Pegasus Tutorial

Mats Rynge

Setup

It is the same (but hosted) as the self-guided tutorial available in the Pegasus documentation: https://pegasus.isi.edu/documentation/user-guide/tutorial.html

Please claim an instance by putting you name next to an unused instance in:
https://shorturl.at/fpsV5 (see Zoom chat for clickable link!)
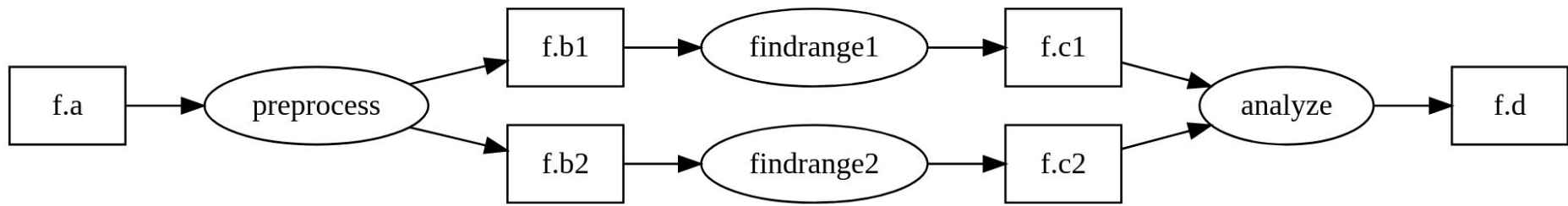
Follow the link next to your name.
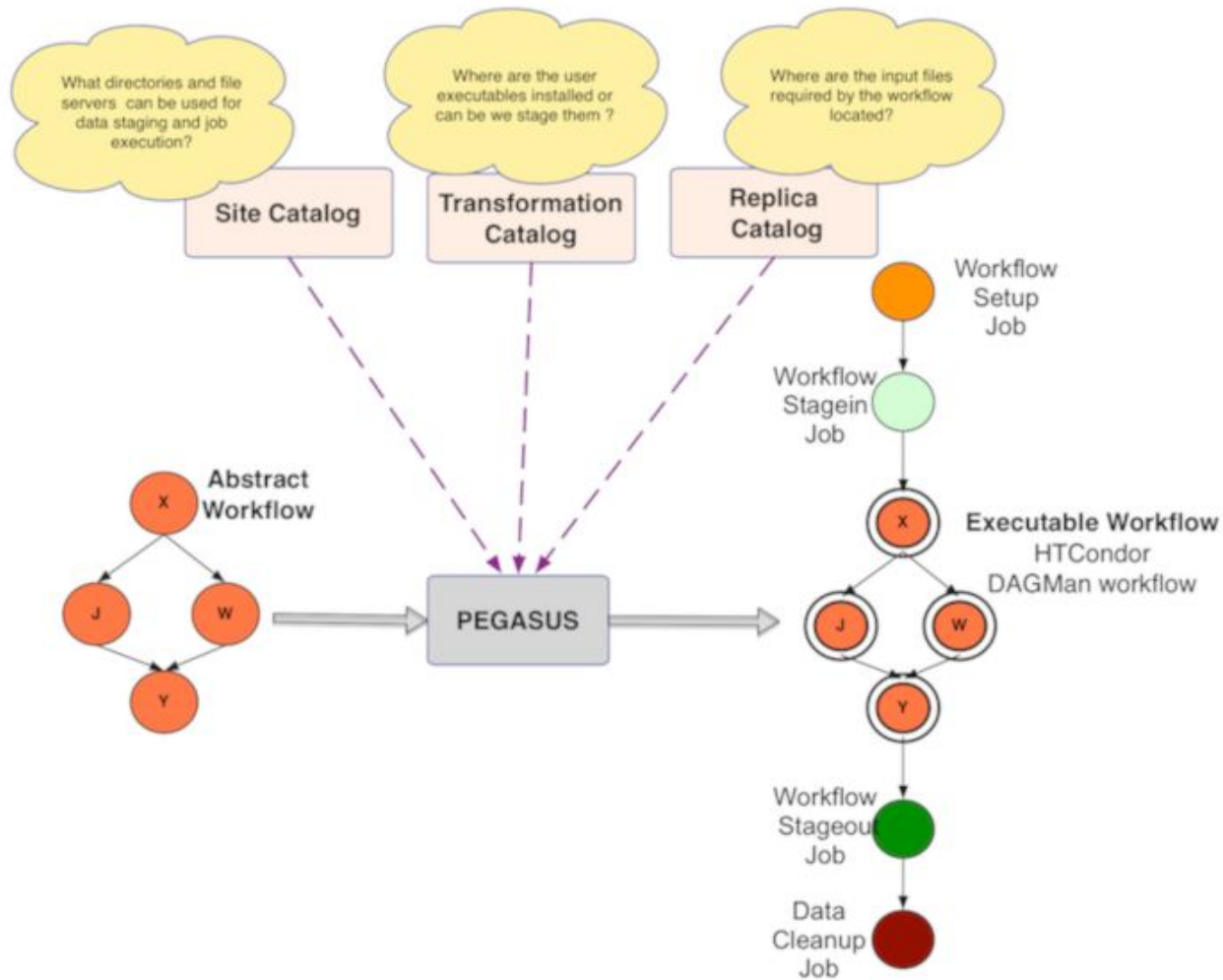
Docker Container / Jupyter Notebook

Container is for tutorial purposes - most production workflows have dedicated submit hosts

Jupyter is optional. You can choose to use just the workflow abstraction API, the full workflow management API, inside or outside Jupyter.

API

```
f.a → preprocess → f.b1 → findrange1 → f.c1 → analyze → f.d
                 → f.b2 → findrange2 → f.c2 →
```

What directories and file servers can be used for data staging and job execution?

Where are the user executables installed or can be we stage them?

Where are the input files required by the workflow located?

Site Catalog

Transformation Catalog

Replica Catalog

Abstract Workflow

X

J W

Y

PEGASUS

Workflow Setup Job

Workflow Stagein Job

Executable Workflow
HTCondor
DAGMan workflow

X

J W

Y

Workflow Stageout Job

Data Cleanup Job

Debugging

Command Line Tools

Summary

Thank you!