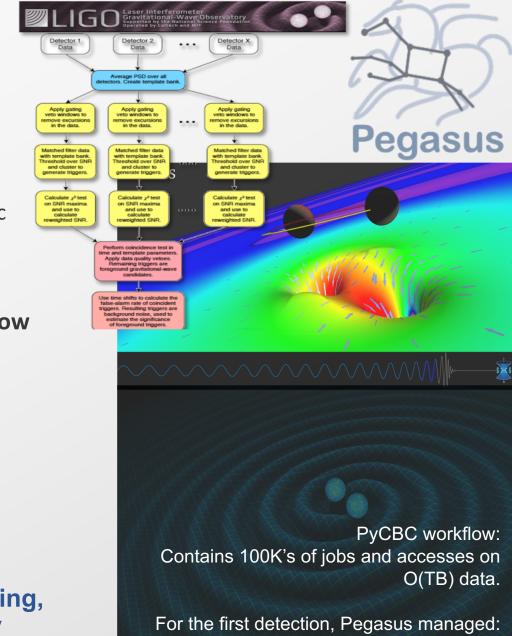# Pegasus: Automate, Recover, Debug

Ewa Deelman, University of Southern California, PI

- **Collaboration with Miron Livny, HTCondor**
- **Construct workflows in your favorite language (also Jupyter)**
- **Portability across heterogeneous infrastructure**
  - Separation of workflow description and execution
  - Support for campus and leadership class clusters, OSG, XSEDE, academic and commercial clouds
  - Can interact with a number of different storage systems (with different protocols)
  - Supports containers
- **Supports data reuse– useful in collaborations and ensemble workflow runs**
- **Reliability:** Recovers from failures, retry, workflow-level checkpointing
- **Scalability:** O(million) task, O(TB) data in a workflow
- **Restructures workflow for performance**
- **Supports reproducibility**
- **Web-based monitoring and debugging tools**
- **Can be included in various user-facing infrastructures**
- **Open source, available on Github**
- Since 2001 used **in astronomy, bioinformatics, climate modeling, earthquake science, molecular dynamics, helioseismology**
- **Funded by NSF under grant #1664162**

http://pegasus.isi.edu

PyCBC workflow:
Contains 100K's of jobs and accesses on O(TB) data.

For the first detection, Pegasus managed:
20,942 Workflows
107,576,294 Tasks
55,915,928 Jobs