghu, 4 months ago

Working on ConnectBook tutorials
Posted by Rob Gardner, 5 months ago

? How do I ... ? (2)

What are the available software on OSG
Posted by Bala Desinghu, 4 months ago ,
Last Reply by Rob Gardner 4 months ago

How do I know if my jobs/applications can be run ...
Posted by Emelie Harstad, 5 months ago ,
Last Reply by Rob Gardner 5 months ago

Feature Requests (2)

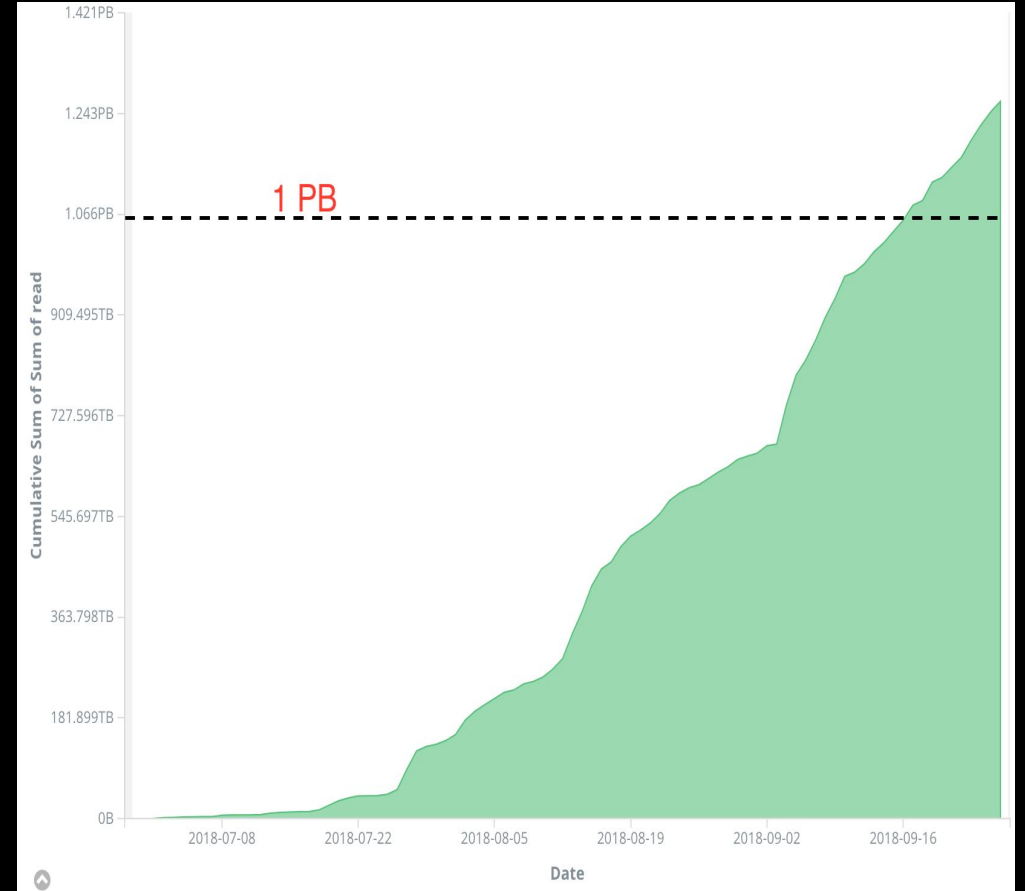
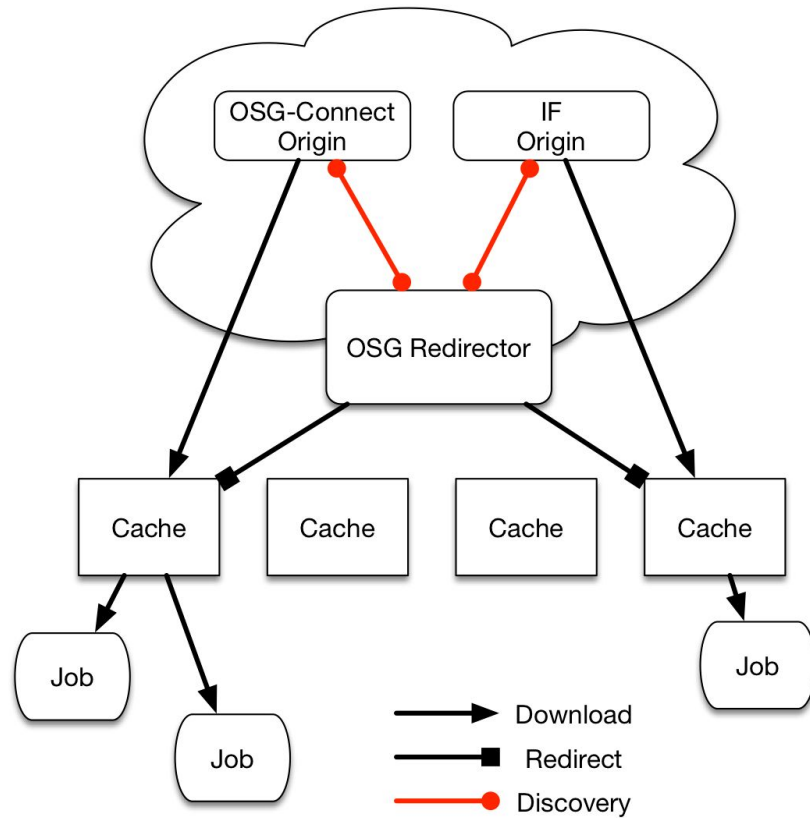
FreeSurfer on the Open Science Grid?
Posted by Rob Gardner, 3 months ago ,
Last Reply by Suchandra Thapa about 1 month ago

What applications are of interest to the medical co...
Posted by Rob Gardner, 4 months ago

Storage service: “Stash”

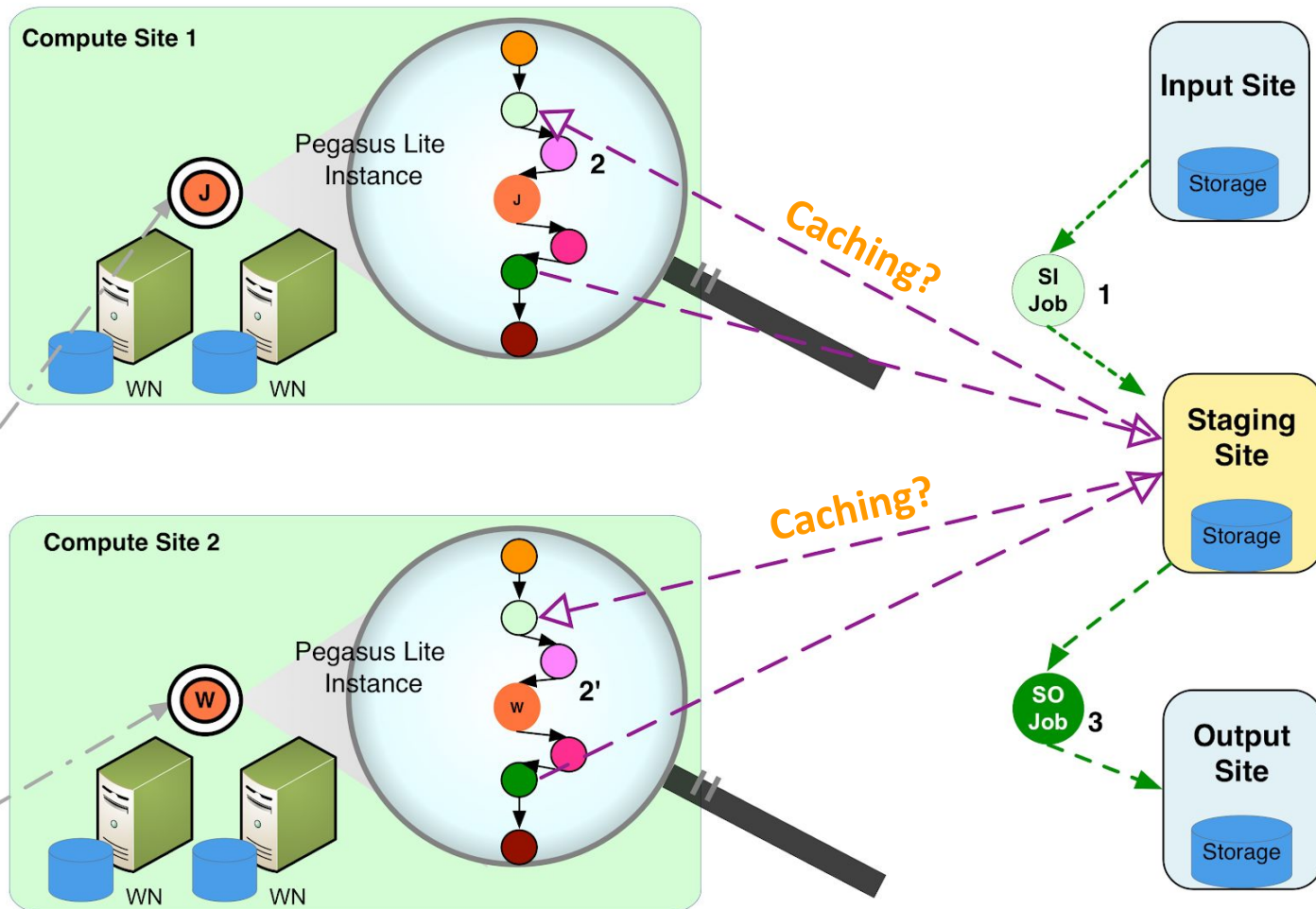
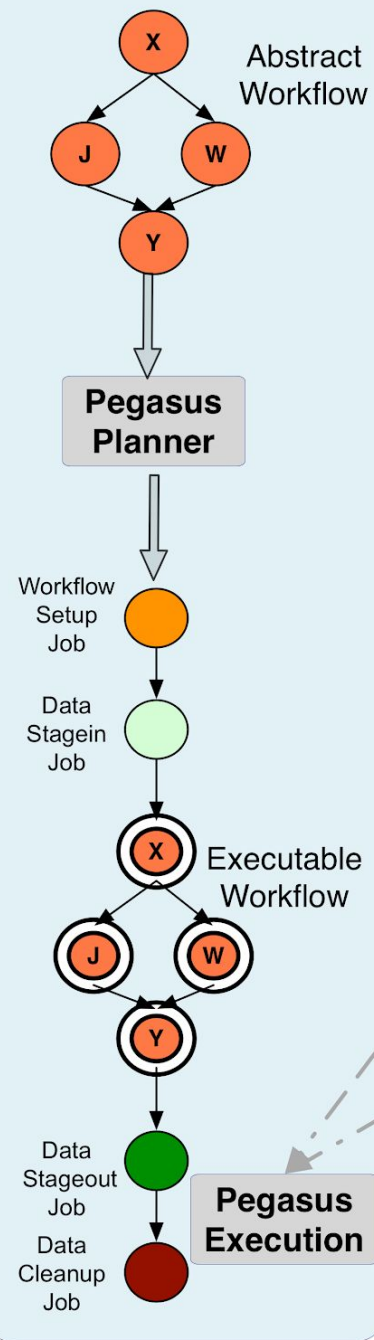
- Provide a quasi-transient storage service for job input/output data
- **POSIX** access provided to the login host
- **Globus Online** Server for managed transfers from campus data services
- Personalized **http** service endpoint
- Can now handle writes!
- Connected to 100 Gbps SciDMZ (I2, ESnet)

StashCache



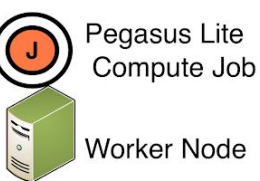
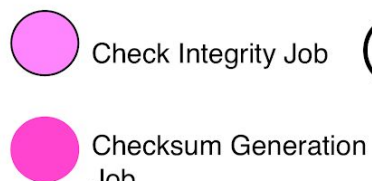
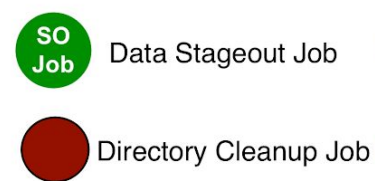
<https://derekweitzel.com/2018/09/26/stashcache-by-the-numbers/>

SUBMIT HOST



Stash

LEGEND



Tutorial

<https://support.opensciencegrid.org/support/solutions/articles/5000639789-pegasus>

```
$ tutorial pegasus
```

```
$ cd tutorial-pegasus
```