MonetDB/SQL Meets SkyServer: the Challenges of a Scientific Database

M. Ivanova *  N. Nes  R. Goncalves  M. Kersten
Centrum voor Wiskunde en Informatica
Kruislaan 413, Amsterdam, The Netherlands

Scientific databases pose new challenges to the database community. SkyServer project (http://cas.sdss.org) is a pioneering work establishing collaboration between database and astronomy research with the goal to provide public access to the Sloan Digital Sky Survey warehouse for astronomers and a wide public. Leveraging on MS SQL Server, the project convincingly showed the potentials of a SQL-based solution for a real-life astronomy application.

The focus of this presentation is the MonetDB/SkyServer project which has been set up as an experimental study of large scale databases based on the MonetDB platform (http://www.monetdb.nl), an open-source database system developed for over a decade at CWI, Amsterdam. The goal of the project is to provide a counter proof for the MS SQL Server implementation and to act as a platform for developing novel algorithms to support the data management challenges in the domain of scientific databases.

In this presentation we describe our experiences in porting the SkyServer application to MonetDB, the functional improvements and adaptations needed to get the original application up and running. Experiments with a representative query workload show performance figures competitive to the reference platform and demonstrate that the column store approach of MonetDB is promising for the scientific domain. We also present an analysis of one month SkyServer query traces that gives insights about the system usage and shows opportunities for performance improvements.

Finally, we present a number of open research questions inspired by the experience in SkyServer porting. Examples include multi-query optimization, physical organization techniques differentiating data depending on the usage frequency, optimization of SQL functions, and efficient support for data mining tasks augmenting the astronomical research.

The presentation is based on the following publication:

*Contact author (M.Ivanova@cwi.nl), post-doctoral researcher at CWI