

## Pegasus WMS

Pegasus is a system for mapping and executing abstract application workflows over a range of execution environments.

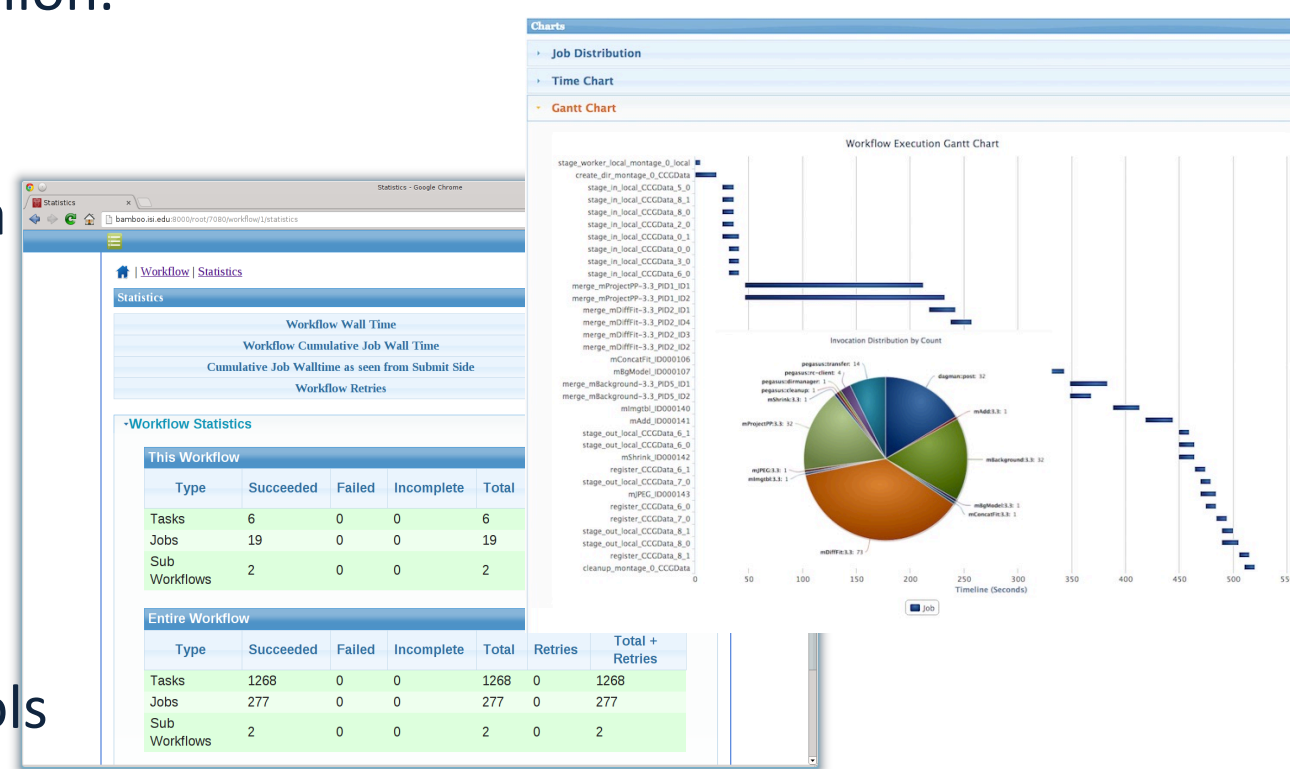
The same abstract workflow can, at different times, be mapped different execution environments such as XSEDE, OSG, commercial and academic clouds, campus grids, and clusters.

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.

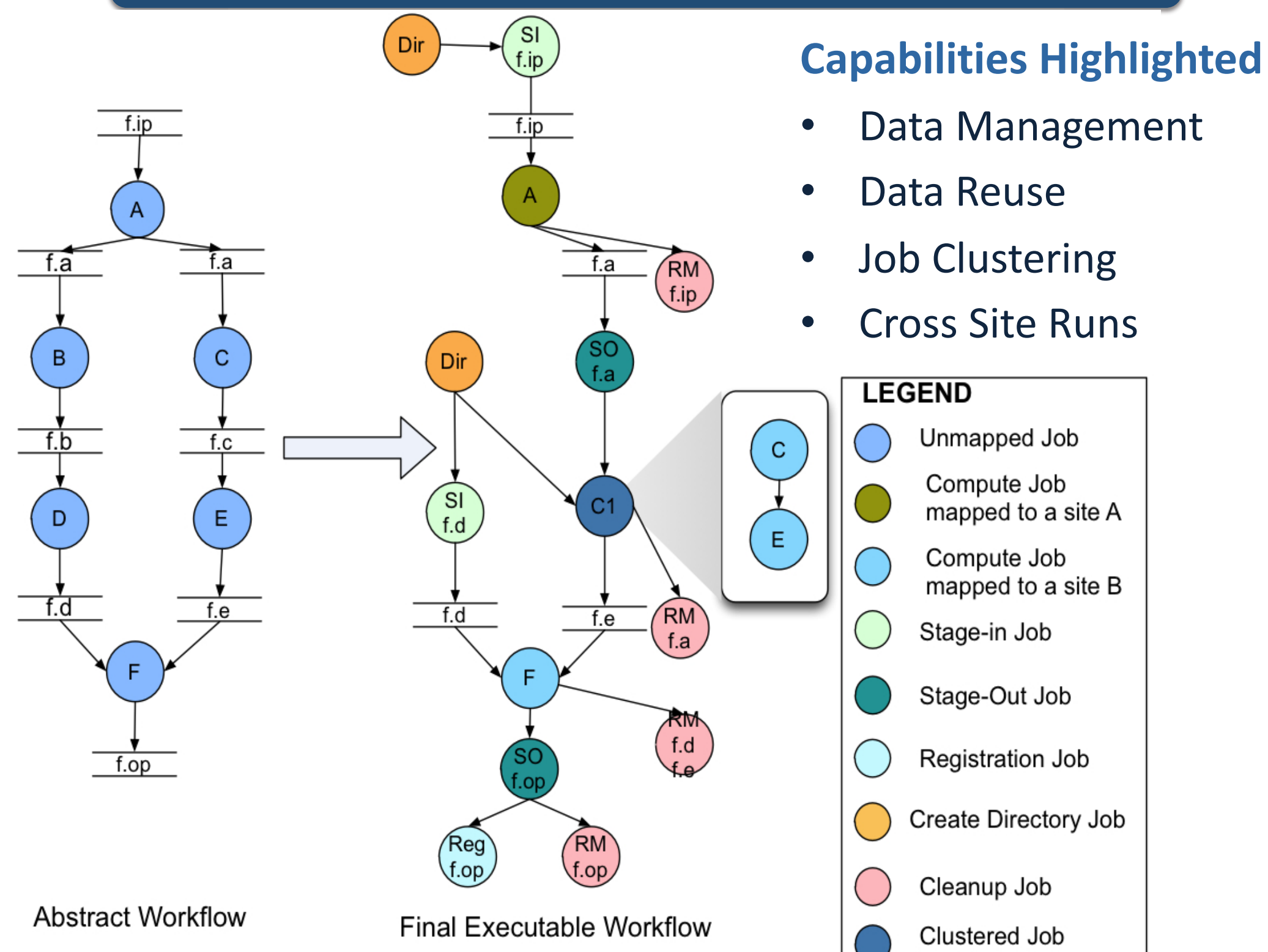
End to End automatic checksumming of workflow data to ensure data integrity.

Stores static and runtime metadata associated with workflow, files and tasks.

Accessible via command line tools and web-based dashboard.



## Canonical Workflow Example



## Applications using Pegasus

### Astronomy and Physics

- Pegasus powered workflows help detect gravitational waves.
- DECAM workflows to detect optical counterparts to LIGO detections
- Periodogram workflows help detect extra solar planets.
- Galactic Plane workflow generates mosaics for astronomy surveys.

### Seismology

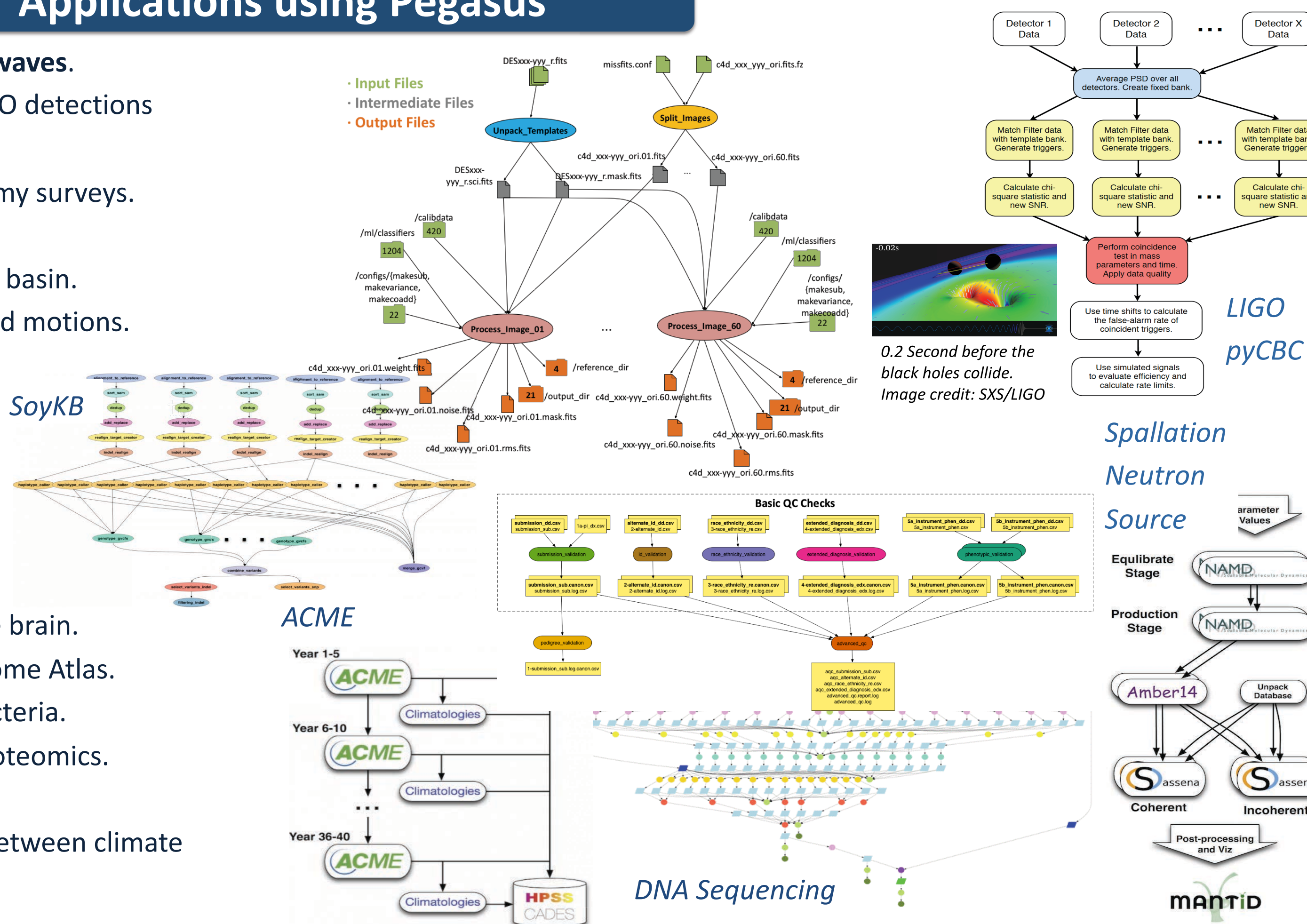
- CyberShake workflows for seismic hazard analysis of LA basin.
- Broadband workflows for accurate predictions of ground motions.

### Bioinformatics

- Quality control workflows for data submissions 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>

- ACME Climate Modeling Workflows study interaction between climate change and societal energy requirements.

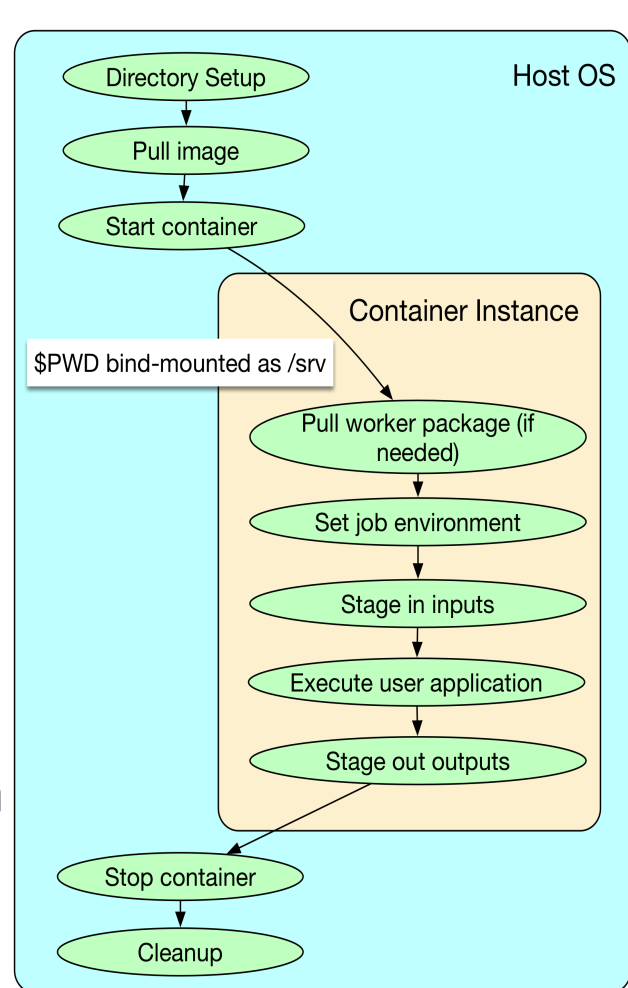


## Containers

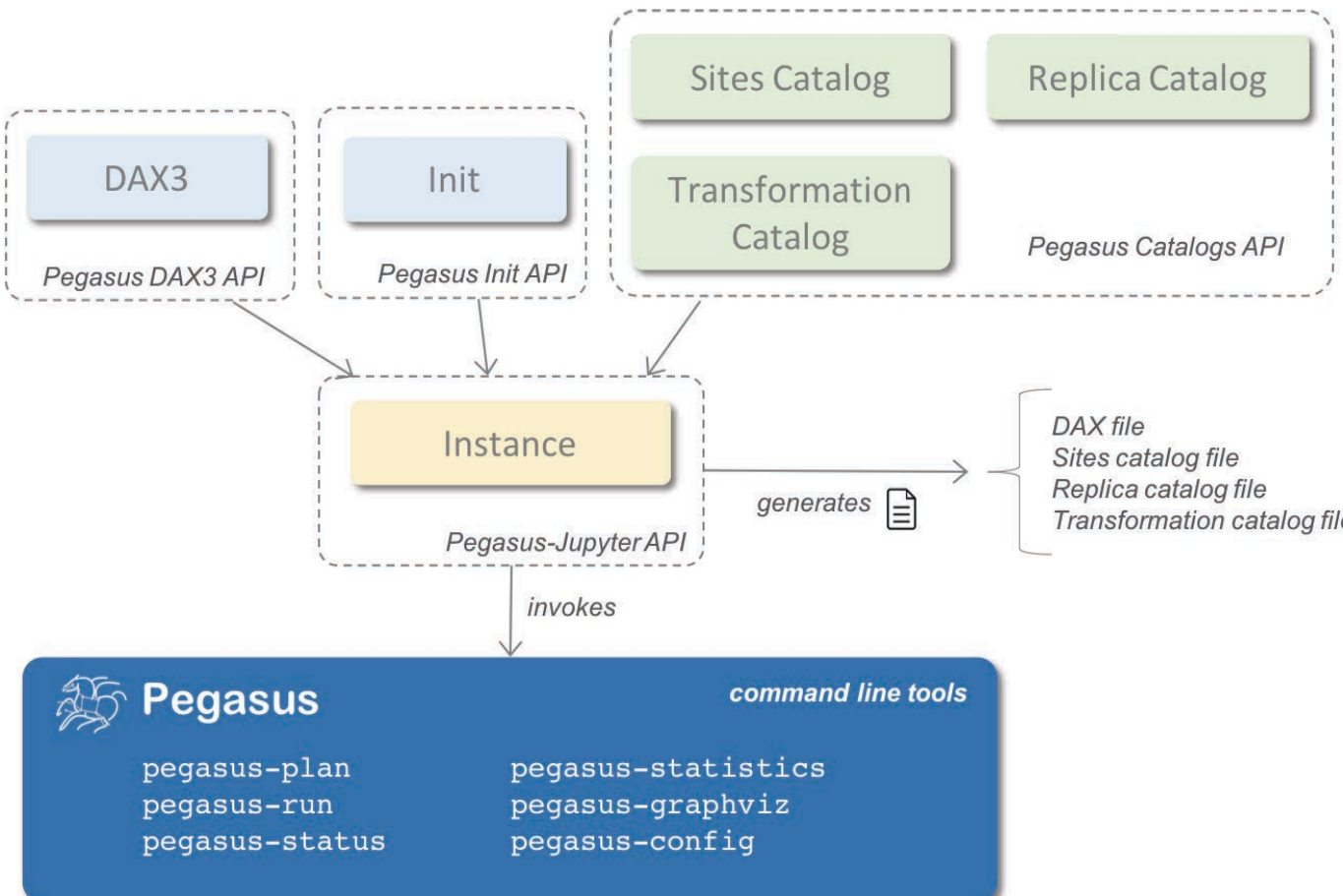
- Application containers provides a solution to package software with complex dependencies to be used during workflow execution
- Users have the option of either using a **different container for each executable** or **same container for all executables**.

### Container Execution Model

- Container image is put in the job directory along with input data.
- Loads the container if required on the node (applicable for Docker)
- Run a script in the container that sets up Pegasus in the container and job environment
- Stage-in job input data
- Launches user application
- Ship out the output data generated by the application
- Shut down the container
- Cleanup the job directory



## Jupyter Notebooks



The Pegasus-Jupyter integration aims to facilitate the usage of Pegasus via Jupyter notebooks.

- In addition to ease of use, notebooks
  - o foster **reproducibility** (all the information to run an experiment is in a unique place) and
  - o **reuse** (notebooks are portable if running in equivalent environments)
- Users can create a notebook and declare a workflow using the Pegasus DAX API, and then create an instance of the workflow for execution.
- This API encapsulates most of Pegasus commands (e.g., plan, run, statistics, among others), and also allows workflow creation, execution, and monitoring.

## Software Availability

### Release Schedule

- Major Release every 9 months. Minor releases every 4 months
- Continuous Integration Testing with Bamboo
- Issue Tracking via JIRA

### Download Options

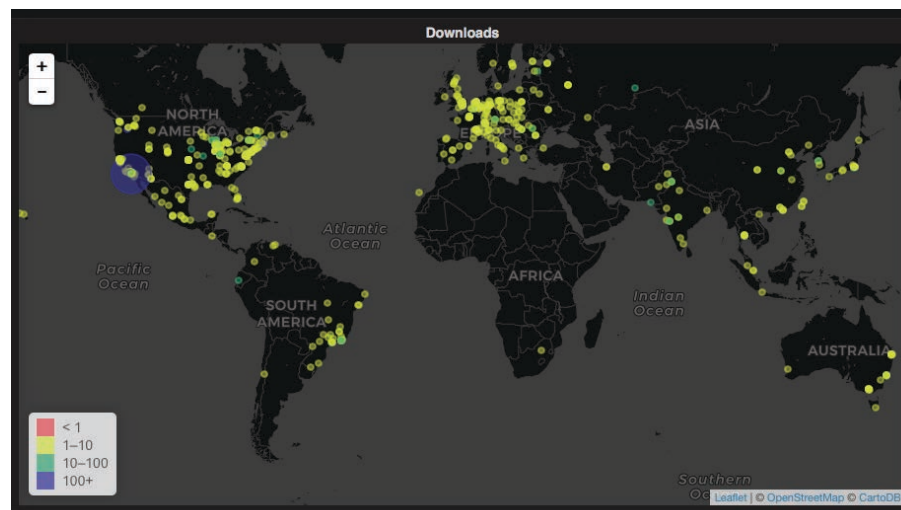
- Source Code publicly hosted on GitHub
- Binary packages for Linux and MAC
- YUM/APT repositories with RPM/DEB packages
- Nightly Developments builds also available

### Documentation / Training Materials

- Tutorials - Virtual Machine, EC2 and Docker images
- Support - Email lists and Online User Guide

### Downloads & Usage Since 2013

- Workflows - 1,741,851
- Tasks - 6,234,974,718
- Jobs - 1,297,468,849



Year	Downloads
2012	62
2013	980
2014	874
2015	1,981
2016	5260
2017	14,474
2018	20,878
2019	23,685