Producing a Multiwavelength Galactic Plane Atlas Using Montage, Pegasus and Amazon Web Services

Mats Rynge\textsuperscript{1}, Gideon Juve\textsuperscript{1}, Jamie Kinney\textsuperscript{2}, John Good\textsuperscript{3}, Bruce Berriman\textsuperscript{3}, Ann Merrihew\textsuperscript{2}, Ewa Deelman\textsuperscript{1}

\textsuperscript{1}USC Information Sciences Institute
\textsuperscript{2}Amazon Web Services
\textsuperscript{3}Infrared Processing and Analysis Center, Caltech
Science Goal

- Multiwavelength image atlas of the Galactic Plane, with coverage of 360° along the galactic plane and ±20° on either side
- 16 different wavelengths from 1 µm to 24 µm
- Each output image is 5° by 5° in size, and have an overlap of 1° with neighboring tiles
- Processed so that they appear to have been measured with a single instrument observing all 16 wavelengths - Cartesian projection
- When complete, the data will be released to the community via an API

<table>
<thead>
<tr>
<th>Survey / Bands (µm)</th>
<th>Coverage of 360° x 40° area</th>
<th>Output Size (TB)</th>
<th>Compute time (1,000s core hours)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2MASS (1.2, 1.6, 2.2)</td>
<td>100%</td>
<td>14.4</td>
<td>87</td>
</tr>
<tr>
<td>GLIMPSE (3.6, 4.5, 5.8, 8.0)</td>
<td>11%</td>
<td>2.0</td>
<td>60</td>
</tr>
<tr>
<td>MIPS-GAL (24)</td>
<td>8%</td>
<td>0.4</td>
<td>3</td>
</tr>
<tr>
<td>MSX (8.8, 12.1, 14.6, 21.3)</td>
<td>35%</td>
<td>6.8</td>
<td>36</td>
</tr>
<tr>
<td>WISE (3.4, 4.6, 12, 22)</td>
<td>100%</td>
<td>19.2</td>
<td>132</td>
</tr>
</tbody>
</table>
The Montage Image Mosaic Engine

- Toolkit written in ANSI-C for creating and managing image mosaics in FITS format.

- Portable and scalable – runs on desktops, grids and cloud computing platforms under *nix platforms.

- Code available through clickwrap license at Caltech.

- Widely adopted by astronomy and IT communities: used on desktops, integrated into processing pipelines, used in development cyber-infrastructure.
Pegasus Workflow Management System

- Builds on top of HTCondor and DAGMan.

- Abstract Workflows - Pegasus input workflow description
  - Workflow “high-level language”
  - Only identifies the computation, devoid of resource descriptions, devoid of data locations

- Pegasus is a workflow planner/mapper (“compiler”)
  - Transforms the workflow for performance and reliability
  - Automatically locates physical locations for both workflow components and data
  - Collects runtime provenance
Galactic Plane Workflow

16 hierarchal workflows
Each one with 1,001 subworkflows
Over 10M input files
45 TB output dataset

Subworkflow generator
Local Tile Setup

1 ... 1001

Montage 5 Degree Workflow

LEGEND

- mProjectPP
- mDiffFit
- mConcatFit
- mBgModel
- mBackground
- mlmgTbl
- mAdd
- mShrink
- mJpeg
System Overview

Published original survey data hosted at IPAC

Caching / rate limiting Squid server at ISI

Intermediate Files

Produced Dataset

Master

Worker

Worker

Worker

Worker

Worker
### Numbers

- Amazon Web Services contributed the computations and storage
  - **hi1.4xlarge instance (the one we used)**
    - Memory optimized, with 2 x SSD ephemeral drives
    - 318,000 core hours
    - Spot instance price: $5,950
  - **cc2.8xlarge instance (benchmarked)**
    - Compute cluster optimized, with 4 ephemeral drives (2 used)
    - 274,000 core hours
    - Spot instance price: $2,200

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

rynge@isi.edu

- **Pegasus** - [http://pegasus.isi.edu/](http://pegasus.isi.edu/)
  - NSF funded
  - Open Source
  - Documentation, tutorial, and support available on website

- **Montage** - [http://montage.ipac.caltech.edu/](http://montage.ipac.caltech.edu/)