Databases for the *Pi of the Sky* experiment

Marek Biskup[a], Lech Mankiewicz[b], Marcin Sokołowski[c], Grzegorz Wrochna[c]

[a] Warsaw University, Institute of Experimental Physics, Hoża 69, 00-681 Warsaw, Poland
[b] Center for Theoretical Physics, Polish Academy of Sciences, Lotników 32-46, 02-668 Warsaw, Poland
[c] A. Soltan Institute for Nuclear Studies, Hoża 69, 00-681 Warsaw, Poland

ABSTRACT

This paper describes databases used in the *Pi of the Sky* experiment, particularly the databases of stars and observations. The *Pi of the Sky* experiment, located at Las Campanas Observatory in Chile searches for rapidly changing optical objects such as Gamma Ray Bursts (GRB). The system consists of two CCD cameras placed on paralactic mount and operated by a PC, equipped with dedicated software. The data acquired is reduced and mainly the brightness of stars is stored in the database. A web interface for the database has been created. It allows easy and quick access to the experiment's data and to other star catalogs, which are also included in our databases.

Keywords: optical transients, stars' database, web interface, *Pi of the Sky* experiment, light curve, sky survey.

1. INTRODUCTION

The *Pi of the Sky* experiment [1] looks for short optical transients in the sky caused by astrophysical objects. For that purpose specialized Charge Coupling Device (CCD) cameras have been designed and build. In the current phase there are two cameras observing the sky each night and taking pictures every 12 seconds. The exposition itself lasts 10 seconds. In the remaining two seconds data transfer is performed. On average there are 11 hours of data acquisition per night, which amounts to over three thousand frames (of size 8MB each) per camera. The pictures are analyzed online and searched for pointlike flashes. Precise sky pictures are useful for studying other astronomical phenomena as well. However this large stream of data cannot be stored on a computer system of reasonable size and cost. Even if it were possible, data analysis on raw sky images would be rather complicated and it would take longer time than the data acquisition. Thus the data collected during one year would be processed in at least one year. To reduce the data and to prepare it in a simple, easy to analyze form, the pictures are further processed offline and only the most important information is kept.

The processing works in the following way. Firstly a dark frame is subtracted from the picture. The dark frame is taken with the camera's shutter closed and it represents electronics' noise. Then the picture is divided by a flat frame, pixel by pixel. The flat frame is a image of uniform background so such a division removes the bias caused by unequal sensitivity of pixels. The dark and flat frames are taken every day. After this preparation *photometry* is performed. For every bright spot its coordinates within the CCD chip and brightness are computed. These give input for the *astrometry* process, which translates them to celestial coordinates and magnitude brightness scale. To be more precise, astrometry creates a mapping (a polynomial of 3rd degree) between those scales using some reference stars.

Each frame contains about 10000 stars for which the data (like brightness, celestial coordinates and some more control data) has to be stored. With two cameras, each night we get around 60 million brightness measurements and within a year of observations this number grows to well over 20 billion. The rest of this paper describes how we store those measurements and how we provide the access to the data.

2. THE DATABASE

In this article we present only the database for brightness measurements and stars. This is, taking into account the volume, the biggest part of stored data in the *Pi of the sky* experiment. There is however a lot of other information, like raw frames kept for several days, configuration data, detected events. These will not be described here.

There are several requirements for the way the data is stored:

- reliability (users should be able to extract information at any time),
- scalability (there is considerable amount of observations' data, 2TB, and in the future the volume will constantly grow),
- stability (errors in the software must not cause lost of all the data),
- performance (data should be processed/extracted in reasonable time),
- manageability (the system must be easy to access and extract information),
- security (the system will be accessible not only from our laboratory but also by external users),
- cost (as in every academic project the funds are limited).

Of course some of the requirements can be relaxed. For instance a failure of the system (as long as the data is not lost) will not result in as serious consequences, as in commercial banking systems. Having taken into account all the pros and cons we have decided to use PostgreSQL database [2]. We are, however, considering future migration to other commercial Database Management System which supports distributed data storage. Relational database is better than proprietary software as it allows precise and quick selection and manipulation of the data and, which was on of the key requirements, such a system is much more stable.

There are several ways the data can be accessed. First of all, for experienced internal users and project members there is the SQL command line interface, `psql` (a part of the PostreSQL distribution), in which SQL queries may be typed directly. For external users we have implemented two possible solutions. The first one is a web interface which allows easy and quick visual browsing of all the information. It will be described in details in section 4. The other possibility is to use a proprietary JAVA [3] application, which executes queries on the database and retrieves results. The results can be stored in a number of formats. Currently we support HTML tables and text files, however other formats are planned to be supported.

3. DATABASE SCHEME

The scheme of the database is presented in Figure 1. The table `Superstar` keeps information about all the stars being observed in the experiment. An entry in that table, superstar, is identified by its unique id, has a name, celestial coordinates, type (the field `star_class`) and links to other popular star catalogs, such as Tycho [4] (`tycho_id`), ASAS [5] (`asas_id`), GCVS [6] (`gcvs_id`; `asas_id` and `gcvs_id` are not presented in the picture). There are some other fields which are important for the internals of the experiment, but will not be described here.

There are two cameras in the *Pi of the Sky* experiment (the next stage will involve 32 cameras). The observations for each of them are separated in the database. Each entry in the `Stars` table contains information concerning a star observed by a single camera. The same star observed by the other camera has another entry in the `Stars` table. The reason for this separation is to compute stars' parameters, such as mean brightness, from data of each camera separately. This allows early detection of erroneous measurements and flaws in cameras. The table `stars` is related to the `Superstar` by a many-to-one relation (each star has just one superstar but a superstar has many stars – the same number as the number of cameras, currently two). Among attributes of a star there are mean magnitude (brightness), celestial coordinates (calculated from all the measurements of this star) and id of the camera.

The last table is called `Measurements`. Each row contains information about a single measurement: the brightness, the time of the measurement, the celestial and CCD coordinates, id of the camera and others. Each measurement is related to one entry in the `Stars` table. There are more tables in our database but they are more technical and store information important for the experiment itself and not for external users.

The database is optimized for fast queries and it contains a lot of redundant information, such as mean magnitude of a star (calculated from magnitude of its measurements), mean magnitude of a superstar (weighed magnitude of stars related to this superstar), position of stars and superstars, number of measurements per star, etc. This information is frequently used in queries selecting particular stars for further analysis. Had they not been precalculated, the queries which use these values would have been unbearably slow. This optimization is the reason for having the `star` table instead of only marking each measurement with an `id` of a camera. Other speed optimizations involve indexes, released constraints and tuned queries in applications. These databases are optimized for fast data access and not necessarily for fast loading of new information.

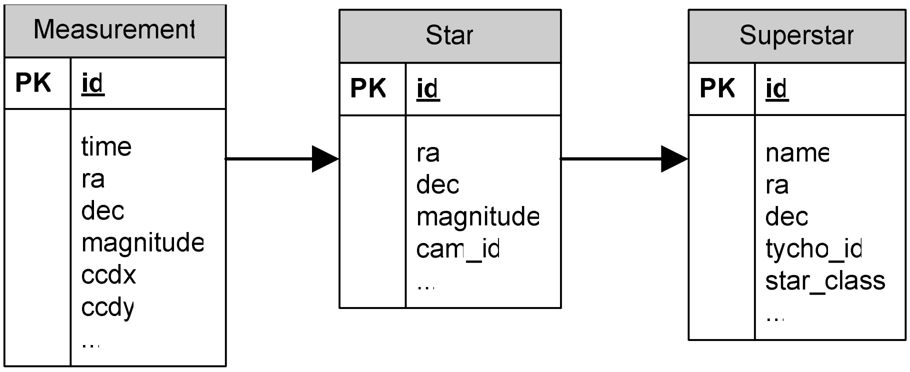

Figure 1. Database scheme

4. EXTERNAL STAR'S CATALOGS

Having a star's database requires a connection to other existing star's catalogs, to be able to compare and publish information about stars. We have chosen Tycho [4], ASAS [5] and GCVS [6] databases as reference catalogs. Tycho is a big database of stars below 11.5 magnitude (around 2.5 million stars). GCVS and ASAS are small databases of variable stars. The reason for taking GCVS and ASAS databases is that we plan to study star's variability on our data. Tycho was chosen as a big reference database. The cross-identification between Pi catalog and Tycho or ASAS was done on the base of coordinates. The tests have proven that the accuracy of measurements is good enough for correct identification of the Tycho and Pi stars. The root-mean-square difference between stars' positions Tycho and Pi databases is around 0.6 arc minutes. The results for ASAS are basically the same. With GCVS there are well known problems with outdated coordinates. We handle them by using existing Tycho-GCVS cross-identification tables.

All the databases share the same schema, so all the software (e.g. the web interface described in the next section) can be used for other catalogs as well, without any modifications.

5. WEB INTERFACE

The web interface is based on *Apache* web server [7], *php* server scripting language [8], *Smarty* template engine [9] and *javascript* browser scripting language. For generating plots we are using the JpGraph [10] library for php.

The main purpose of the web interface is to search star's catalog for stars of given parameters such as coordinates, magnitude, dispersion of the magnitude, number of observations and period (Figure 2). When searching the user has to specify ranges of those parameters. There are three ways in which the stars can be presented. The first one is just the number of stars found. The second is a list of stars limited to a specific number of stars per page (printing thousands of stars would overload the server as well as the browser). The table can be sorted in several ways, e.g. by magnitude, number of measurements, star's name. The last possibility of presenting results is to plot two diagrams (Figure 3): sky map and dispersion plot. The sky map is a plot with celestial coordinates `Ra` and `Dec` on the axes. The dispersion plot is a plot of root-mean-square error of magnitude's measurements versus the magnitude itself. The latter is helpful when searching for variable stars. The plots as well, for efficiency reason, can hold no more than 1000 stars. The rest of the stars will just be ignored. Both plots are interactive. If the user selects an area by drawing a rectangle, the plot will be zoomed in, i.e. a new query with new parameters will be executed on the server.

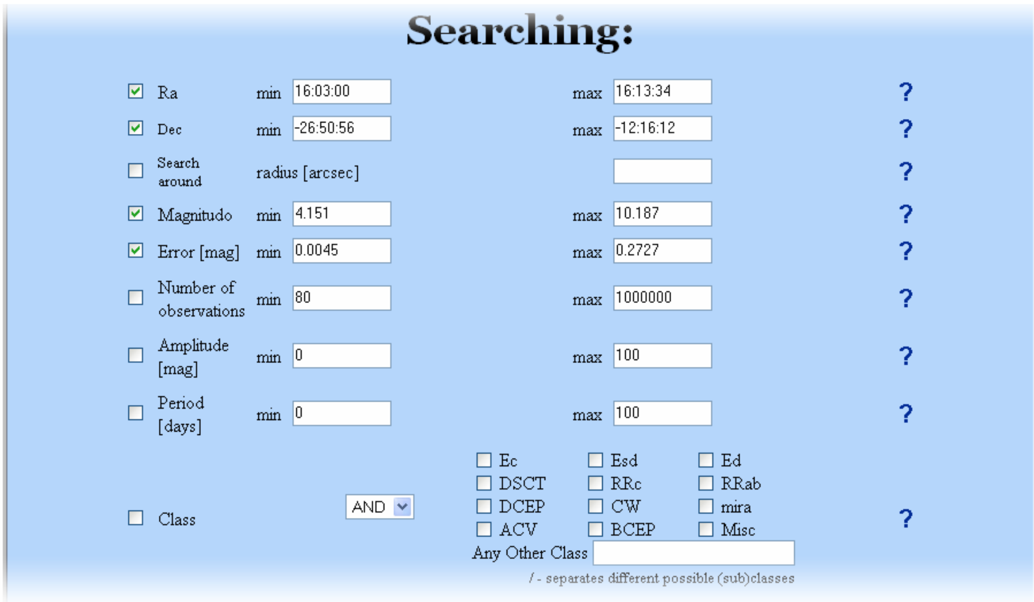

Figure 2. Web interface – search form

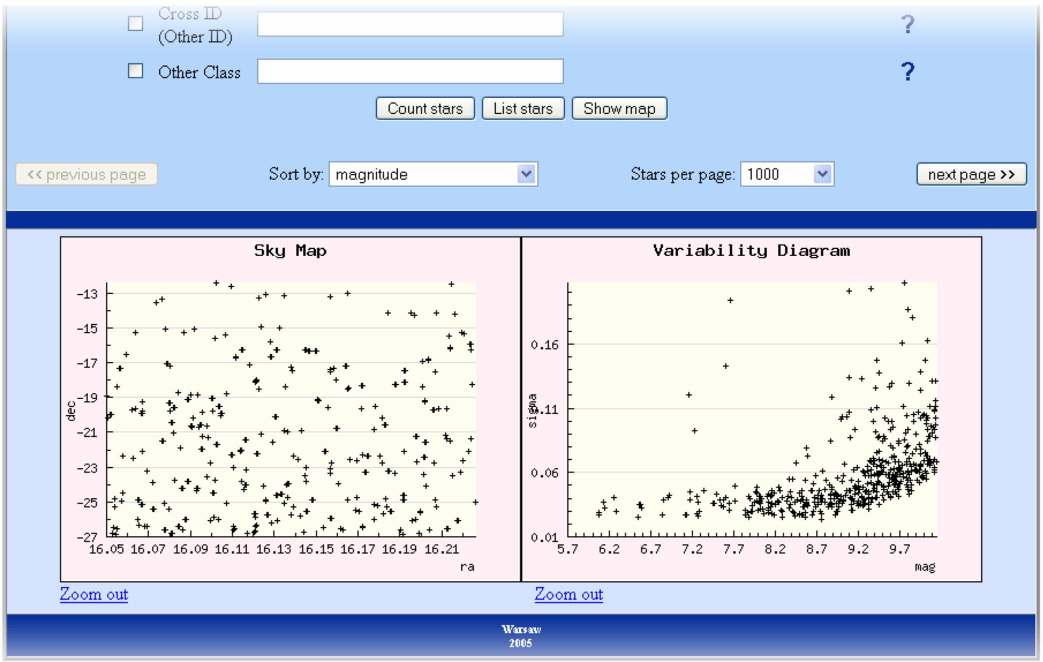

Figure 3. Web interface - sky map and variability diagram

After clicking on a star, either in the list or in any of the plots, the user will be directed to a web page dedicated for that star. The main part of that page is the light curve, i.e. a plot of measured magnitude versus time (Figure 4). To make the analysis easier, the plot shows also the measurements of the other camera. This allows an easy check of results from the cameras. The light curve is zoomable as well – after selecting a region it will be magnified. From the plot we can extract the data as a simple HTML table with just time and measured magnitude, as a larger table with more details or as a macro for the ROOT framework [11] which plots the light curve in a nice, ready-for-publication way. The plot may also be folded modulo a given period. This feature is useful when studying periodic variables. Apart from the plot there is more information about the star displayed. For instance there is a link to the star from the other camera and a link to stars in external databases. That link is actually done in two ways. First of all, there is a script which dynamically looks for stars in other databases in a given radius around the star (in our experiment the accuracy of the position measurements is well below 2 arc seconds, so the radius chosen is 2 arc seconds, which is the angular size of two pixels of our cameras) and shows stars identifiers as links to the interface of the other database. The script is executed after clicking on a link with the name of one of the supported databases. And if links to other databases are included in the `Superstar` table (see section 3), that information is displayed and it points directly to the web page of this star in that other database.

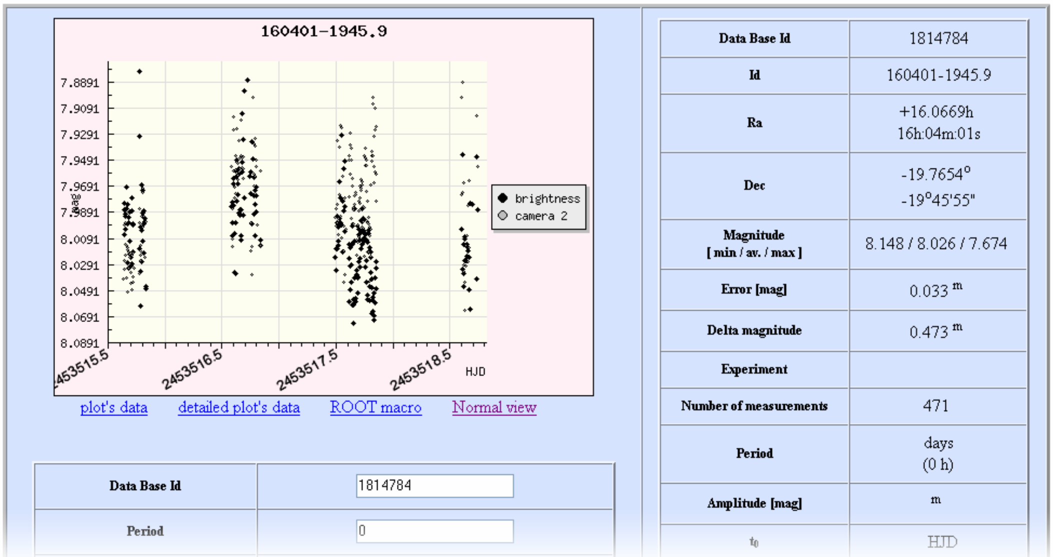

Figure 4. Web inteface - Star's page with a lightcurve

6. SUMMARY

The “Pi of the Sky” experiment is constantly growing and the amount of data gathered is larger and larger. The data require reliable, fast and easy to access storage. We have chosen PostgreSQL relational database for storing the information and Apache+Php as a platform for the user interface, which provides easy, interactive and visual browsing of all the data. Apart from the data from the *Pi of the Sky* experiment we have adopted existing stars’ catalogs such as Tycho, ASAS, and GCVS, and put them into the same schema. The advantage is that we are able to compare our results with other databases and we can use the same analysis software for processing data of every database. The whole system has proven to be efficient, comfortable to use and reliable.

7. ACKNOWLEDGEMENS

This work was financed by Polish Ministry of Science in 2005-2006 as a research project. We are very grateful to B. Paczyński and G. Pojmański for encouraging and many practical advices. We would like to tank the staff of the Las Campanas Observatory for their help during the installation of the apparatus.

REFERENCES

[1] M. Cwiok, et. al., *Search for Optical Counterparts of Gamma Ray Burst*, in Acta Physica Polonica B, Vol. 37, No. 3, March 2006, page 919
[2] PostgreSQL web page: http://www.postgresql.org
[3] Java web page: http://java.sun.com
[4] E. Hog et. al.,*The Tycho-2 catalogue of the 2.5 Milion Brightest Stars*, Astronomy and Astrophysics, v.355, p.L27-L30 (2000)
[5] Pojmanski, G. 2002*, The All Sky Automated Survey. Catalog of Variable Stars. I. 0h - 6h Quarter of the Southern Hemisphere*. Acta Astronomica, 52, 397.
[6] Kholopov et. al., *Combined General Catalogue of Variable Stars*, 4.1 Ed (II/214A). (1998)
[7] Apache httpd server, http://httpd.apache.org/
[8] Php: Hypertext Preprocessor, http://www.php.net/
[9] Smarty Template Engine, http://smarty.php.net/
[10] Jpgraph: Object-Oriented Graph creating library for PHP, http://www.aditus.nu/jpgraph/
[11] Rene Brun and Fons Rademakers, *ROOT - An Object Oriented Data Analysis Framework*, Proceedings AIHENP'96 Workshop, Lausanne, Sep. 1996, Nucl. Inst. & Meth. in Phys. Res. A 389 (1997) 81-86. See also http://root.cern.ch/.