



# AstroPortal: A Science Gateway for Large-scale Astronomy Data Analysis

## Ioan Raicu

Distributed Systems Laboratory  
Computer Science Department  
University of Chicago

## Joint work with:

**Ian Foster:** Univ. of Chicago, CS & Argonne National Laboratory, MCS

**Alex Szalay:** Dept. of Physics and Astronomy, Johns Hopkins University

**Gabriela Turcu:** Univ. of Chicago, CS

**TeraGrid Conference 2006**

June 13<sup>th</sup>, 2006

# Introduction



- Science Portals: gateway to Grid resources
- Potential Applications Characteristics
  - Large data sets
  - Large number of users
  - Easy (but not necessarily trivial) parallelization
- Applicable fields:
  - Astronomy
  - Medicine
  - Others

# Astronomy Field



- Astronomy datasets (i.e. SDSS) are the crown-jewels
  - SDSS DR4
    - 1.3M images
      - 300M+ objects
      - 3TB compressed images (2MB x 1.3M)
      - 8TB raw images (6.1MB x 1.3M)
    - 100K worldwide potential users (100s of big users)
- Applications:
  - Stacking
  - Montage



A screenshot of a Mozilla Firefox browser window displaying the AstroPortal Stacking Service interface. The browser's address bar shows the URL: `http://people.cs.uchicago.edu/~iraicu/research/AstroPortal/`. The page title is "AstroPortal Stacking Service".

The main content area features a large, semi-transparent watermark that reads "DEFWAO". Below the watermark, there is a table of data with the following values:

194.940	132658	2.98	4884	1
194.993	538067	2.95	3381	1
194.993	485523	2.89	4869	326 1
194.941	099309	2.97	5258	417 1
194.988	214584	2.97	17907	681 1
194.997			217682	

Below the table, there is a link labeled "Upload Description File".

The interface includes a login form with the following fields and buttons:

- User ID:**
- Password:**
- Stacking ID:**
- Upload Description File:**
- Buttons:** Submit, Reset, Browse...

At the bottom of the browser window, there is a "Done" button.

For more information about the AstroPortal, please see the [About Page](#).

# AstroPortal: Stacking Service Results

User ID: iraicu  
Password: \*\*\*\*\*  
Stacking Description: [stacking\\_description.txt](#)  
Stacking Size: 20  
AstroPortal Web Service Location: <http://fig-viz-login.uc.teragrid.org:50001/wsrff/services/AstroPortal/core/WS/APFactoryService>

RESULT:  
  
Size: 43 KB  
Dimensions: 100x100 pixels  
Download result: [stacked\\_result.fit](#)

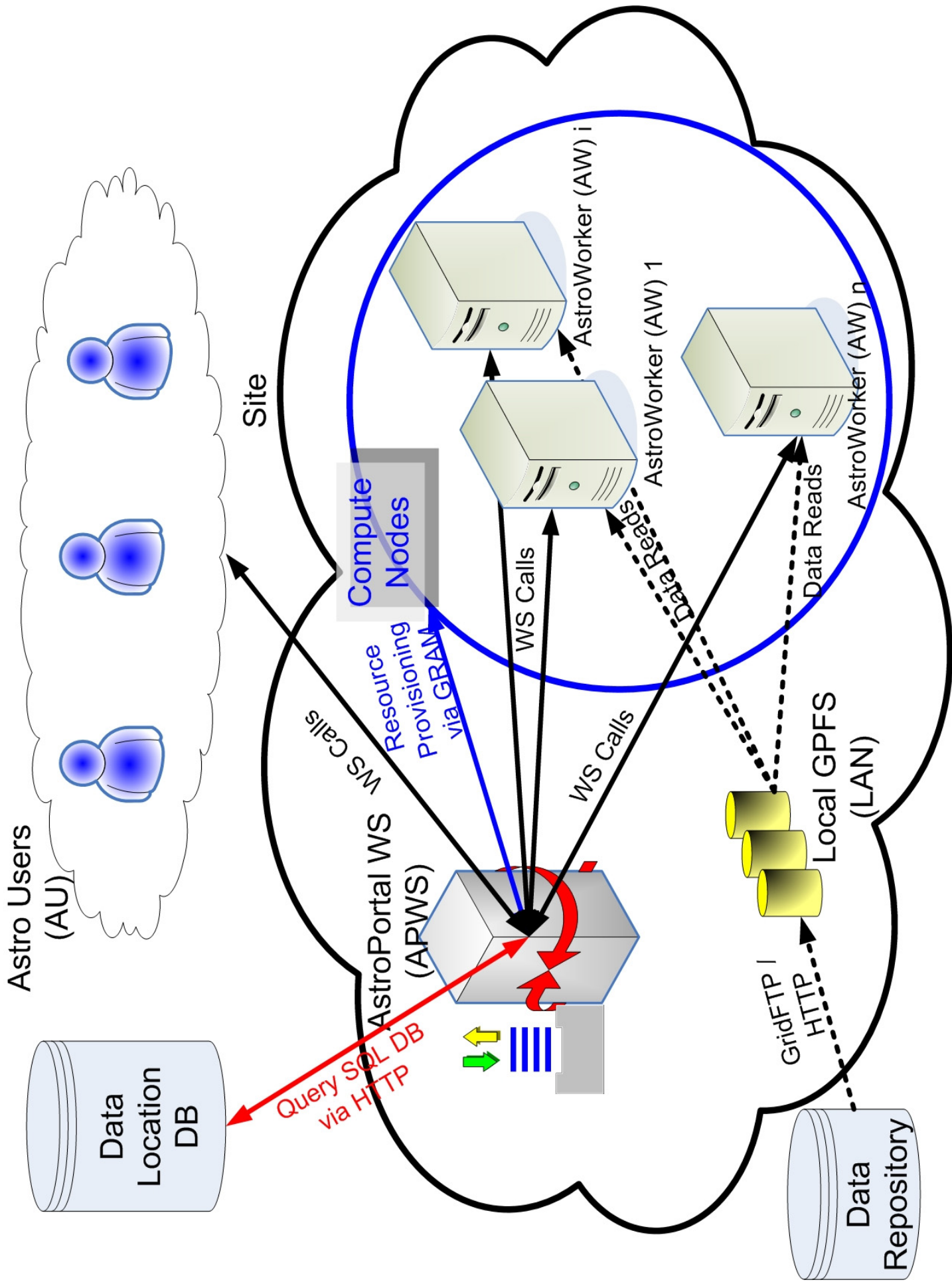
Time to complete Stacking: 5.164 seconds  
Number of physical resources utilized: 16  
Number of Stacking completed successful: 18  
Number of Star Objects not found in the SDSS dataset: 1  
List of Star Objects [ra, dec, band] not found:

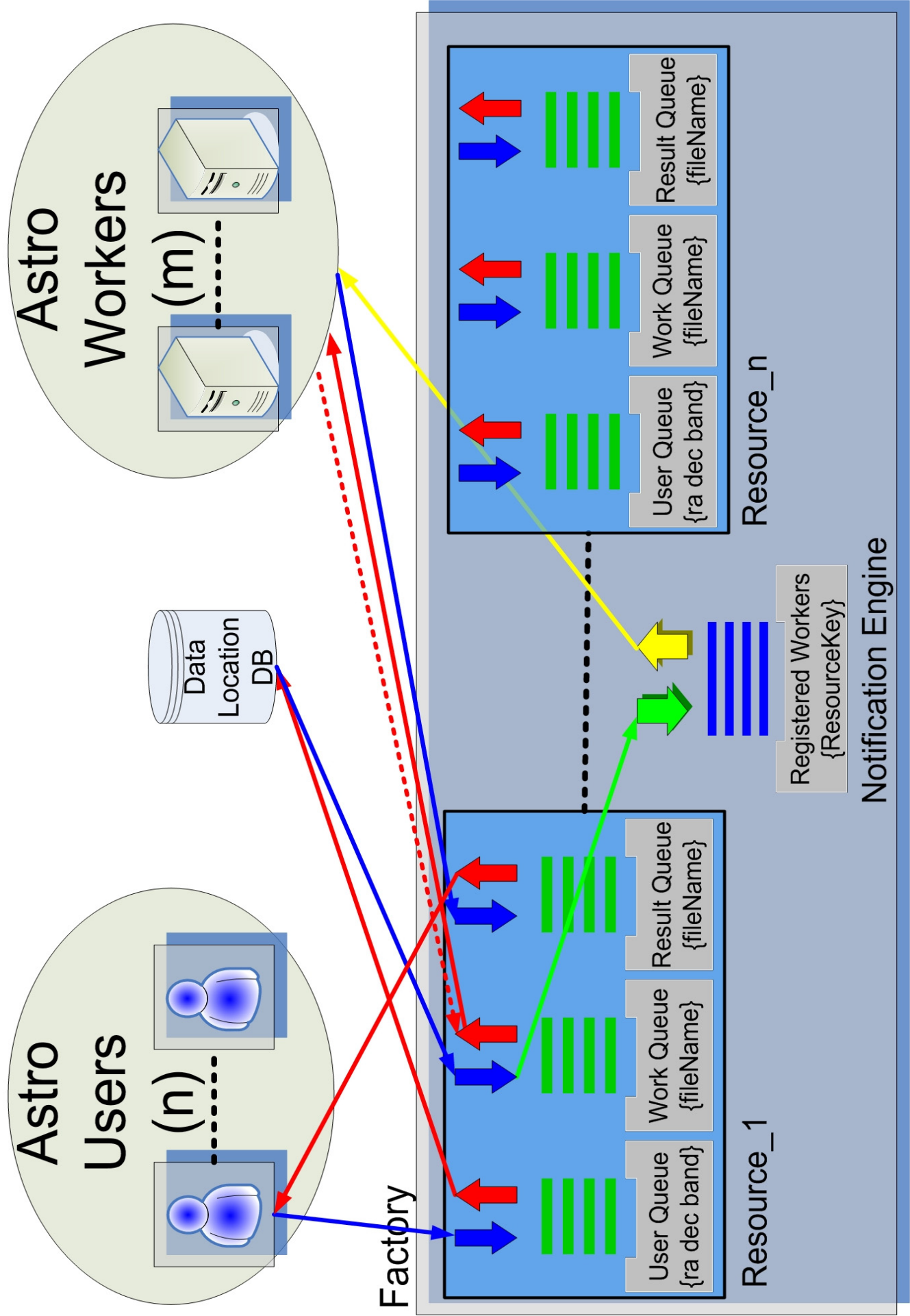
- [194.969060213455, -13.90189344168167, r]

Number of Data Objects not found in the data cache: 1  
List of Data Objects {ra, dec, band} filename [x\_coord x y\_coord] not found:

- ([194.969705877549, 2.93855950426612, r]  
/disks/scratchgfs1/iraicu/sdss\_gz/das.sdss.org/DR4/data/imaging/752/40/corr/6/fig-C-000752-r6-0245.fit.gz [0 x 0])

To start a new stacking, go back to the main [Stacking Service](#).





# AstroPortal Web Service

# Open Research Questions



- Site level
  - advanced reservations
  - resource allocation
  - resource de-allocation
- Data management
  - Data location and replication
  - Data caching hierarchies
- Resource management
  - Distributed resource management between various sites



# Questions?



- More information: <http://people.cs.uchicago.edu/~iraicu/research/>
- Related materials and further readings:
  - Ioan Raicu, Ian Foster, Alex Szalay, Gabriela Turcu. “**AstroPortal: A Science Gateway for Large-scale Astronomy Data Analysis**”, to appear at TeraGrid Conference 2006, June 2006.
  - Alex Szalay, Julian Bunn, Jim Gray, Ian Foster, Ioan Raicu. “**The Importance of Data Locality in Distributed Computing Applications**”, NSF Workflow Workshop 2006.
  - Ioan Raicu, Ian Foster, Alex Szalay. “**Harnessing Grid Resources to Enable the Dynamic Analysis of Large Astronomy Datasets**”, under review at SuperComputing 2006.
  - Ioan Raicu, Ian Foster, Alex Szalay, Gabriela Turcu, Catalin Dumitrescu. “**Enabling Large-scale Astronomy Data Analysis with the AstroPortal**,” under preparation for the HPC Analytics Challenge at SC06.
  - Ioan Raicu, Ian Foster, Elizeu Santos-Neto, John Bresnahan. “**3DcacheGrid: Dynamic Distributed Data Cache Grid Engine**,” under preparation for the HPC Storage Challenge at SC06.



THE UNIVERSITY OF  
**CHICAGO**

AstroPortal

