

## First ASTERICS-OBELICS Workshop, 12-14 December 2016



Contribution ID: 68

Type: **not specified**

### Tutorial: MonetDB in the context of the new (high-cadence) facilities

*Tuesday, 13 December 2016 14:55 (30 minutes)*

Optical and radio telescopes planned for the near future will generate enormous data streams to meet their scientific goals, e.g., high-speed all-sky surveys, searches for rapid transient and variable sources, cataloguing the multi-millions of sources and their thousands of measurements. These high-cadence instruments challenge many aspects of contemporary data management systems. However, no database system exists yet, that keeps pace with and stores these huge amounts of scientific data, nor would it be capable of querying the data scientifically with acceptable response times.

The open source relational database management system MonetDB is built upon column-store technologies which have many advantages in different Big Data science domains. MonetDB is a mature main-memory database system, compliant to the SQL2003 standard, and has APIs to C, Java, Python and R. The ease of extending its functionality with UDFs written in SQL, C, R and recently Python are other strong points. Furthermore, support of SQL management of external data makes loading of binary data, e.g., FITS files, extremely fast.

With the experience and lessons learnt from high-cadence radio astronomy (LOFAR) further development was triggered to meet the database needs that characterise the optical regime where source densities are orders of magnitude larger. MonetDB is a key component in the automated full-source pipeline for the optical BlackGEM telescopes, currently under construction.

In this tutorial talk I will give an overview of the properties of column stores, the fundamental differences between MonetDB and the well-known mainstream row-stores and how it is being used in astronomical pipelines and archives. I will discuss an embedded implementation of MonetDB in the experimental infrastructure of the SciLens platform, a tiered 300+ nodes locally distributed cluster focussed on massive I/O, instead of raw computing power, where remote and merge tables play a crucial role. In the context of BlackGEM, I will show examples and promising results for source cross-matching using alternative multi-dimensional tree indexes built inside the database engine.

**Presenter:** Dr SCHEERS, Bart (Postdoctoral researcher in the Database Architecture, CWI Amsterdam)