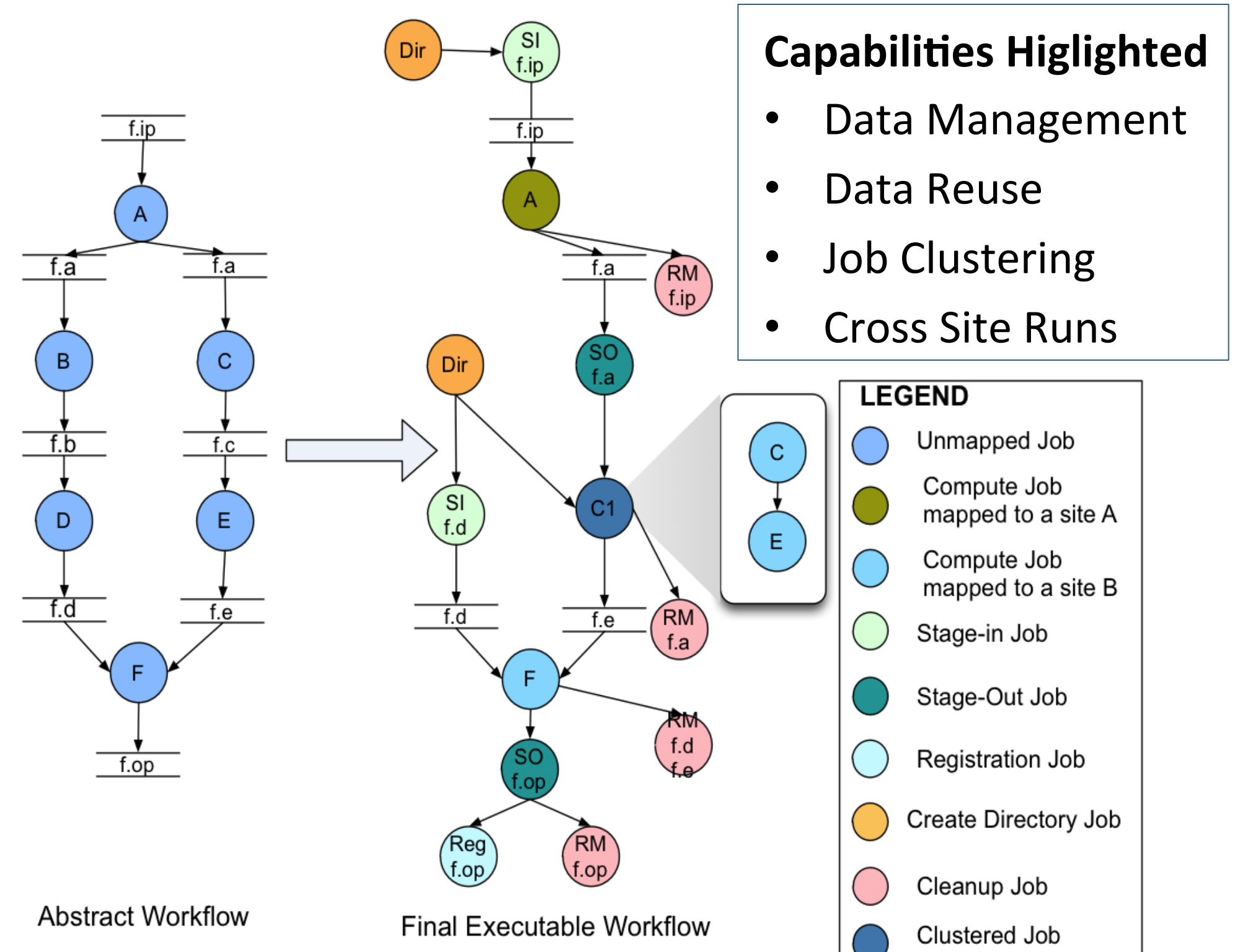


## Pegasus WMS

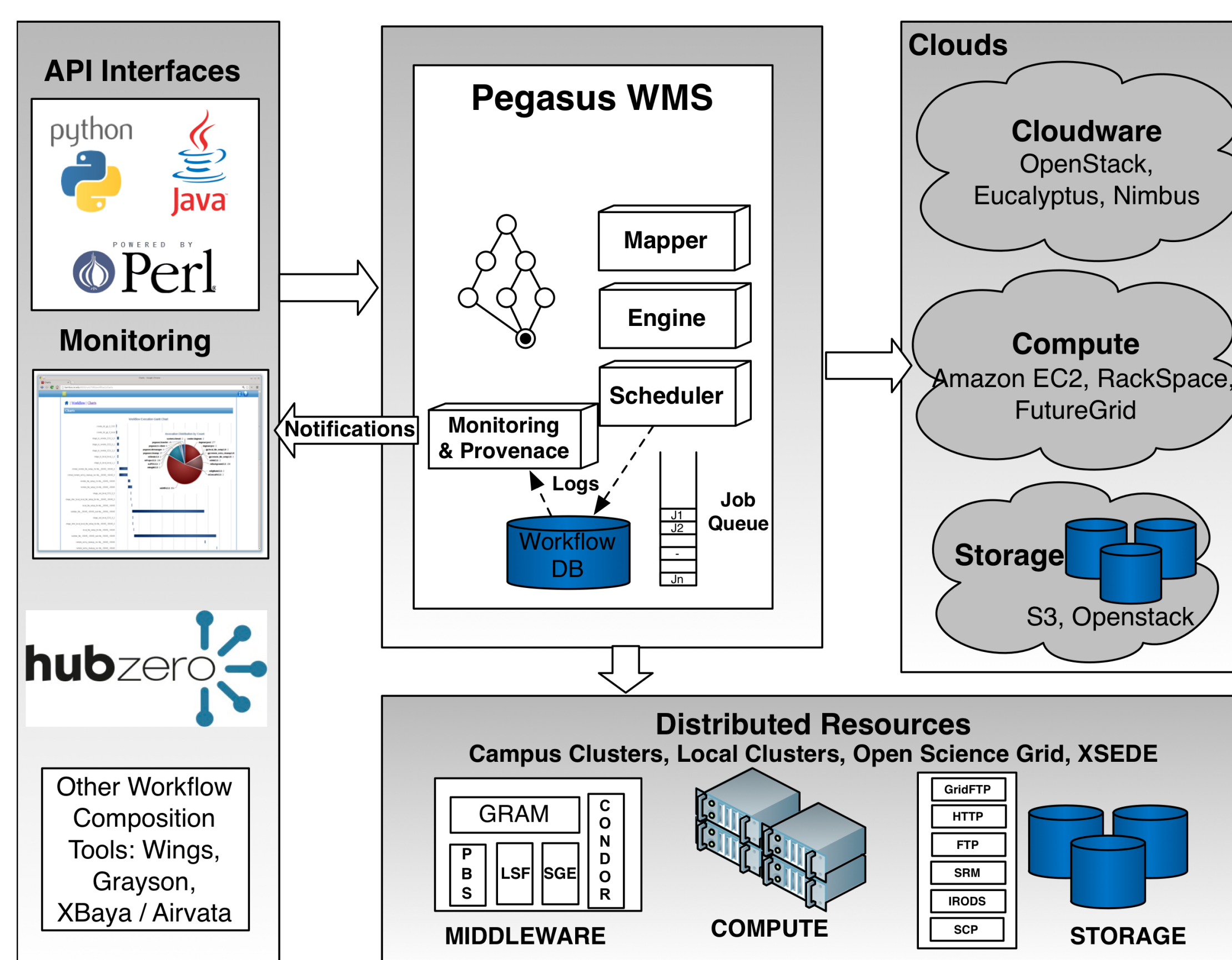
- Pegasus is a system for mapping and executing abstract application workflows over a range of execution environments.
- The same abstract workflow can, at different times, be mapped different execution environments such as XSEDE, OSG, commercial and academic clouds, campus grids, and clusters.
- Pegasus can easily scale both the size of the workflow, and the resources that the workflow is distributed over. Pegasus runs workflows ranging from just a few computational tasks up to 1 million.
- Pegasus Workflow Management System (WMS) consists of three main components: the Pegasus Mapper, HTCondor DAGMan, and the HTCondor Schedd.

## Canonical Workflow Example



## System Architecture and Features

- Optimized data transfers and support for many protocols
- Automatic data cleanup to reduce data footprint
- Retries computations in case of failures
- Workflow-level checkpointing through data reuse and DAGMan
- Command line and web-based monitoring and debugging tools to support large workflows
- workflow and task-level notifications
- Stores provenance of data used and produced, and executable invocations
- Pegasus-MPI-Cluster enables fine-grained task graphs to be executed efficiently on HPC resources



### Software Availability

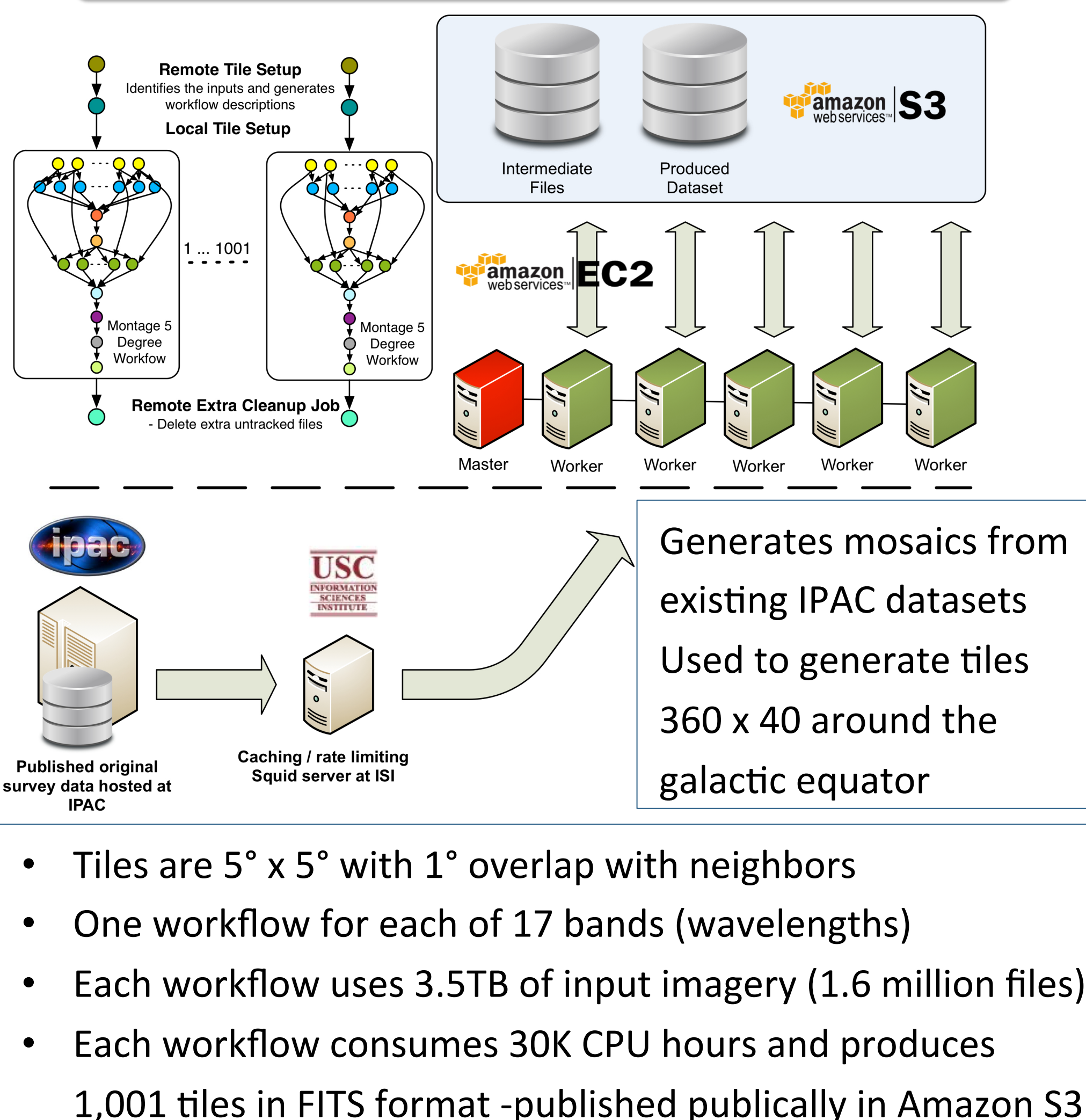
#### Download Options

- YUM repository with RPM packages
- APT repository with DEB packages
- Binary packages for Linux and Mac

#### Documentation / Training Materials

- User Guide
- Quickstart Guide
- Tutorial with Virtual Machine
- Software Carpentry Module

## Montage Galactic Plane



## Applications using Pegasus

### Astronomy and Physics:

- Galactic Plane workflow generates mosaics for astronomy surveys
- LIGO workflows help detect gravitational waves
- Periodogram workflows help detect extra solar planets

### Seismology:

- CyberShake workflows for seismic hazard analysis of LA basin
- Broadband workflows for accurate predictions of ground motions

### Bioinformatics:

- Quality control workflows for data submissions to NRGR repository and PAGE consortium.
- Imputation workflows on PAGE data
- Workflows for Genome and Transcriptome free analysis of RSEQ
- Brain span workflows help study gene expression in the brain
- RNA Sequencing workflows for generating Cancer Genome Atlas
- SIPHT workflows to predict sRNA encoding genes in bacteria
- Proteomics workflows for mass spectrometry based proteomics

### Others:

- <http://pegasus.isi.edu/applications>