CREATING A MULTI-WAVELENGTH GALACTIC PLANE ATLAS WITH AMAZON WEB SERVICES

G. Bruce Berriman, John Good

IPAC, California Institute of Technology

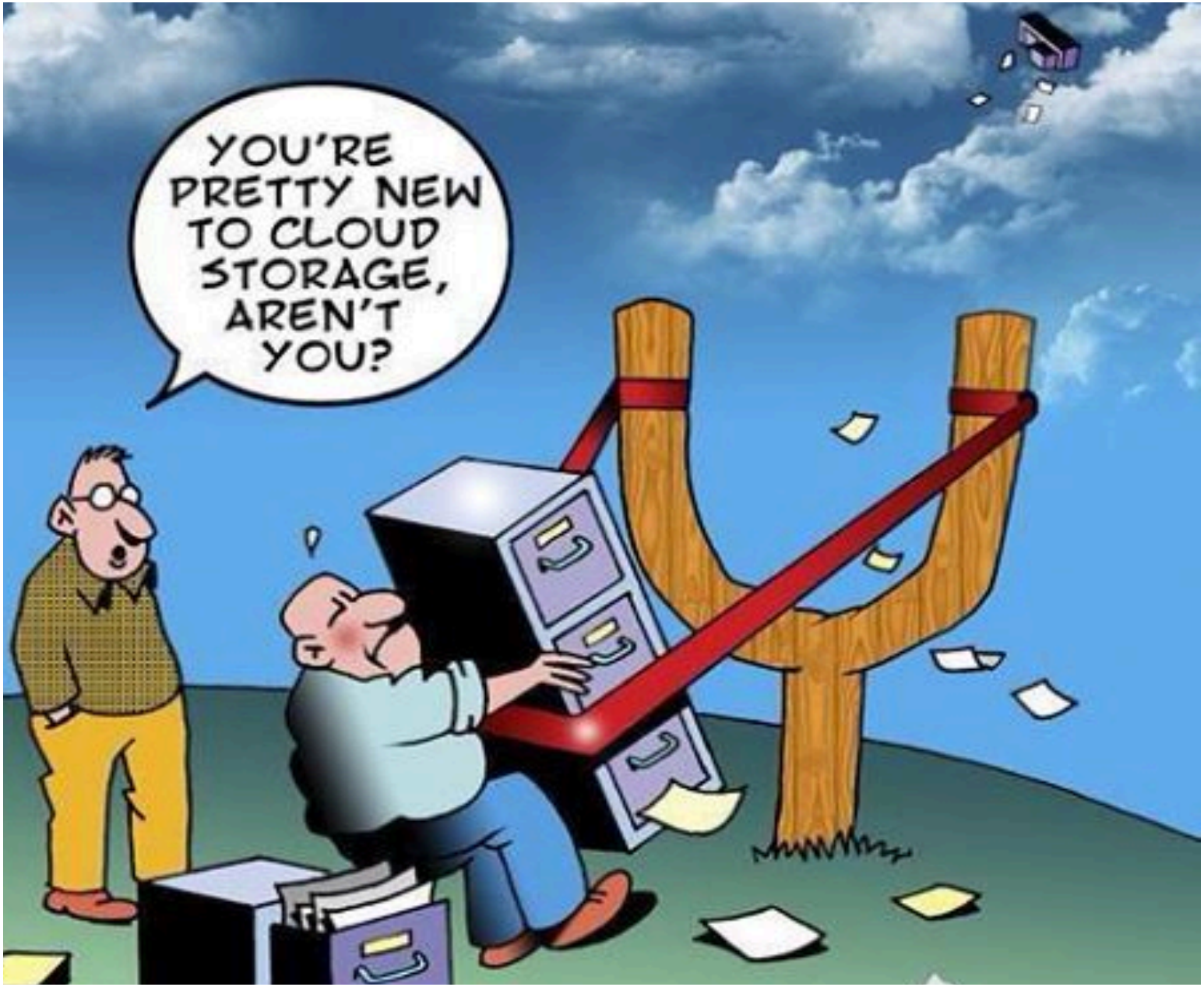
Ewa Deelman, Gideon Juve, Mats Rynge

ISI, University of Southern California

Jamie Kinney, Ann Merrihew

Amazon Web Services

Caltech ipac JPL USC University of Southern California NSF NASA


YOU'RE PRETTY NEW TO CLOUD STORAGE, AREN'T YOU?

How Can Astronomers Use The Cloud?

- How well can we …
  - Manage virtual machines and operate a virtual cluster for processing data.
  - Abstract the technical details of processing from astronomers.
  - Optimize processing and manage costs.
  - Exploit cloud storage in operations.
  - Create turnkey tools for astronomers to exploit cloud platforms.

A Multi-wavelength Image Atlas of the Galactic Plane

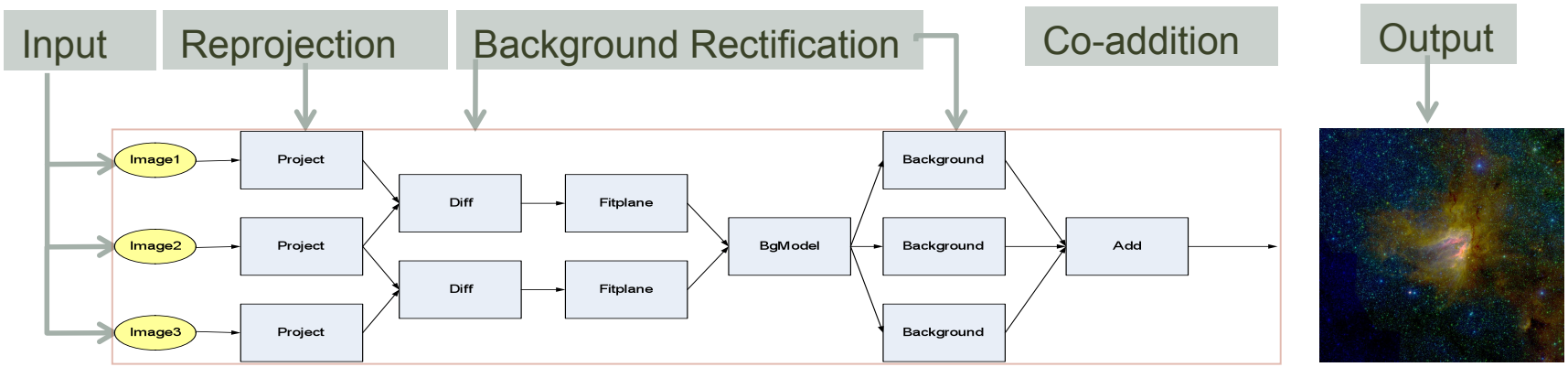
- Use **open source tools** to create a new data product on the Amazon Elastic Compute Cloud (EC2).
- 16 wavelengths from 1μm to 24 μm.
- Coverage l = 0° - 360°; b = ±20°.
- Processed at 1 arcsec spatial sampling, Cartesian projection, Galactic coordinates.

Survey / Bands (μm)	Coverage of 360°x40° area	Output Size (TB)	Compute time (1,000s core hours)
2MASS (1.2, 1.6, 2.2)	100%	14.4	87
GLIMPSE (3.6, 4.5, 5.8, 8.0)	11%	2.0	60
MIPSGAL (24)	8%	0.4	3
MSX (8.8, 12.1, 14.6, 21.3)	35%	6.8	36
WISE (3.4, 4.6, 12, 22)	100%	19.2	132

The Montage Image Mosaic Engine

- ANSI-C Toolkit for creating and managing image mosaics in FITS format.
- Portable and scalable – runs on all Unix platforms.
- BSD 3-clause license.
- Widely adopted by astronomy and IT communities.

Three-color Spitzer mosaic of the LMC created from 300,000 frames. Seven degrees on a side.
Blue - 3.6 μm; green- 8 μm red - 24 μm

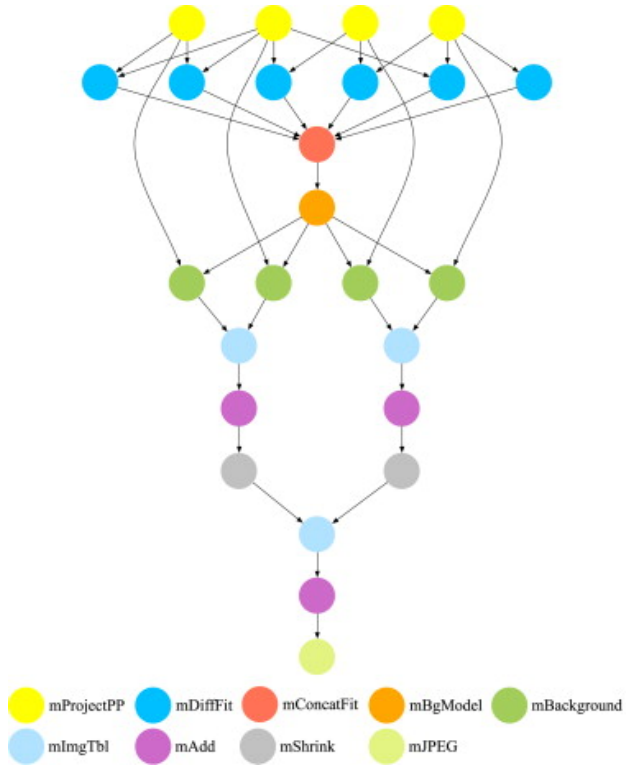


Montage Workflow

http://montage.ipac.caltech.edu

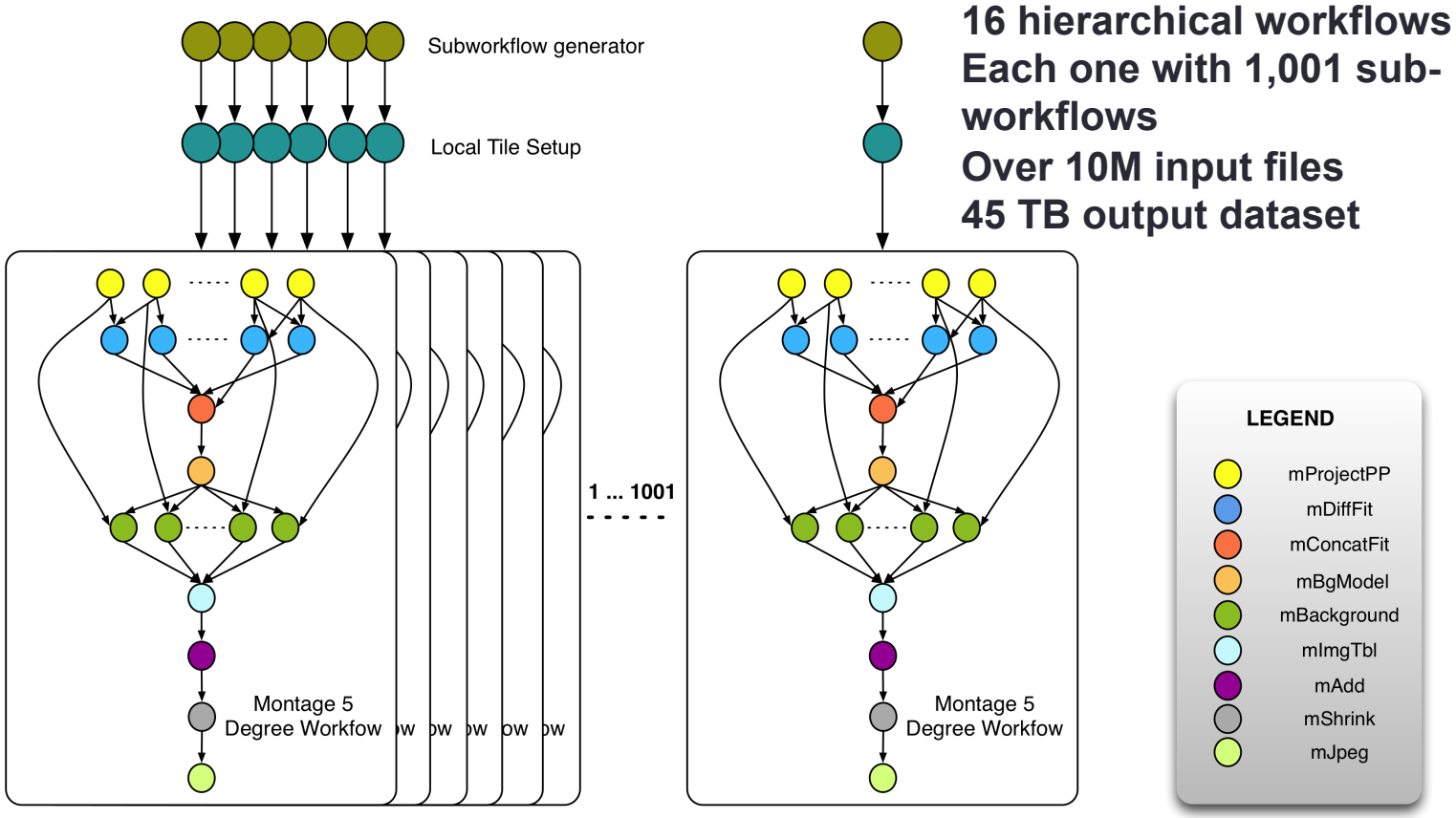
The Pegasus Workflow Management System

- Pegasus is a workflow planner
  - Mature and used in many fields.
  - Takes *abstract* descriptions of workflows and sets up and runs a processing plan for an environment whose configuration is incorporated into workflow manager.
  - Layered on HTCondor (scheduler) and DAGMan (submits jobs to HTCondor in the correct order).
  - EC2 environment is already known to Pegasus.
- Abstract Workflows
  - Directed Acyclical Graphs
  - Identifies the computational flow and dependencies.

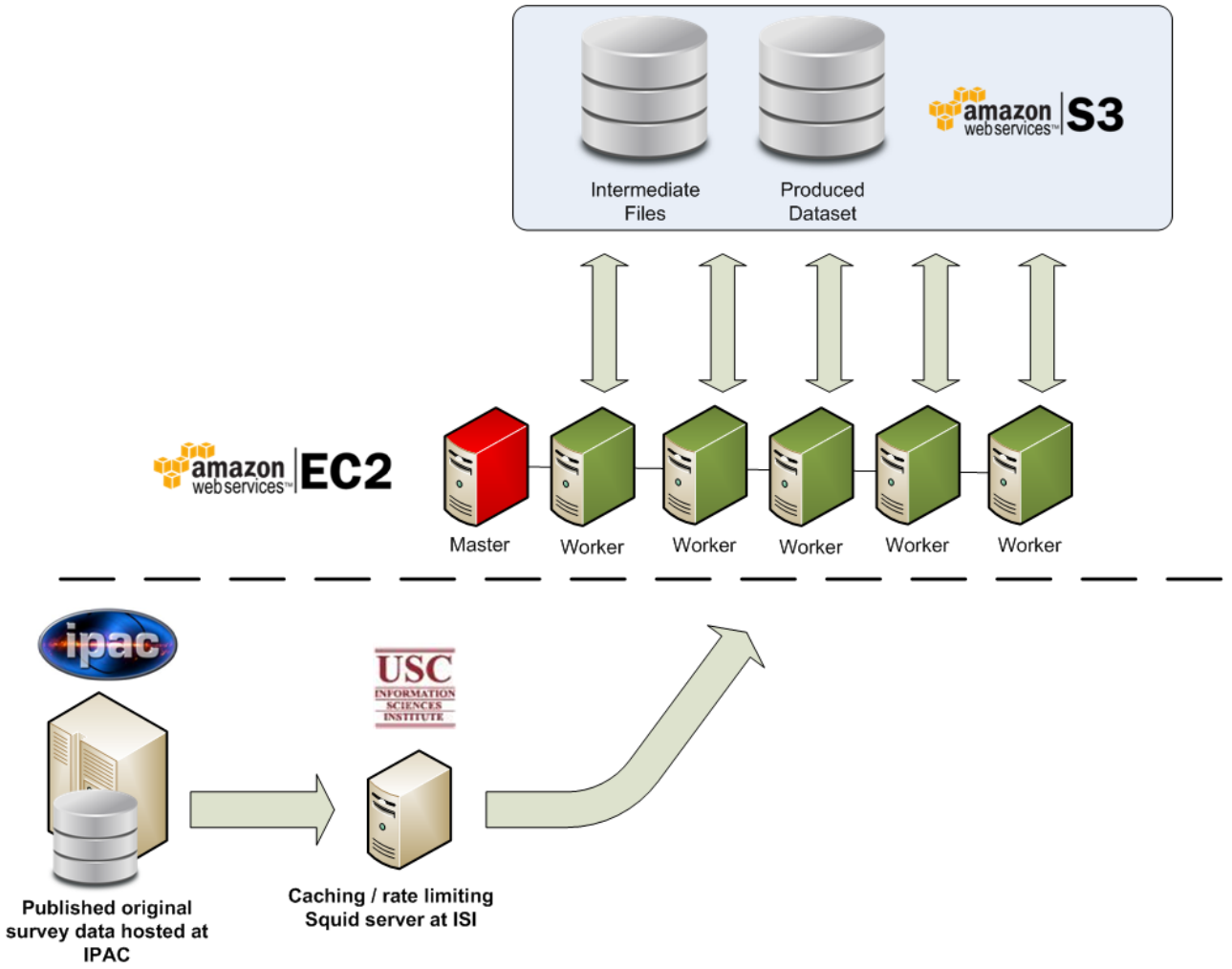


http://pegasus.isi.edu

Galactic Plane Workflow



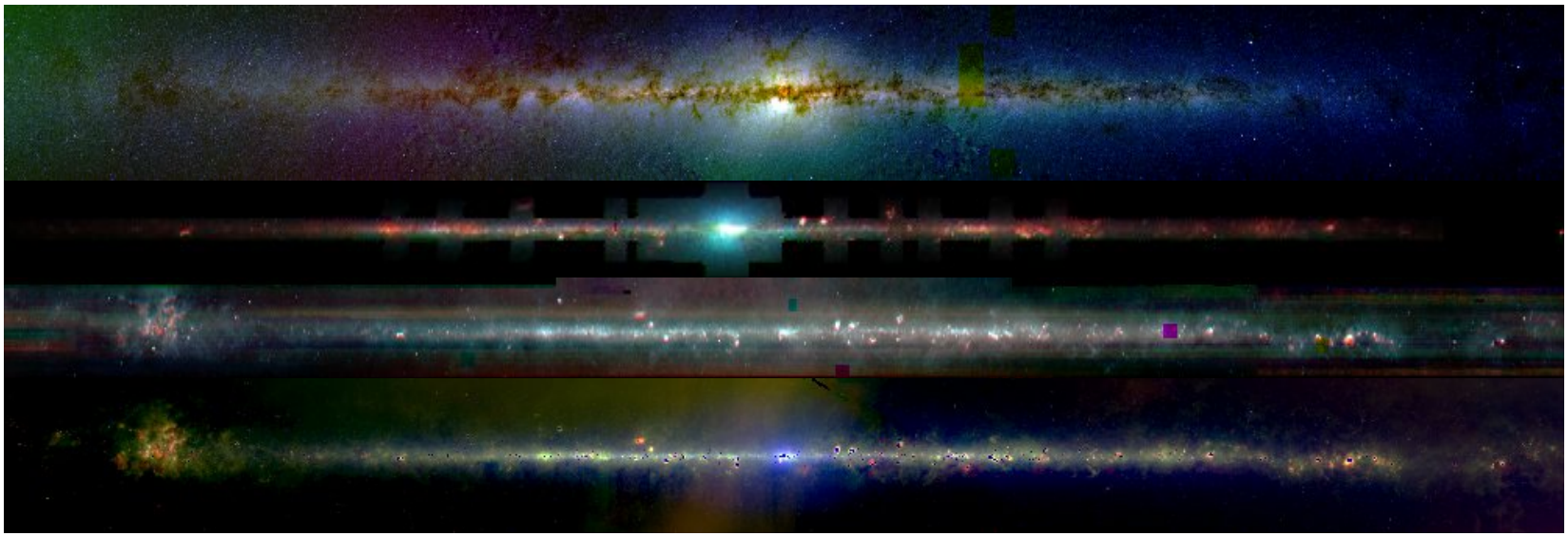
System Architecture Design



Be careful about your choices.

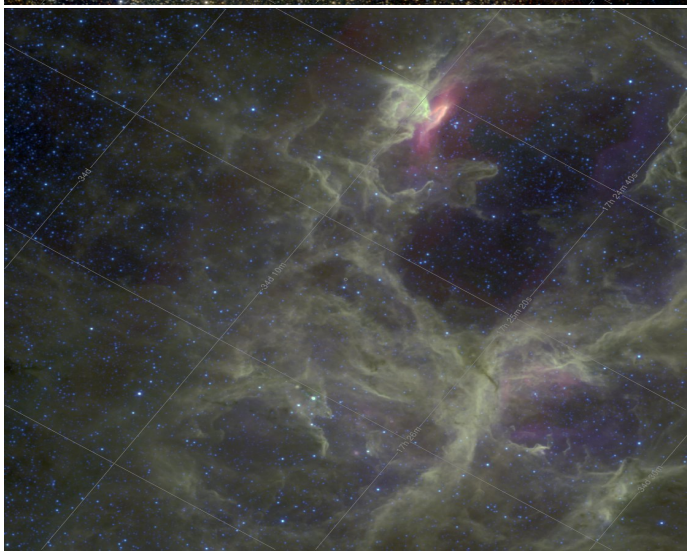
- Amazon offers a wide range of instance types and costs of running applications can vary widely.
- **hi1.4xlarge** (the one we used)
  - Memory optimized, with 2 x SSD ephemeral drives.
  - 318,000 core hours.
  - Spot instance price: US $5,950.
- **cc2.8xlarge** (benchmarked)
  - Compute cluster optimized, with 4 ephemeral drives (2 used.)
  - 274,000 core hours.
  - Spot instance price: US $2,200.
- Best practice: do a cost benefit analysis or benchmarking.

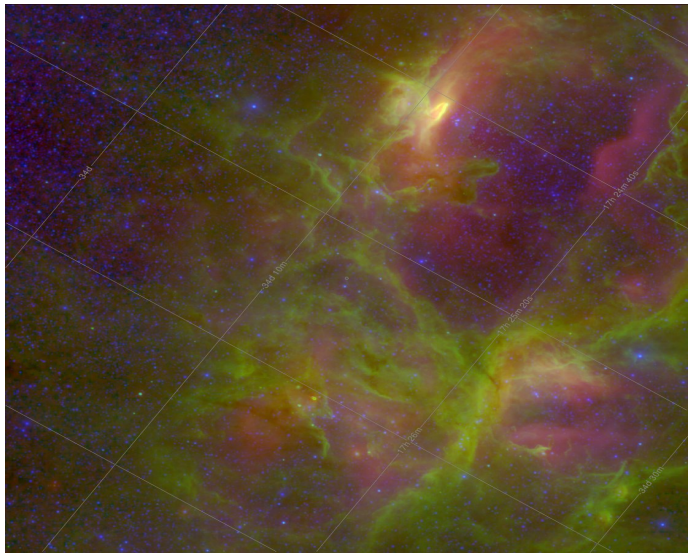
Multi-Color Mosaics



Top to bottom: 2MASS, GLIMPSE, MSX, WISE

Three color mosaics near NGC 6537

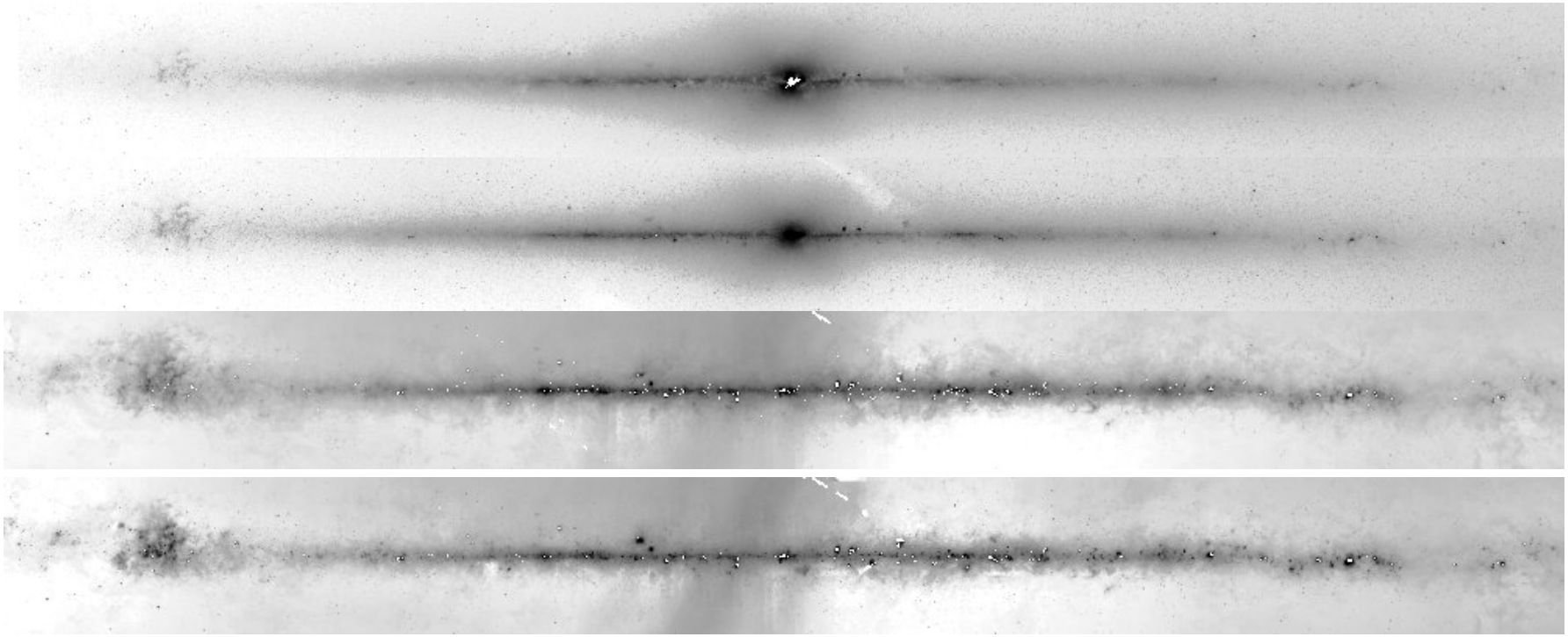




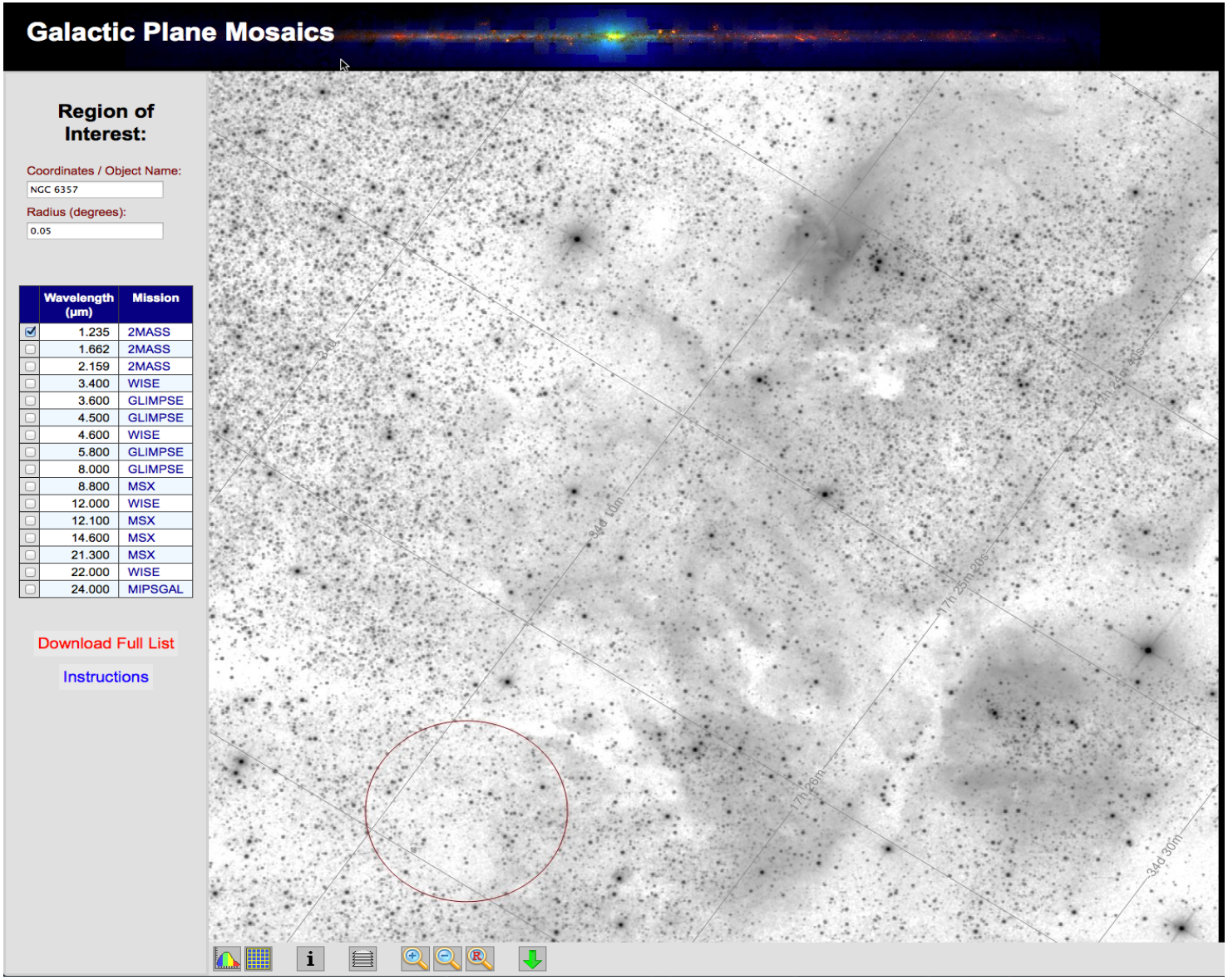
Three color-infrared images of the Galactic Plane, extracted from the 16-wavelength Galactic Plane Atlas.

Top left: 2MASS 1.2 μm, 1.6 μm, 2.2 μm
Top right: GLIMPSE 4.5 μm, 5.8 μm and 8 μm
Bottom left: 2MASS 1.2 μm, GLIMPSE 8.0 μm, MSX 21.3μm

WISE Mosaics



Top to bottom: 3.4 μm, 4.6 μm, 12 μm, 22 μm.



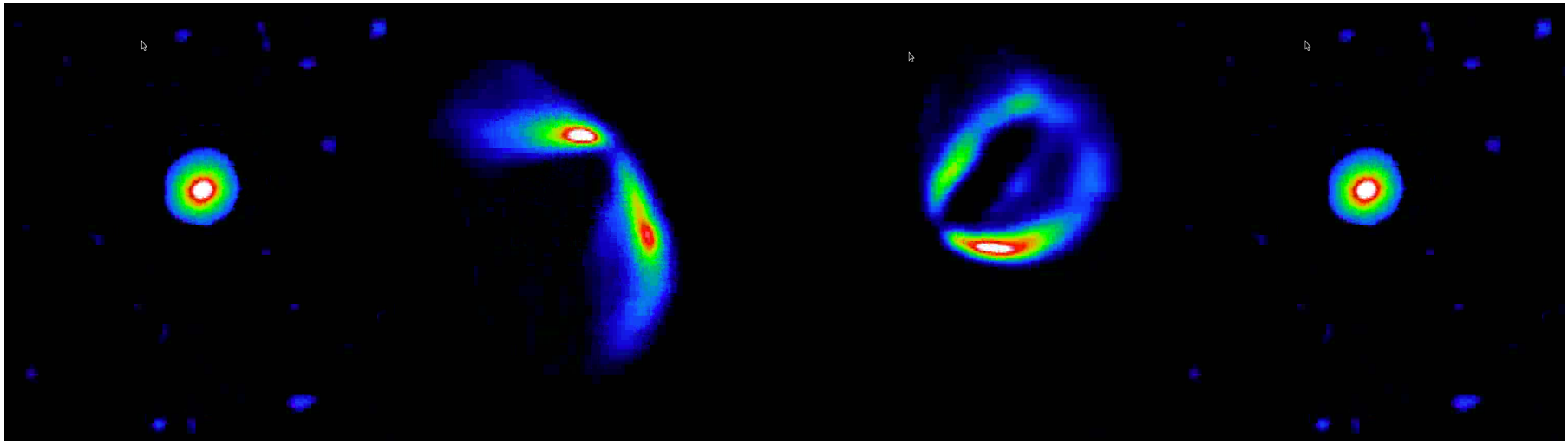
http://vaoweb2.ipac.caltech.edu/cgi-bin/GalPlane/nph-gpInit

What Have We Learned?

- Amazon EC2 is a powerful platform for astronomical computing.
- Open Source tools are invaluable for taking advantage of it.
- Optimizing operations requires human intervention.
- Be mindful of pay-as-you-go: need to do cost benefit analyses, benchmarking for optimization.
- High latency in recovering images from S3 storage.

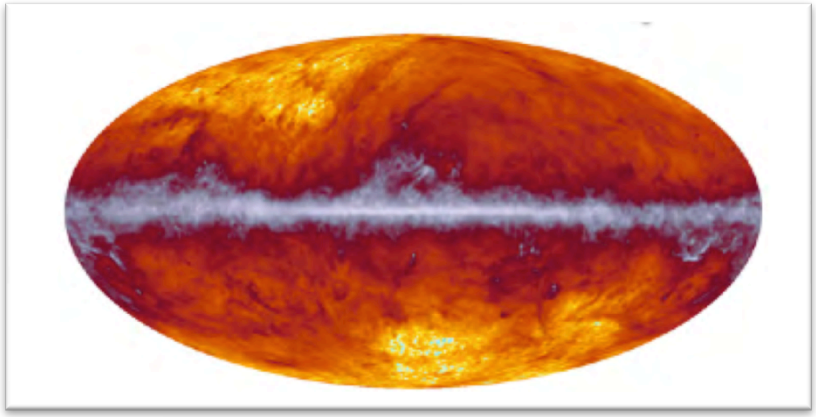
Next Generation Montage: Data Cubes

The structure of a molecular disk wind in HD 163296, measured by ALMA (PI: M. Rawlings).
Re-projection by **Montage** of a **data cube** of the star that covers multiple velocities relative to the center of the CO J=3-2 line. Shown are images at four velocities.

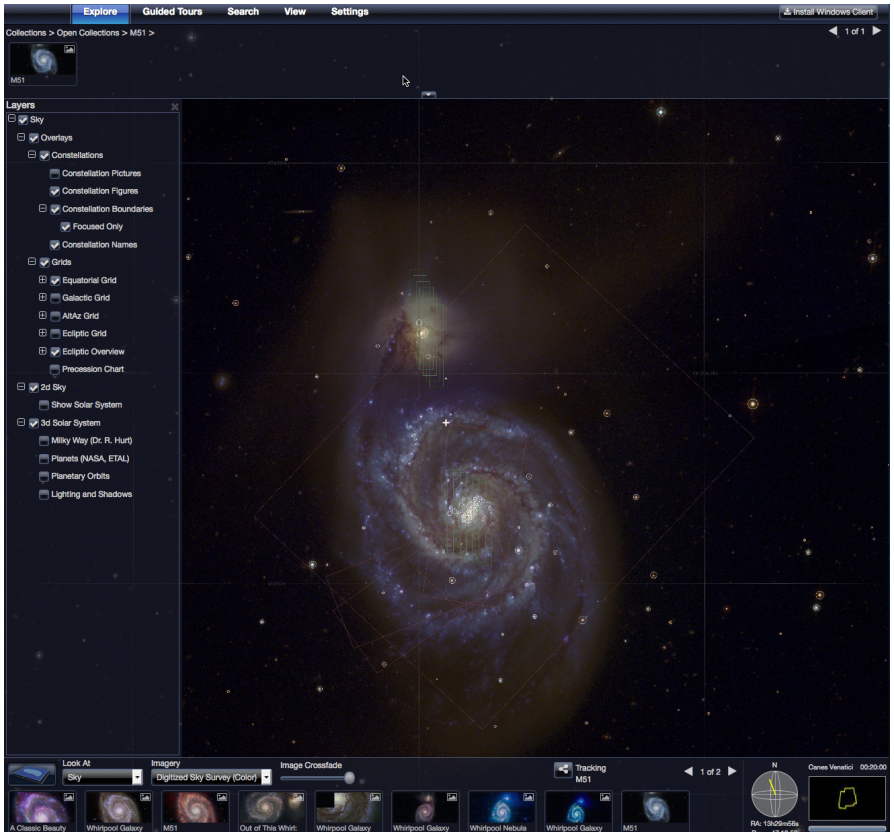


Next Generation Montage: Sky Partitioning Schemes

Transform **HEALPix** images, standard in cosmic background missions, for multi-wavelength image analysis and visualization.



353 GHz all-sky map measured by Planck.



Re-project images to **TOAST** and visualize in World Wide Telescope