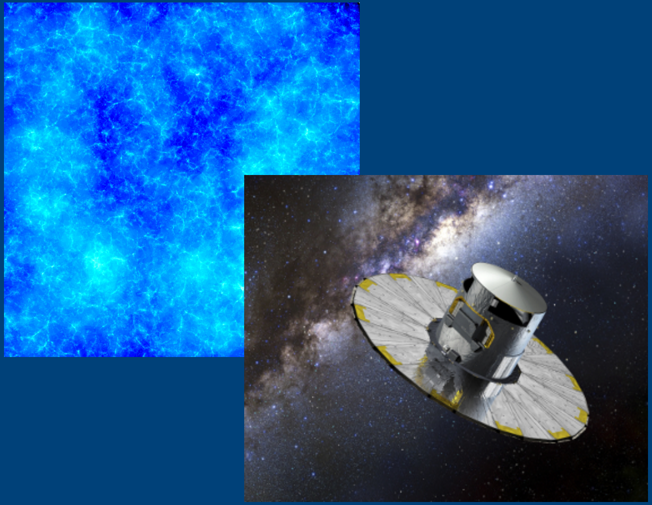
AIP Leibniz-Institut für Astrophysik Potsdam

# Daiquiri – an VO ready solution for medium size data providers



Anastasia Galkin
Jochen Klar
Gal Matievic
Harry Enke

Asterics data provider forum, 27.06.2018

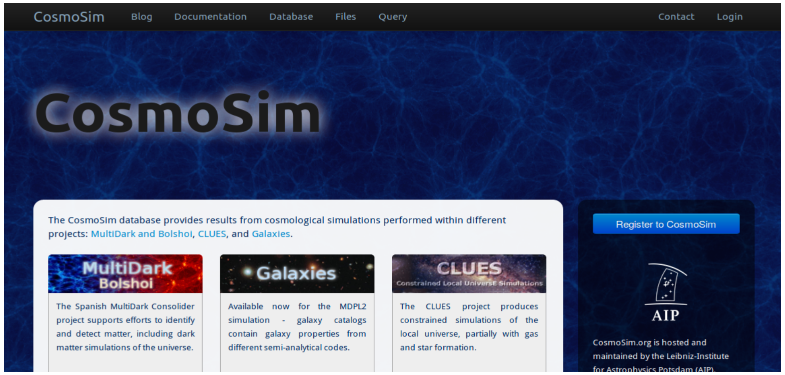
# Daiquiri

## A framework for the publication of scientific databases

- Allows for highly customizable web applications
- Based on a common easily maintainable code base
- Separated into an app and the daiquiri library
- Features:
  - SQL web interface to relational databases
  - User space
  - User management and user registration work flow
  - Metadata management and access control
  - First glance in-browser plotting
  - Table download and file service
- Employs VO protocols and standards

# Public Databases hosted @AIP

- RAVE database query interface for DR3 (Siebert et al., 2011)
- MultiDark Database (Riebe et al., 2011)



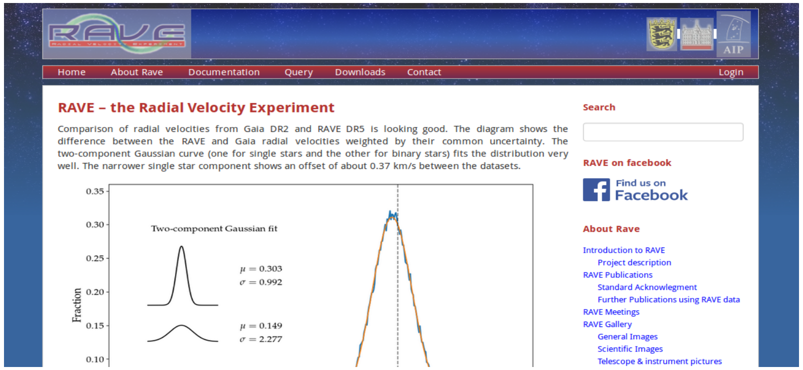
The CosmoSim database provides results from cosmological simulations performed within different projects: MultiDark and Bolshoi, CLUES, and Galaxies.

The Spanish MultiDark Consolider project supports efforts to identify and detect matter, including dark matter simulations of the universe.

Available now for the MDPL2 simulation - galaxy catalogs contain galaxy properties from different semi-analytical codes.

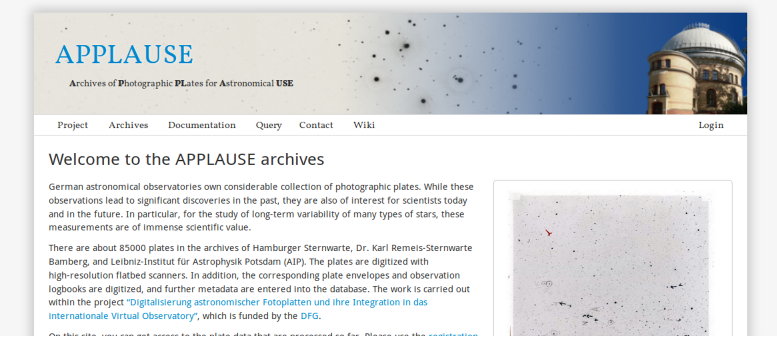
The CLUES project produces constrained simulations of the local universe, partially with gas and star formation.

CosmoSim.org is hosted and maintained by the Leibniz-Institute for Astrophysics Potsdam (AIP).



RAVE – the Radial Velocity Experiment

Comparison of radial velocities from Gaia DR2 and RAVE DR5 is looking good. The diagram shows the difference between the RAVE and Gaia radial velocities weighted by their common uncertainty. The two-component Gaussian curve (one for single stars and the other for binary stars) fits the distribution very well. The narrower single star component shows an offset of about 0.37 km/s between the datasets.

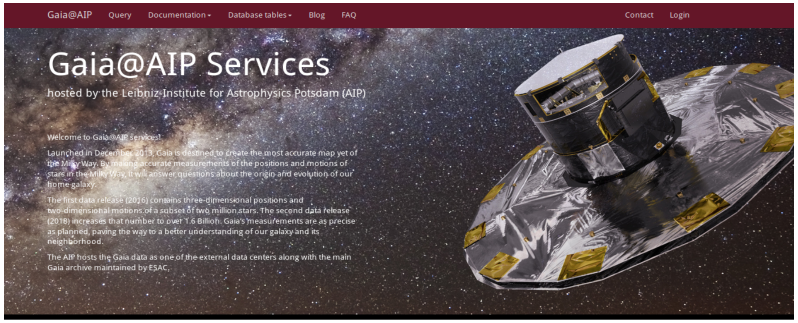


Welcome to the APPLAUSE archives

German astronomical observatories own considerable collection of photographic plates. While these observations lead to significant discoveries in the past, they are also of interest for scientists today and in the future. In particular, for the study of long-term variability of many types of stars, these measurements are of immense scientific value.

There are about 85000 plates in the archives of Hamburger Sternwarte, Dr. Karl Remeis-Sternwarte Bamberg, and Leibniz-Institut für Astrophysik Potsdam (AIP). The plates are digitized with high-resolution flatbed scanners. In addition, the corresponding plate envelopes and observation logbooks are digitized, and further metadata are entered into the database. The work is carried out within the project "Digitalisierung astronomischer Fotoplatten und ihre Integration in das internationale Virtual Observatory", which is funded by the DFG.

On this site, you can get access to the plate data that are processed so far. Please use the registration



Gaia@AIP Services

hosted by the Leibniz-Institute for Astrophysics Potsdam (AIP)

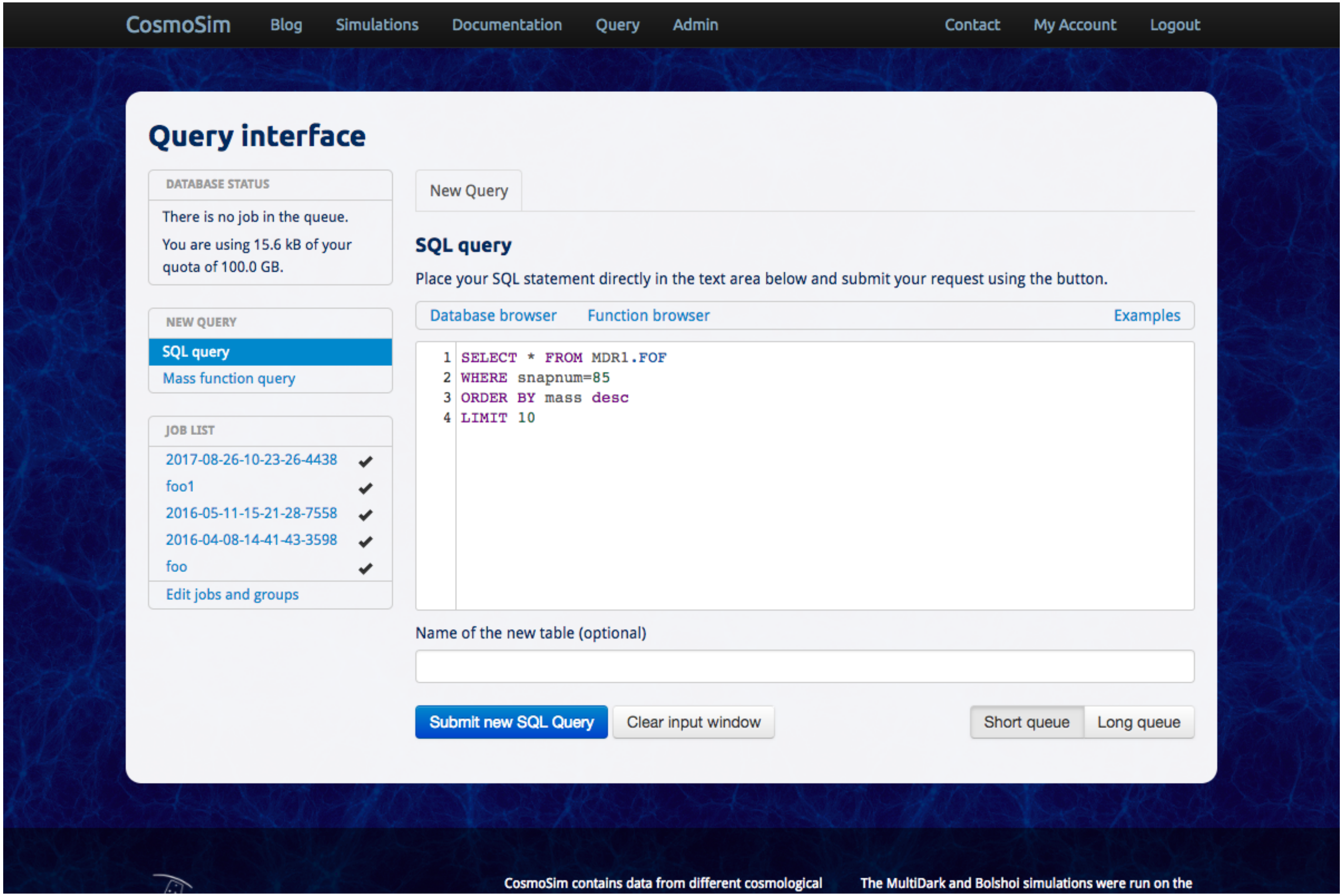
Welcome to Gaia@AIP services!

Launched in December 2013, Gaia is destined to create the most accurate map yet of the Milky Way. By making accurate measurements of the positions and motions of stars in the Milky Way, it will answer questions about the origin and evolution of our home galaxy.

The first data release (2016) contains three-dimensional positions and two-dimensional motions of a subset of two million stars. The second data release (2018) increases that number to over 1.6 Billion. Gaia's measurements are as precise as planned, paving the way to a better understanding of our galaxy and its neighborhood.

The AIP hosts the Gaia data as one of the external data centers along with the main Gaia archive maintained by ESAC.

- RAVE (2013), CosmoSim (2014), APPLAUSE (2015), Gaia@AIP (2016)
- Future: 4MOST public archive



# Query interface

DATABASE STATUS

There is no job in the queue.

You are using 15.6 kB of your quota of 100.0 GB.

NEW QUERY

- SQL query
- Mass function query

JOB LIST

- 2017-08-26-10-23-26-4438 ✓
- foo1 ✓
- 2016-05-11-15-21-28-7558 ✓
- 2016-04-08-14-41-43-3598 ✓
- foo ✓
- Edit jobs and groups

## SQL query

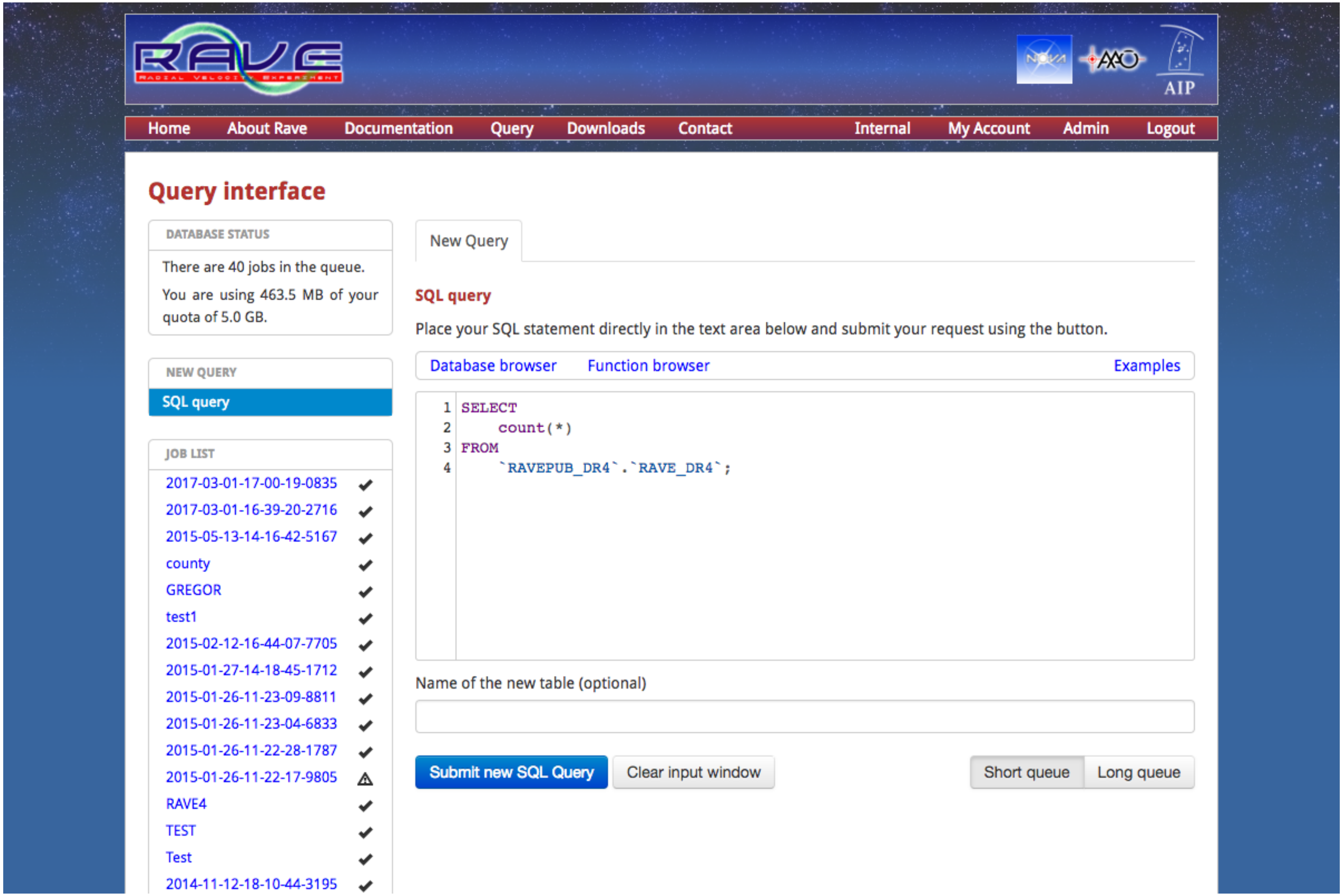
Place your SQL statement directly in the text area below and submit your request using the button.

```
SELECT * FROM MDR1.FOF
WHERE snapnum=85
ORDER BY mass desc
LIMIT 10
```

Name of the new table (optional)

CosmoSim contains data from different cosmological

The MultiDark and Bolshoi simulations were run on the



## Query interface

DATABASE STATUS

There are 40 jobs in the queue.

You are using 463.5 MB of your quota of 5.0 GB.

JOB LIST

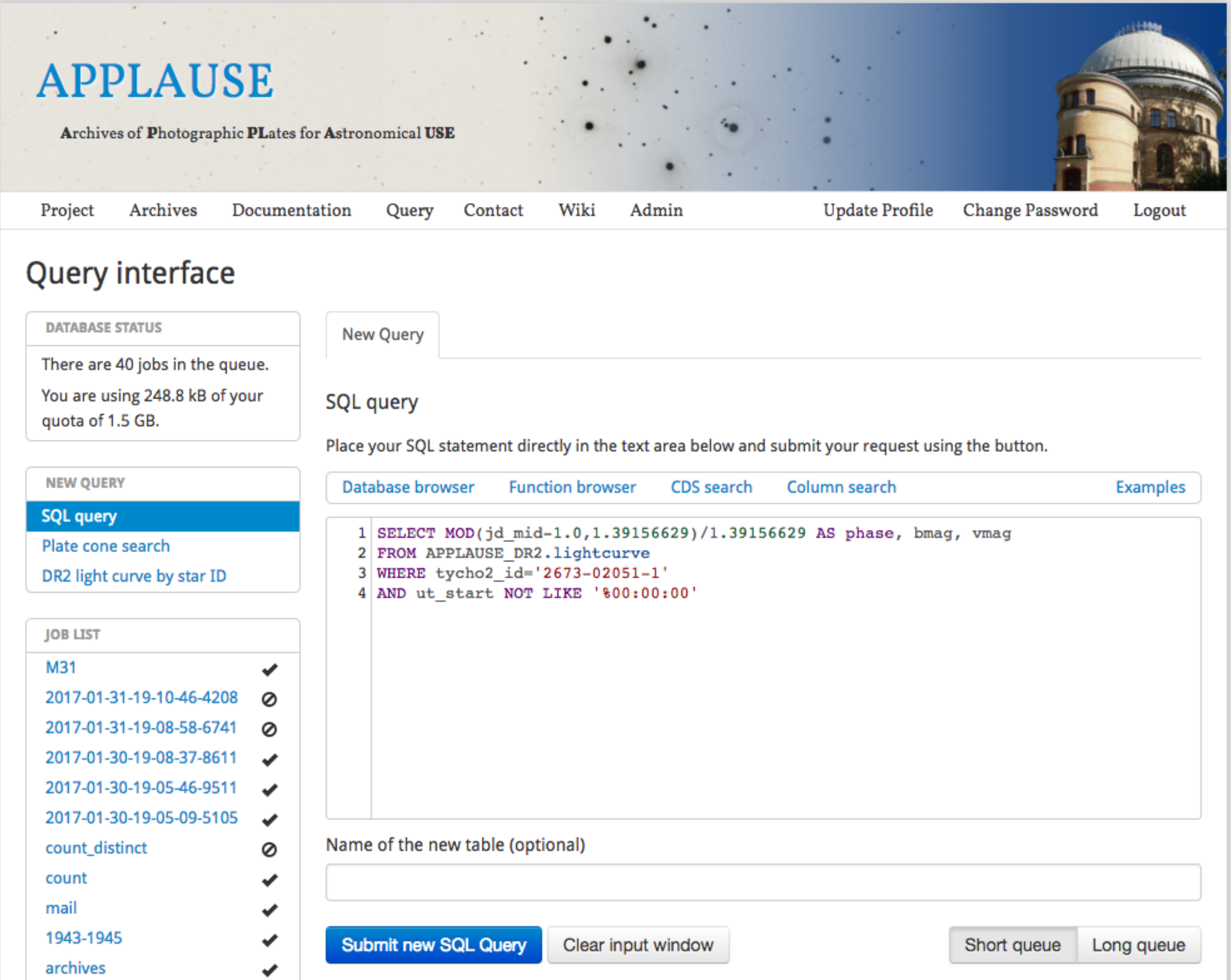
- 2017-03-01-17-00-19-0835 ✓
- 2017-03-01-16-39-20-2716 ✓
- 2015-05-13-14-16-42-5167 ✓
- county ✓
- GREGOR ✓
- test1 ✓
- 2015-02-12-16-44-07-7705 ✓
- 2015-01-27-14-18-45-1712 ✓
- 2015-01-26-11-23-09-8811 ✓
- 2015-01-26-11-23-04-6833 ✓
- 2015-01-26-11-22-28-1787 ✓
- 2015-01-26-11-22-17-9805 ⚠
- RAVE4 ✓
- TEST ✓
- Test ✓
- 2014-11-12-18-10-44-3195 ✓

SQL query

Place your SQL statement directly in the text area below and submit your request using the button.

```
SELECT
    count(*)
FROM
    `RAVEPUB_DR4`.`RAVE_DR4`;
```

Name of the new table (optional)



## Query interface

There are 40 jobs in the queue.

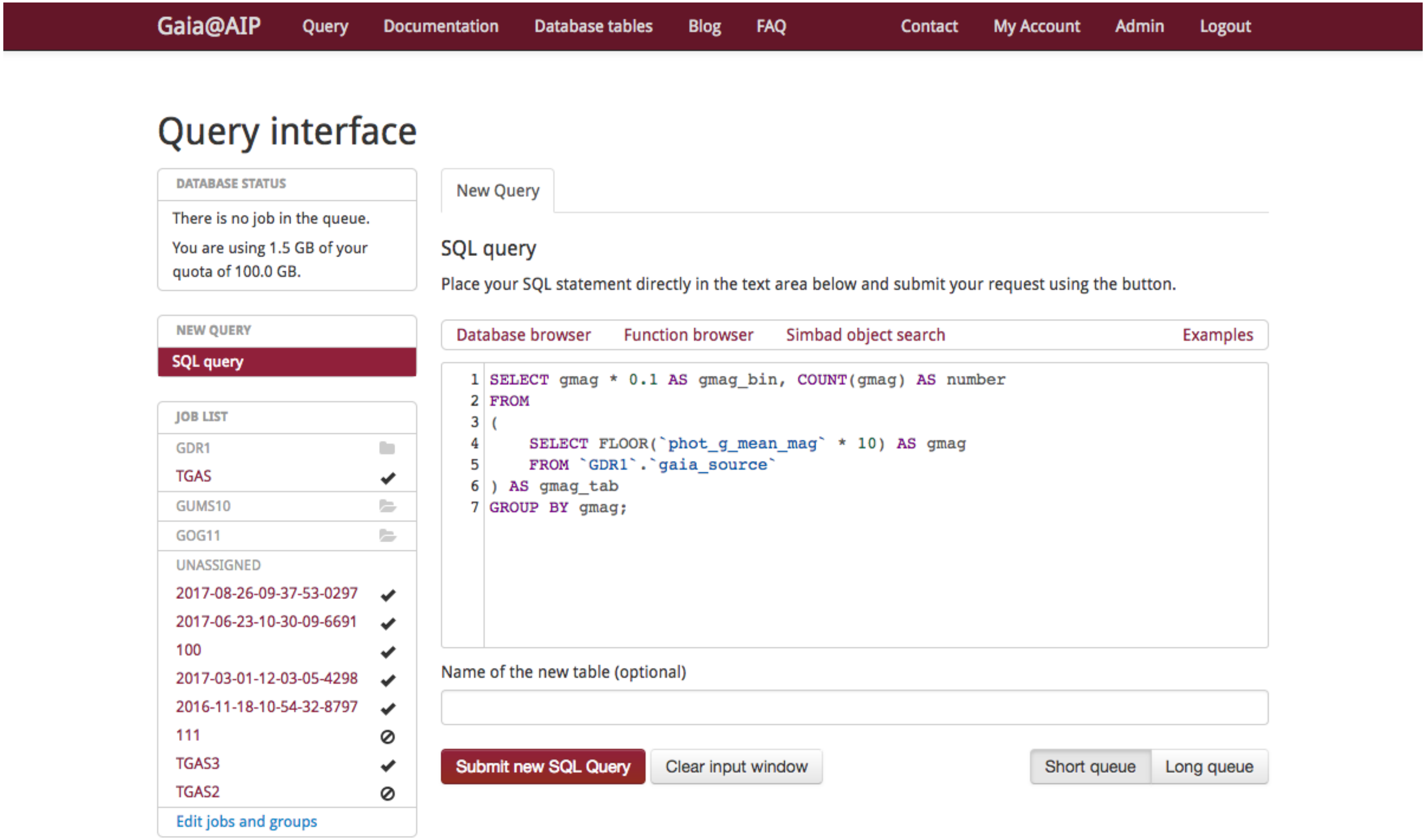
You are using 248.8 kB of your quota of 1.5 GB.

### SQL query

Place your SQL statement directly in the text area below and submit your request using the button.

```
SELECT MOD(jd_mid-1.0,1.39156629)/1.39156629 AS phase, bmag, vmag
FROM APPLAUSE_DR2.lightcurve
WHERE tycho2_id='2673-02051-1'
AND ut_start NOT LIKE '%00:00:00'
```

Name of the new table (optional)



# Query interface

DATABASE STATUS

There is no job in the queue.

You are using 1.5 GB of your quota of 100.0 GB.

NEW QUERY

SQL query

JOB LIST

GDR1

TGAS ✓

GUMS10

GOG11

UNASSIGNED

2017-08-26-09-37-53-0297 ✓

2017-06-23-10-30-09-6691 ✓

100 ✓

2017-03-01-12-03-05-4298 ✓

2016-11-18-10-54-32-8797 ✓

111 ⊘

TGAS3 ✓

TGAS2 ⊘

Edit jobs and groups

New Query

## SQL query

Place your SQL statement directly in the text area below and submit your request using the button.

```sql
SELECT gmag * 0.1 AS gmag_bin, COUNT(gmag) AS number
FROM
(
    SELECT FLOOR(`phot_g_mean_mag` * 10) AS gmag
    FROM `GDR1`.`gaia_source`
) AS gmag_tab
GROUP BY gmag;
```

Name of the new table (optional)

# Language and framework

## Python and Django

- Python: todays preferred scripting language, widely used in astronomy
- Django: full MVC framework with everything included, huge community
- Django REST framework: de-facto standard for REST interfaces in python
- Django-allauth: local and social authentication, registration work flows
- astropy: community python library for Astronomy
- ...

# Front-end

## AngularJS and Bootstrap

- AngularJS 1: awesome since Daiquiri v1
- Bootstrap 3: responsive layout, mobile friendly

# Understanding queries

## Queryparser based on Antlr

- Antlr: parser generator for structured text or binary files
- queryparser: Antlr generated python code to parse query strings
- ADQL translator using ADQL grammar to translate to MySQL or PostgresQL syntax
- MySQL and PostgresQL parsers using MySQL / PostgresQL grammar for parsing an SQL query
- open source and available on GitHub and PyPI (python2, python3)
- Using mysql_sphere to translate ADQL functions into MySQL
- developed and maintained by Gal Matijevic (AIP)

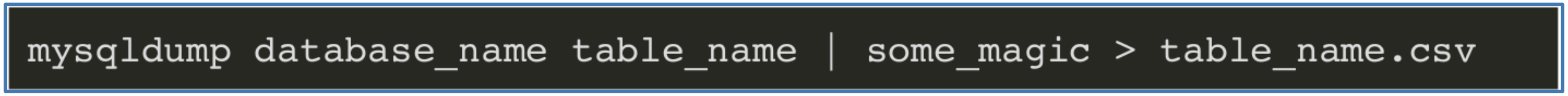
# Asynchronous jobs

## Celery and RabbitMQ

- Celery: asynchronous task queue queue in Python, widely adopted
- RabbitMQ: message broker in Erlang
- redis: in-memory data structure store for task results, can also be used for
- caching
- systemd: new init system for Linux, make it easy to deploy daemons

# Downloading tables

## Celery and RabbitMQ and old-school unix pipes



```
mysqldump database_name table_name | some_magic > table_name.csv
```

# Features

## Implemented

- SQL query interface (with examples, job list, plotting, …)
- Full ADQL + SQL syntax of PostgresQL – pgSphere integration
- Customizable data query forms
- Contact messages + ~mangement for the support staff
- DOI integration and landing pages for databases and tables
- File service and zip-archive creation
- Registration and log-in using Oauth2 (facebook, twitter, GitHub, Google, ORCID)
- WordPress as CMS for project presentation and documentation
- VO protocols:
  - Data Access Layer Interface (DALI)
  - Table Access Protocol (TAP)
  - Universal Worker Service Pattern (UWS)
  - Cone search
- File access, filtering and download for observatory archives (MUSE WIDE)
- Cut-out service for images and datacubes

# Features

## Upcoming

- FITS tables download
- Use of sharded databases (paqu v2)
- Management of project/collaboration meetings
- VO protocols:
    - Simple Image Access (SIA)
    - Simple Spectral Access (SSA)
    - Provenance Data Model (ProvSAP, ProvTAP)

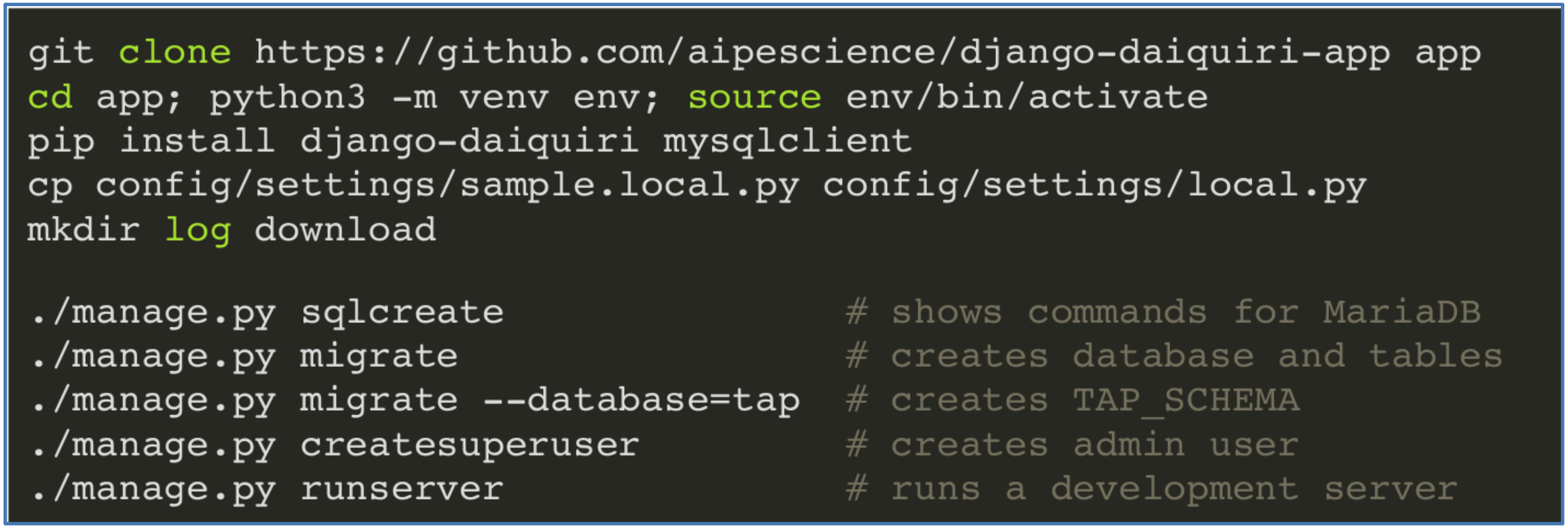
# Try django-daiquiri!

## as a user

Production version of the Gaia@AIP Services: [https://gaia.aip.de/](https://gaia.aip.de/)

## as a provider

README: https://github.com/aipescience/django-daiquiri/blob/master/README.rst



```
git clone https://github.com/aipescience/django-daiquiri-app app
cd app; python3 -m venv env; source env/bin/activate
pip install django-daiquiri mysqlclient
cp config/settings/sample.local.py config/settings/local.py
mkdir log download

./manage.py sqlcreate                # shows commands for MariaDB
./manage.py migrate                  # creates database and tables
./manage.py migrate --database=tap   # creates TAP_SCHEMA
./manage.py createsuperuser          # creates admin user
./manage.py runserver                # runs a development server
```

# Daiquiri v2

**django-daiquiri (since 2016)**

Python (>= 2.7 and >= 3.4) using the Django framework (1.11)

Responsive front-end written in AngularJS 1 and Bootstrap 3

Using:

- MariaDB 10.1 or PostgresQL 9.6/10
- queryparser and Antlr to parse and translate queries
- Celery, RabbitMQ, redis and systemd for asynchronous tasks (not only queries)
- Old school unix pipes to create files using mysqldump or pgdump

open source and available on

GitHub and

PyPy

to deploy daemons

AIP

Leibniz-Institut für
Astrophysik Potsdam

# Questions?

Anastasia Galkin

agalkin@aip.de

github.com/aipescience

escience.aip.de