Astronomical Data Analysis Software and Systems IX
ASP Conference Series, Vol. 216, 2000
N. Manset, C. Veillet, and D. Crabtree, eds.

On-line Access to Very Large Catalogues

S. Derriere, F. Ochsenbein, D. Egret

CDS, Observatoire Astronomique de Strasbourg, UMR 7550, 11 rue de l'Université, 67000 Strasbourg, France

Abstract. Dedicated tools have been developed at CDS in order to face the challenge of fast on-line access to very large astronomical catalogues. Powerful compression methods, keeping the direct access to the data on the basis of their celestial position, allow very fast queries; their usage for CDS services such as VizieR or Aladin is presented.

1. Introduction

Astronomy is entering the era of very large catalogues. Digitizations of photographic plates (USNO, GSC II) or current digital sky surveys (DENIS, 2MASS, Sloan DSS) lead to 10^8—10^9 object catalogues. And the number of parameters for each object (positions, magnitudes, errors, flags, etc.) is also considerably increasing.

Gigabyte is the basic size unit for the resulting catalogues, and it becomes impossible for each astronomer to copy locally the full amount of data. Queries over the net, extracting smaller samples with pertaining information, are thus a requisite for such catalogues.

CDS[1] already provides fast on-line access to some very large catalogues which are fully integrated into other CDS services.

2. Compression and Query Mechanism

2.1. Binary Compression

When running queries on databases of several hundred million objects, response time becomes critical. A fast access to the database requires that i/o be reduced as far as possible, and thus to have a powerful compression tool to downsize the amount of data.

We use, for each catalogue, a dedicated and optimized compression method to store the data into binary files:

- All parameters are first converted into integer numbers, and all unused bits are removed. The records are then grouped by their position on the sky, meaning that contiguous records have quite similar values for α and δ;

[1] `http://cdsweb.u-strasbg.fr/`

- A "reference" position is then stored as a "header" of a group of records, and individual positions are stored as offsets (i.e. very small integers) from the "reference" position;
- Additional flags can be added if they help in compressing the data. For example, if a catalogue contains measurements in several photometric bands which are not necessarily available for each object, we can add a single bit, telling if there are data in the band, to every record. That way, we can avoid coding empty values.

E.g., simply by storing only position offsets, the 6 GB of the USNO A1.0 catalogue (Monet et al. 1997) were compressed down to 3.4 GB (Ochsenbein 1998).

2.2. Query Processing

Most of the queries include a selection based on celestial positions, and thus the compressed binary database is indexed on position. Additional filters (on colors or flags) may be applied to the selected sample if necessary.

When a query to such a large catalogue is submitted to VizieR[2], the CDS database of catalogues and tables (Ochsenbein et al. 2000), a dedicated program handles the request, extracts and decodes the matching data, and sends its output back to VizieR.

An overview of the system is shown in Figure 1.

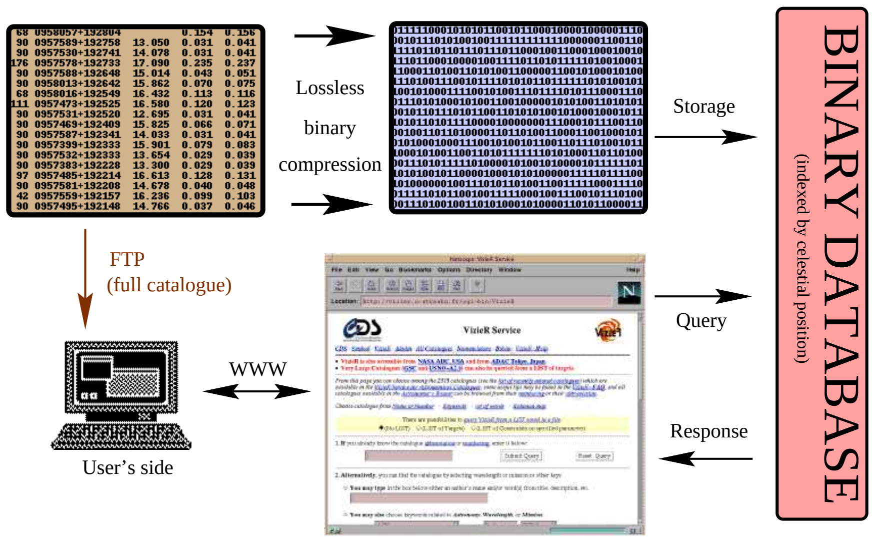

Figure 1. Construction of the database and queries processing.

Table 1 gathers some figures for queries performed on some large catalogues using a Sparc-20 station (72 MHz). On this machine, the complete USNO A1.0 is read in about 40 minutes.

[2]`http://vizier.u-strasbg.fr/`

Table 1. Example of performances achieved.

Catalogue	Record length (bytes)	Reading time (per object)
USNO A1.0	7	$5\mu s$
DENIS PSC	42	$40\mu s$
2MASS PSC	60	$138\mu s$

3. Integration in CDS Services

Thanks to the specific tools that have been developed, very large catalogues can be fully integrated in CDS services such as VizieR or Aladin.

For example, one can select in VizieR all the USNO sources around a given target, with additional constraints on the magnitudes. In Aladin[3], the interactive digitized sky atlas at CDS (Bonnarel et al. 2000), very large catalogues can be overlaid on digitized sky images just like any other catalogue, as illustrated in Figure 2.

The currently (November 1999) available large catalogues include:

- The Guide Star Catalog (Lasker et al. 1990, Russel et al. 1990, Jenkner et al. 1990) (25,258,765 entries), in 3 different versions (GSC 1.1, GSC 1.2, GSC-ACT);
- The USNO catalogues (Monet et al. 1997) versions 1.0 and 2.0 (488,006,860 and 526,280,881 sources, respectively);
- The first DENIS release (17,471,214 point sources);
- The 2MASS (Skrutskie et al. 1997) Spring 1999 release (20,204,378 point sources).

For the DENIS project (Epchtein et al. 1999), beyond the access to the point source catalogue, the CDS provides other on-line services[4], like access to a database of observation strips (quality, number of sources, etc.).

Note that the DENIS and 2MASS catalogues are growing as new data become public.

4. Conclusions

Thanks to dedicated tools, large-size catalogues can now be handled as easily as the more "classical" ones. This will allow further integration of upcoming very large catalogues into CDS services.

[3] `http://aladin.u-strasbg.fr/`

[4] `http://cdsweb.u-strasbg.fr/denis.html`

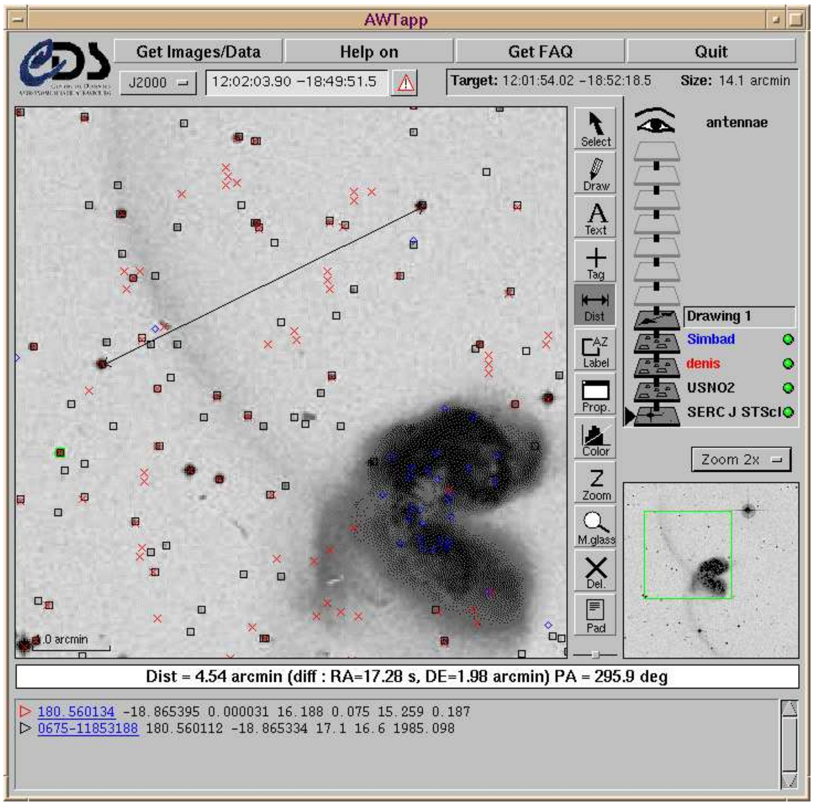

Figure 2. Projection of DENIS (crosses) and USNO A2.0 (squares) sources on an DSS-I blue plate image using Aladin (the field displayed is a 7 arcmin box in the Antennae galaxies).

References

Bonnarel, F., Fernique, P., Bienaymé, O. et al. 2000, A&A, in preparation

Epchtein, N., Deul, E., Derriere, S. et al. 1999, A&A, 349, 236

Jenkner, H., Lasker, B. M., Sturch, C. R. et al. 1990, AJ, 99, 2082

Lasker, B. M., Sturch, C. R., McLean, B. J. et al. 1990, AJ, 99, 2019

Monet, D., Bird, A., Canzian, B. et al. 1997, USNO-A1.0, (U.S. Naval Observatory, Washington DC)

Ochsenbein, F. 1998, in ASP Conf. Ser., Vol. 145, Astronomical Data Analysis Software and Systems VII, ed. R. Albrecht, R. N. Hook, & H. A. Bushouse (San Francisco: ASP), 387

Ochsenbein, F., Bauer, P., Genova, F. et al. 2000, A&A, in preparation

Russell, J. L., Lasker, B. M., McLean, B. J. et al. 1990, AJ, 99, 2059

Skrutskie M. F., Schneider S. E., Stiening R. et al. 1997, Proc. Workshop "The Impact of Large Scale Near-IR Sky Surveys", 25