



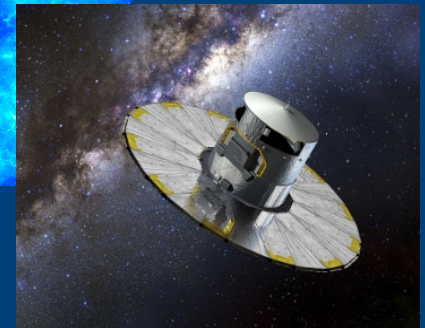
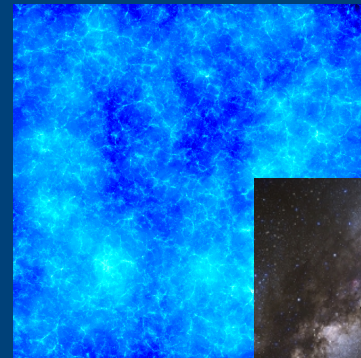
Leibniz-Institut für
Astrophysik Potsdam

Hosting astronomical data in sharded SQL databases

Parallel MariaDB databases for terrabytes of data

Anastasia Galkin
Kristin Riebe
Jochen Klar
Harry Enke

ADASSXXVI, 18.10.2016

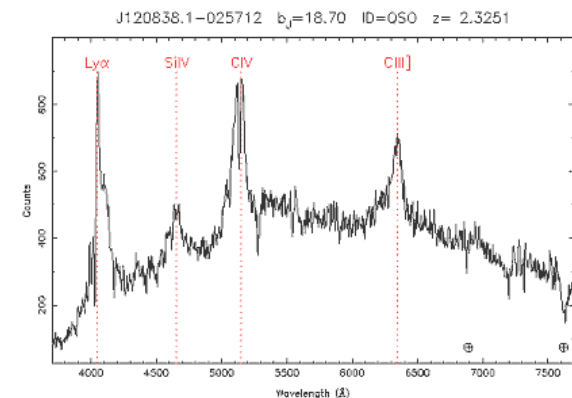
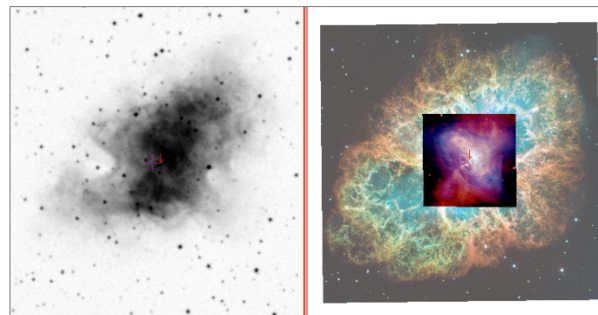
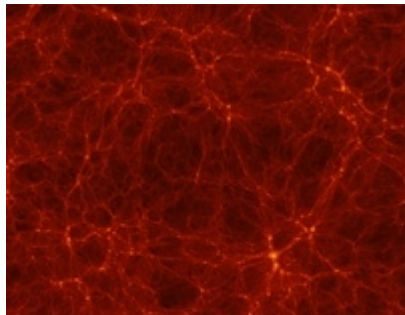
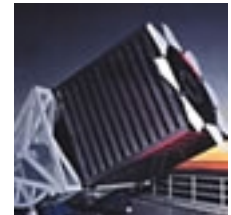


Outline

- Astrophysical data and SQL in astrophysics
- The setup: Daiquiri, Paqu and the parallel MariaDB databases
- Future development

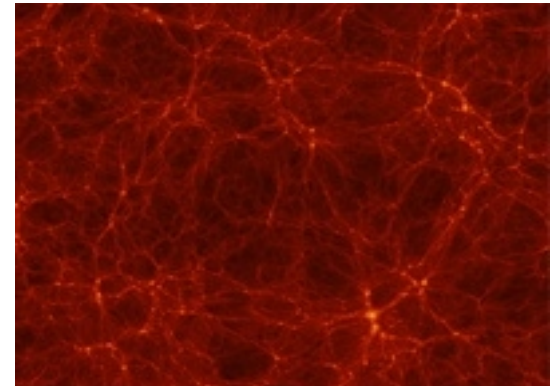
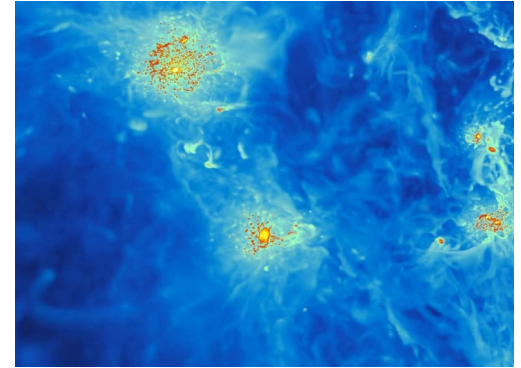
Terabytes of astrophysical data

- Observational data
 - Gaia@AIP gaia.aip.de, incl. Tycho-2 and RAVE for crossmatch purposes
 - RAVE – radial velocity project
 - APPLAUSE – historical photometric plates
- Simulations output
 - CosmoSim.org: MultiDark, Bolshoi



Why SQL?

- Very clear structured data
- Retrieve specific subset
- Need for complex queries with astrophysical functions
- Quasi-standard in astronomy





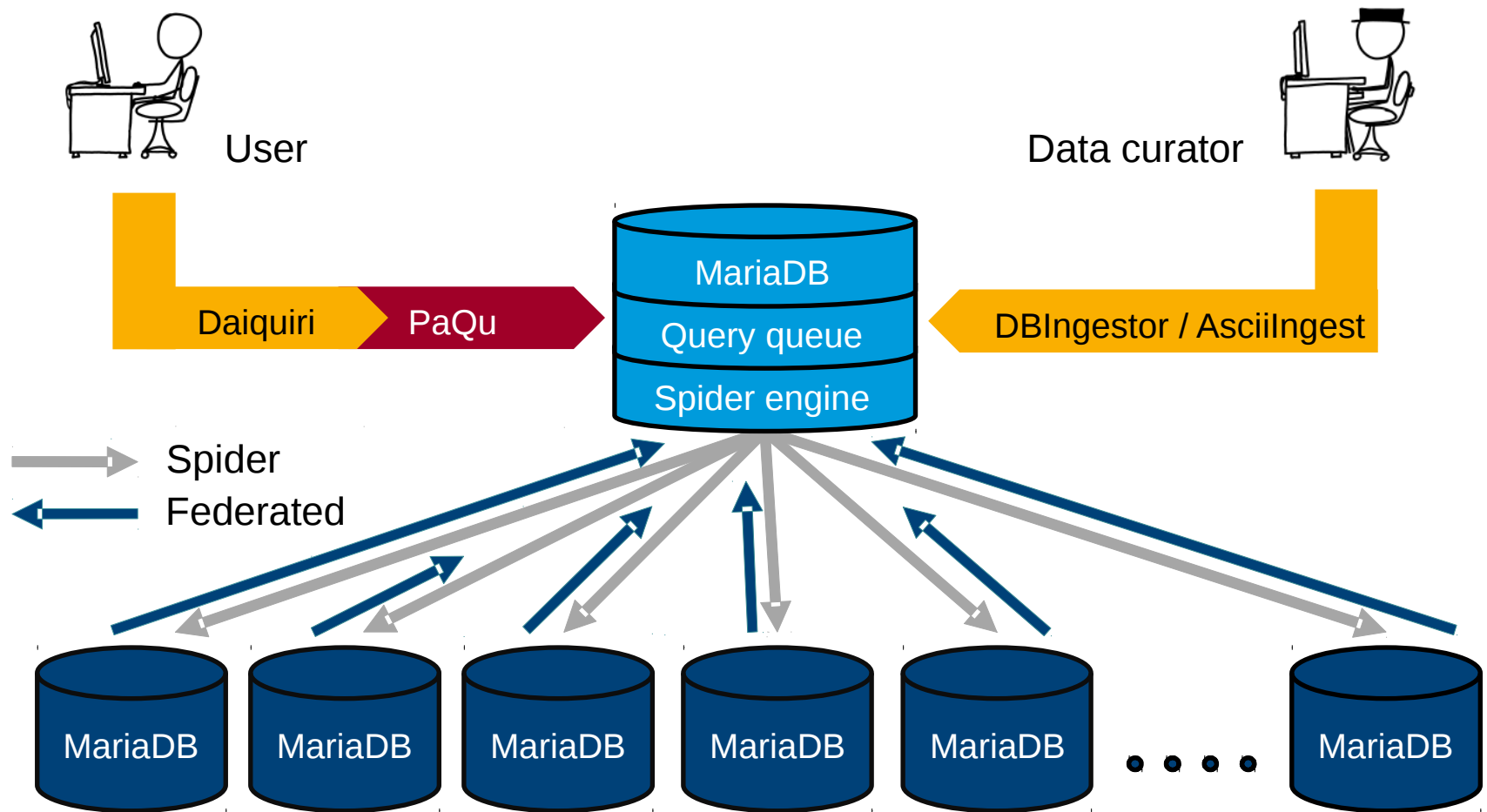
It's a question of size...

- Currently hosted 45TB public available data
- Number of columns: up to 60
- Number of entries:
 - CosmoSim has ca. 40TB with 354 billion rows
 - Gaia DR1 244GB with 2 billion rows (TGAS 2 million)

... and the complexity of the queries.

- Selection of specific properties on billions of rows
- Combined with specific astrophysical functions

The full picture



Queries running in parallel

MariaDB

- Data is distributed on 10 shard nodes with MyISAM engine
- orchestrated by a head node with Spider engine
- Random number plugin



PaQu

- reformulates MySQL queries
 - joins, aggregates, functions
 - e.g. count rows
 - count on each node
 - sum on head node
- head node collects data via federated tables



Daiquiri

A customizable framework for database publication

- SQL queries
 - SQL query assistance
 - Can be used with PaQu with the Spider engine setup
 - Query result table viewer, plotting tool
- UWS interface
- User database space
- Download of data in different formats
- Project documentation via Wordpress
- Administration tools:
 - Database management
 - User management
 - Contact messages
 - Meetings management



CosmoSim


[Home](#) [About Rave](#) [Query](#) [Downloads](#) [Contact](#) [Intranet](#) [Login](#)

RAVE – the Radial Velocity Experiment

Search

APPLAUSE



4MOST — 4-metre Multi-Object Spectroscopic Telescope

[Home](#) [News](#) [Science](#) [Surveys](#) [Facility](#) [Publications](#) [Consortium](#) [Team](#) [Login](#)

Overview

[Gaia@AIP](#) [Query](#) [Documentation](#) [Database tables](#) [Blog](#) [FAQ](#) [Contact](#) [Login](#)

Gaia@AIP Services

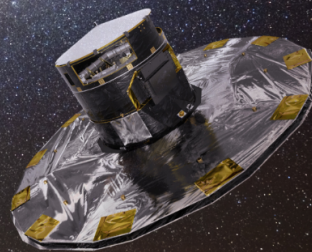
hosted by the Leibniz-Institute for Astrophysics Potsdam (AIP)

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 intermediate data release, containing among other things three-dimensional positions and two-dimensional motions of a subset of two million stars, demonstrates that Gaia's measurements are as precise as planned, paving the way to create the full map of one billion stars to be released towards the end of 2017.

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



Gaia first data release (GDR1)

The first release catalog includes positions and G-band magnitudes for more than a billion sources. The five-parameter astrometric solution (positions, parallaxes and proper motions) is available for over 2 million stars that are shared between Gaia and Tycho-2 and Hipparcos catalogs. More information about the release can be found [here](#).

RAVE DR5 and TYCHO2

RAVE Data Release 5 crossmatched with TGAS is available now. Additionally, Gaia@AIP offers TYCHO2 catalog for the crossmatch purposes.

Getting started

- Register via the [Registration Form](#).
- After registration, you can use the [Query Form](#) using plain SQL.
- Use the example queries in the Query Form to get started.
- [Contact us](#) if you need further assistance.

- Gaia@AIP
- CosmoSim
- RAVE
- APPLAUSE plate archive
- CosmoDB @ Tartu
- JUBILEE @ Madrid
- GREGOR project
- CLUES user management
- Workshops: Dwarfs, 4Most
- Coming next: 4Most

Future development

- Daiquiri – reloaded in python with Django framework
- Improved rewrite process for parallel queries
- Improved queue for query jobs
- Nodes daemons communication to improve the performance
- Frontend: MySQL, ADQL
- Backends: MySQL/MariaDB (sharding), Postgres



Coming up in 2017

- Deployment at AIP: newest MariaDB versions with ARIA, SPIDER and FEDERATED database engines

Questions?

escience.aip.de

Anastasia Galkin agalkin@aip.de

Gaia@AIP gaia.aip.de

www.cosmosim.org