

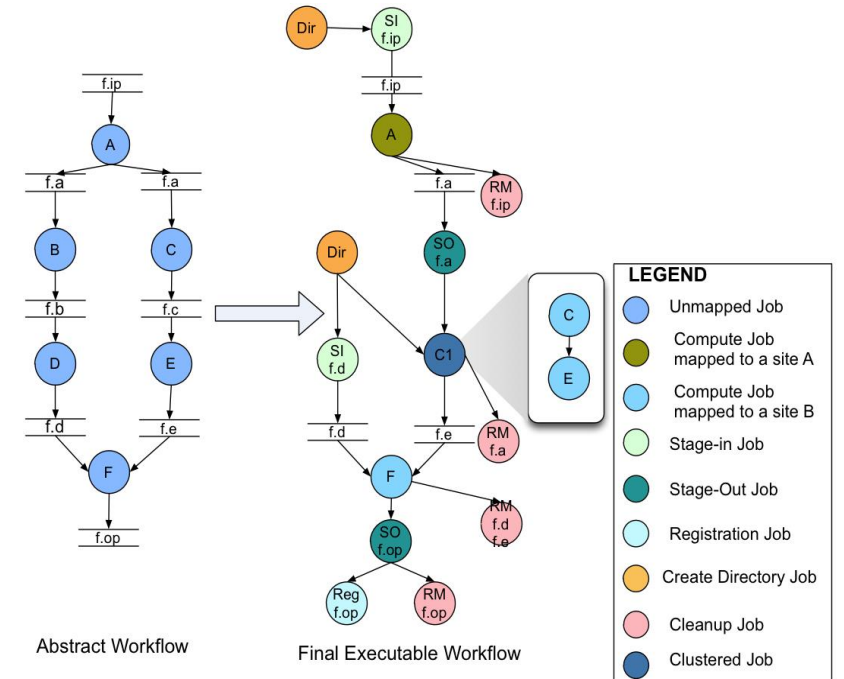


Pegasus: Automating Data Processing and Analysis for Large Facilities

Scientific workflow management system

Distributed computing – automatic data management

API for generating workflows – easy tie in with data discovery APIs



Features Highlights

- Data Management – workflow engine handles data placement and transfers
- Data Reuse – workflows are pruned to not redo already completed parts of the workflow
- Job Clustering – increased efficiency for high throughput workloads
- Cross Execution Environments – run across supercomputers, grids, and commercial and academic cloud providers at the same time

<https://pegasus.isi.edu>

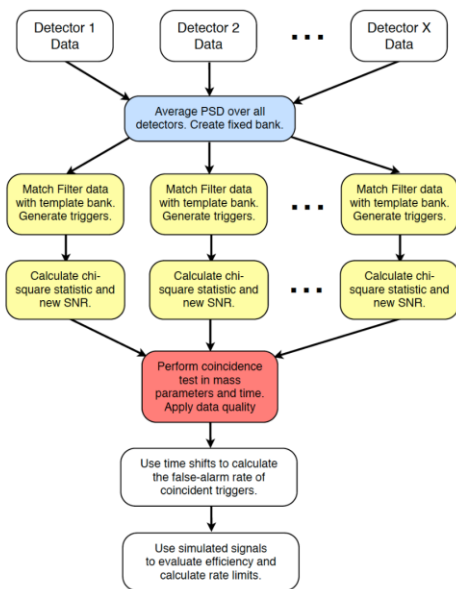


NSF SI2 program
grants #1664162,
#1148515

Pipelines to measure the statistical significance of data needed for discovery

Contains 100's of thousands of jobs and accesses on order of terabytes of data

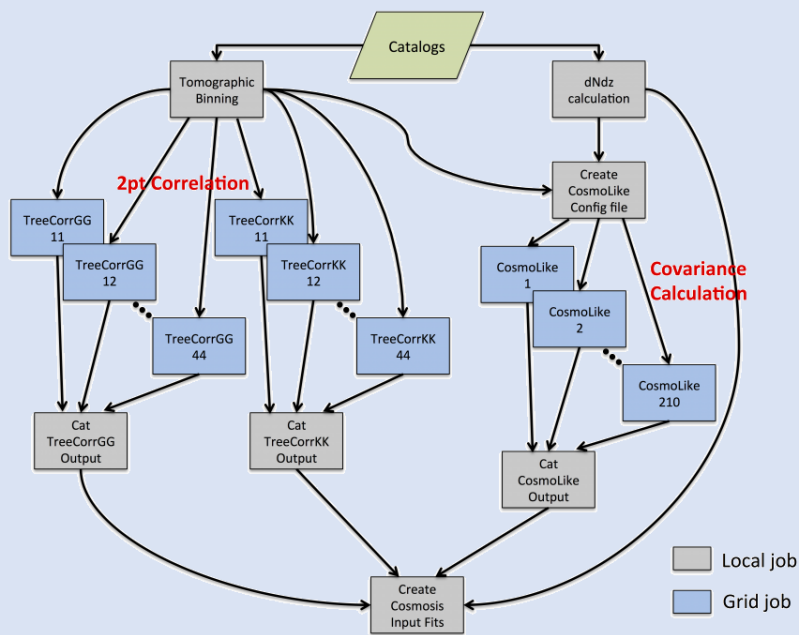
Data from multiple detectors – automatic data management



DES Weak Lensing Pipeline

Uses publicly available Science Verification catalogs of the Dark Energy Survey (DES-SV)

Application codes are bundled into Singularity containers, managed by Pegasus.



Earthquake regional simulation of 1.8 million buildings

Workflow is hierarchical with 50k buildings per subworkflow

OpenSees + UQ

AGAVE interface

