User Defined Functions for the ISO Post Mission Archive

Ekkehard Wieprecht

Max Planck Institut fuer extraterrestrische Physik, Giessenbachstrasse 1, D-85748 Garching, Germany

Bart Vandenbussche

Katholieke Universiteit Leuven, Instituut voor Sterrenkunde,Celestijnenlaan 200 B, B-3001 Heverlee, Belgium

Erich Wiezorrek, Helmut Steinle, Otto H. Bauer

Max Planck Institut fuer extraterrestrische Physik, Giessenbachstrasse 1, D-85748 Garching, Germany

Christophe Arviset, Jose Hernandez, Arno M. Plug, Richard Saxton

ISO Data Centre, ESA SSD, Villafranca, 28080 Madrid, Spain

Abstract.

A major problem with designing archives in astronomy is defining the data resolution. In most cases the smallest element in an archive is the object 'observation'. Attributes of this object contain meta-information about the observation that should be quickly queriable and are thus stored in the database engine. The data of the observation is stored external to the database engine as one big object. During the design of the archive it is often difficult or even impossible to anticipate what meta-information of an observation should be queriable.

We present a solution for this problem as adopted for the Post Mission Archive of the Infrared Space Observatory (ISO) mission. Expert users can create User Defined Functions (UDFs) that extract virtual attributes from files external to the database engine. These virtual attributes can then further be used in queries inside the database engine.

We discuss the concept of User Defined Functions, system design and implementation.

1. The ISO Post Mission Archive

During ISO's (Kessler M. F. et al.) operational period (November 1995 to May 1998), its four instruments obtained a wide variety of astronomical observations at infrared wavelengths from 2.4-240 microns. About 25,000 observations are stored in the ISO Post Mission Archive (Saxton,R. et al.). This archive is providing on-line access to the data and supporting documentation and software through a WWW interface.

The user interface is HTML and JAVA based and issues SQL queries to the database via jConnect for JDBC (trademark of Sybase Inc.). The basic interface allows to select observations defined by standard criteria such as coordinates, source name, observing mode, exposure time, wavelength etc. Users can also select data by engineering, trend and housekeeping parameters.

2. Virtual data resolution with User Defined Functions

As for all astronomical databases data resolution is a critical issue in the ISO Post Mission Archive. The smallest object in the database is often an 'observation' which is stored as a file external to the Database Management System (DBMS) or as a Binary Large Object (BLOB) in the DBMS. Only meta-information such as coordinates, average flux, pointing, etc. and relations to other objects like observer, instrument, etc. are stored and queriable (SQL, OQL, ...) inside the DBMS.

For many instrument analysis and scientific purposes, queries and correlations on a much smaller scale are needed, e.g. finding trends in detector noise for a specific instrument status, measuring line fluxes of one particular line of many sources in a vaste set of broadband spectra, comparing dark current with glitch activities.
Anticipating all queries that will be needed during the design phase of the database is impossible. A scheme where all relevant data is implemented as attributes in the database is therefore impossible. This need could be accommodated in a database system where not only meta-data is stored in a queriable form, but also the data proper. In the case of spectra, a table containing wavelengths and fluxes would then be defined in the database.

Present off-the-shelve RDBMS (Relational Database Management Systems) systems seem not to be robust enough to handle typical sizes of astronomical databases nowadays. Therefore for the ISO Post Mission Archive the concept of virtual data resolution with user defined functions has been designed and implemented as a novel database paradigm.

The traditional division is kept where only meta-information about observations and related files is stored in the relational database, the data files themselves are stored external to the database. But the user is allowed to define functions which extract attributes from any data file in the archive. This includes the possibility to apply Unix commands (sed, awk, grep, ...), TCL or PERL scripts to the ASCII-files, and for instance Interactive Data Language (IDL - a trademark of Research Systems Inc. RSI) functions to the FITS-files. Therefore it is possible to use features provided by the instrument interactive analysis systems (which are coded in IDL).

3. System design

The definition of an User Defined Function is entered in the system via a web interface. A UDF takes one file as input, optional additional input parameters and outputs the virtual attributes for that file.

A defined UDF can be executed on a a number of datafiles. The list of datafiles on which the UDF has to be run and the additional input parameters are prepared by the user in a temporary database table. The UDF is executed on every filename in the table and the the extracted virtual attributes are stored in a second table. The second table can be either inserted again in the database, where it can be used for further queries or it can be sent via e-mail to the user.

4. Implementation - how the ISO PMA user uses UDFs

If an expert user needs additional attributes to files (eg. observations) that are not provided by the standard RDBMS scheme he might use UDFs.

The interface to define and execute user defined functions is web-based. Figure 1 gives an overview of the interface structure. To take advantage of UDF processing the user needs an UDF user identification and password.

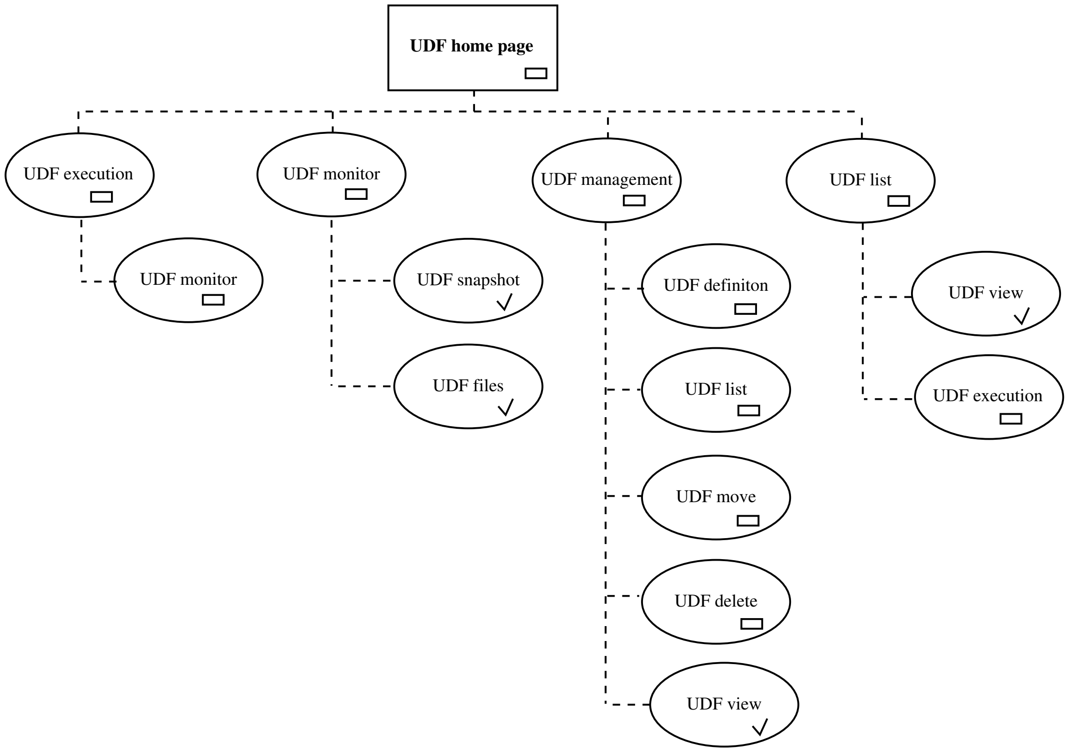

Figure 1. UDF WEB pages

The UDF list page gives an overview about existing UDFs in the public and private areas. But the user can also write his/her own new UDF and install it in his/her privat area on the server side. Further specifications of the UDF (like input parameters, ...) have to be given using the UDF definition page. This specifications can be displayed by the UDF view page.

If a UDF is stable and useful for the community it can be moved from the privat UDF area to the public area by the UDF manager using a WEB page interface : UDF move.

The UDF execution page is used to start the UDF process. The user has to specify the fixed input parameters and whether he wants to select required files via an SQL statement creating a temporary RDBMS table or use an existing database table.

The required files are ordered to guarantee most efficient file access and are copied to the work directory taking into account internal database limitations.

An UDF works on one file at a time. The result (table with filenames and virtual attributes) is available as a temporary RDBMS table, which can be used for further SQL processing or sent via e-mail.

The UDF process can be monitored using the UDF monitor page where the user might get additional informations like a snap shot of the result or which files have been copied from the archive at a certain point of time.

The UDF manager can monitor all the processes, start and stop or even kill them.

5. Conclusion

We presented concept, design and implementation of User Defined Functions on the ISO Post-Mission archive. We showed that UDFs allow users to query the archive using virtual attributes, of which the need cannot always be anticipated during the design phase of the archive.

Acknowledgments. The ISO Post Mission Archive (ISO PMA) is a development at the ISO Data Centre, ESA Astrophysics Division, Villafranca del Castillo, Spain.

Bart Vandenbussche acknowledges financial support from the Belgian Federal Services for Scientific, Technological and Cultural Affairs and from the Onderzoeksfonds K.U.Leuven grant OT/94/10

References

Kessler,M.F. et al., 1996, A&A 15 The Infrared Space Observatory (ISO) mission

Saxton, R.et al., in ASP Conf. Ser., Vol. 145, Astronomical Data Analysis Software and Systems VII, ed. Rudolf Albrecht, Richard N. Hook, & Howard A. Bushouse (San Francisco: ASP)The ISO Post Mission Archive