

Embedded Astrophysics Query Support Using Informix Datablades

A. Zhang, T. Handley

*Infrared Processing and Analysis Center, California Institute of
Technology, Pasadena, CA 91125*

Abstract. The 1.2 billion stars in the Two Micron All Sky Survey (2MASS) working dataset provide significant science opportunities and accompanying database challenges. Effective and efficient access to large datasets such as these is an important service of the Infrared Science Archive (IRSA). By embedding domain-specific query support into a query engine, IRSA provides a significant step toward more efficient queries. This paper describes IRSA's new generation query support in which the Informix server has domain-specific embedded query support, e.g., datablades modules with astronomical functionality. The first IRSA Astronomical datablade developed supports coordinate conversions.

This Astronomical datablade provides scientists, projects and the public with embedded coordinate conversions among common astronomical coordinate systems. Supported conversions include Equatorial, Ecliptic, Galactic, and Super Galactic, including the conversion between Julian and Besselian. This enables data retrieval with no intermediate or client-side processing steps – the user retrieves data from the database as usual. This capability is being deployed to enhance the current IRSA general query support services.

1. Introduction

With dramatic data volume growth in astronomical observations, a new millennium of information retrieval is fast approaching. Such a new era challenges the astronomical community to move to more advanced computer technologies in data mining, data management, data archiving and data analysis. A more creative and efficient way of data retrieval combined with required data analysis is needed. IRSA's astronomical datablade module is one of the solutions for addressing such technical challenges.

2. Datablade Technology

The fundamental datablade module is a software package. It can define any functionality required. Essentially an embedded module is used to extend the intrinsic functionality of Informix Server by implementing user-defined data types and their supporting routines. The science community or new missions are able

to define their own data objects and manipulate those database objects using their own analytic methods in a natural, flexible way.

The basic datablade includes a set of Structural Query Language (SQL) statements and a set of supporting code written in an external language such as C. The datablade accepts user-defined database objects that extend the SQL syntax and its commands.

In addition to the above, generally speaking, datablades provide better performance and simpler client-side applications. Datablade modules handle code for manipulation and storing data, so the application does not have to include low-level resources. Furthermore, datablade module routines and data types can be accessed using SQL as other intrinsic functions and data type. Finally, datablade modules are easy to upgrade.

IRSA has extended the Informix database server with an astronomical coordinate conversion capability. Astronomical coordinates are converted and processed within the database server instead of within a client-side application.

3. Coordinate Conversion Datablade

Scientists, researchers and engineers who work in astronomical projects deal with coordinate conversions on a daily basis. In the conventional way, the coordinate related data are just loaded into the database. In most cases, in order to consider the efficiency of data storage, archives ingest coordinates in a common coordinate system, such equatorial J2000. In order to support such processing in a pipeline or data analysis, coordinate conversions are required.

If the coordinates that users need are not stored in the database table, a client-side program is required to accomplish the coordinate conversion. This step cannot be eliminated, since the database serves only as a storage and search machine.

In some cases, database tables are designed to give a certain degree of flexibility by containing additional coordinates. This approach only partially solves the problem and may actually raise other archive issues.

Coordinate conversion is a complicated process. The resulting coordinate pair (RA, Declination) depends on observation time, epochs of the FROM and TO coordinate systems, and various correction conditions. Ingesting more than one coordinate system may meet some users' immediate needs. However, this does not satisfy the general users' demands, since there is no way to ingest all of the coordinate values required. One extra coordinate in a database table, like the 2MASS working database table, would require an additional 20 GB of disk space. Just storing extra columns to meet scientific requirement is not the solution for the problem by its nature. And, overgrown tables will result in serious efficiency and storage problems. However, a database table that does not store extra columns can result in more steps to accomplish a job.

4. Datablade Design and Functionality

The coordinate conversion datablade (CNV) transforms coordinates among common astronomical coordinate systems including the conversions between Julian

and Besselian. It also supports the conversion with specified proper motions, unknown proper motion or radio source, or without proper motion. For position angle calculation, the datablade handles both epochs of position angle, which of course may differ. In addition, CNV provides a set of conversion corrections, such as FK4-FK5 systematic correction, elliptic aberration E-term correction, photometric magnitude correction, or any combination.

Whenever alternative coordinates are required, only an SQL statement is needed. Calling a set of SQL functions in either dbaccess, an Informix provided query tool, or in application software does the conversion. Since these functions are part of the server, they are transparent to users. Users only need to know the SQL.

5. SQL Function Calls

Based on the nature of the conversion, the functions in IRSA's astronomical datablade are divided into three subsets with convenient default values: (1) general conversion functions; (2) conversion functions with at least one galactic or super galactic coordinate system; and (3) conversion between galactic and super galactic coordinate systems.

Functions accept input values of RA and declination in decimal degrees or sexagesimal degrees (depending on the type of transformation desired), and output values in either decimal or sexagesimal degrees, according to the functions invoked.

- (1) General Conversion Functions – These functions transform any astronomical coordinate into another coordinate with a defined transformation.
- (2) Conversion Functions with at Least One Galactic or Super Galactic Coordinate System – This set of functions is used to convert a non-Galactic or a non-Super Galactic coordinate system to a Galactic or Super Galactic coordinate system, or vice versa (in this category, one less argument is required).
- (3) Conversion Functions between Galactic and Super Galactic Coordinate Systems – Any function in this set can be performed by the above two sets of conversion functions.

6. Enhanced Archive Architecture Using the CNV Datablade

Currently IRSA provides a rich service for astronomical catalog query and image archive retrieval. When a user wants to perform a positional query or cone search, the speed of searching is dependent on whether the Informix optimizer chooses indexing path. To enhance regional searching, spatial index columns are added to the tables. The table index is built on those tables where the input columns contain ra and dec on coordinate Equatorial Julian 2000. This design adds a constraint, i.e., positional values must be in J2000 Equatorial coordinates.

The architecture of the Data Ingest and Upload service within IRSA for cross-comparison is enhanced by deploying the CNV datablade. The new data ingestion system can convert any astronomical coordinate, whether it is in decimal degree or sexagesimal degree, to Equatorial J2000 system if spatial indexing is required.

7. Astrophysics Query Using Embedded CNV

The Informix server with CNV datablade promotes flexible data processing at IRSA. By launching CNV datablade, users are able to retrieve data in any coordinate system regardless of what is stored in the database. Users are allowed to select proper observation time, input epoch and output epoch. Users are even given the flexibility of selecting different correction terms or conditions, e.g., with or without proper motion.

Uploading tables for cross-identification comparisons is one of the important features that IRSA provides. J2000 Equatorial coordinates in decimal degree was previously the only coordinate supported by this application, but with the advent of the CNV datablade, the upload utility is more flexible. Users can now load any coordinates in their list, and IRSA will convert them internally and return the objects of interest in the requested coordinate system.

8. Conclusion and Future Work

The embedded astronomical query which combines a data searching engine with coordinate conversion capability is an efficient tool for the astronomical community. It reduces intermediate steps and makes scientists and engineers' work simpler. Future work includes further optimization of the datablade based on usage experience and full deployment within IRSA and the Space Infrared Telescope Facility.