

## Cross-matching the sky using an RDBMS cluster

*Wednesday 29 May 2013 10:30 (20 minutes)*

In the demo I will present the new generation SkyQuery, a tool designed for astronomers working on multi-wavelength projects that require cross-matching celestial objects across multiple multi-TB catalogs. Cross-match problems are formulated in a slightly extended version of SQL. Stars and galaxies are associated based on spherical coordinates affected by measurement errors. A Bayesian approach is used to determine the appropriate cuts on spherical distances when evaluating possible matches. Cross-matching of more than two catalogs is done iteratively, one catalog at a time, which makes our algorithm scalable. We built our system on a cluster of servers running Microsoft SQL Server. A generic-purpose data warehouse API was developed to manage the cluster servers and execute distributed and partitioned queries in parallel on multiple machines for load balancing. The API includes modules for SQL parsing, workflow management, job queuing, distributed query execution and web-based user interfaces for system managers and end-users.

**Author:** DOBOS, Laszlo (E)

**Co-authors:** Prof. SZALAY, Alexander (Johns Hopkins University); Prof. CSABAI, Istvan (Eotvos Lorand University); Dr BUDAVARI, Tamas (Johns Hopkins University)

**Presenter:** DOBOS, Laszlo (E)

**Session Classification:** Big Data in Astronomy / Astrophysics