Cornell University
Center for Advanced Computing

CAC and the Arecibo Observatory

CAC
we enable your success

Searching for Pulsars in Very-Large Databases

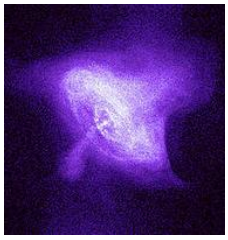
How do you identify pulsars searching through hundreds of terabytes of telescope data?

Finding the Answer

The Arecibo Observatory is collaborating with CAC to collect Arecibo survey data and to provide online tools that will allow scientists to explore pulsar data efficiently and effectively.

Searching for Pulsars

"The Arecibo pulsar surveys will be the deepest ever undertaken and are expected to yield not only about 1,000 new pulsars, but also exotic objects, including millisecond pulsars spinning near the break-up of a neutron star; neutron stars in compact binaries with orbital periods of a few hours or less; and, companion stars that are other neutron stars or black holes," explained James Cordes, Professor of Astronomy, Cornell University.



Improved Research

Research Metrics

- Centralized data storage: Leverage SQL Server database expertise and central storage capabilities at CAC for rapidly growing data sets
- Fast, efficient data exploration: Develop cross-platform tools for data analysis
- Availability: Refine data and provide data access to global scientific community 24x7 via Web services and the National Science Foundation's TeraGrid

Research Challenge

The volume of information being gathered in astronomy is expected to double every 1.5 years. This huge growth in data volume is accompanied by a great increase in data complexity. Cornell astronomers, and astronomers from around the world, use the Arecibo telescope to conduct data intensive surveys. These surveys produce thousands of terabytes of data.

The search for pulsars is complicated by the fact that their signals are weak and dispersed: pulses arrive differentially in frequency and sometimes intermittently. Some pulsars are in orbit around another star, causing a variable Doppler shift that must be accounted for, and sometimes they are almost completely obscured by interference from terrestrial radar sources.

Another major complication is the sheer volume of data. When the radio telescope is pointed at a particular direction, the measured signals are sampled more than 10 thousand times per second in several hundred different frequency channels. A three-hour observing session generates about 1 TB of data. However, most of it is just background sky and receiver noise combined with terrestrially generated radio interference. Good techniques for analyzing and mining pulsar data are paramount.

Solution

CAC developed several relational databases and data tools for Arecibo users. The data tracking database contains information on each data set as it moves through the processing pipeline. Some of the steps in the pipeline include tracking disk shipments from the Arecibo Observatory in Puerto Rico to the Space Sciences Building at Cornell, managing the backup server as the data are loaded into the archive, running the signal processing code on newly arrived datasets, and loading the final data products into a large relational database. This information is also accessible online for simple access and real time tracking. The raw data are also sent on to partner institutions involved in the Pulsar-ALFA Consortium, made up of institutions around the world, but concentrated in North America.

After the data have been stored in the destination database, scientists can access the data through a variety of methods. CAC created a Web site with different queries and plots available to aid in the search for pulsars. Because of the large volume of data, it is often useful to graph the data and allow the user to visually search for certain trends. To facilitate this, online tools for plotting data are provided. Plots include histograms, scatter plots, and line plots that can be generated through a simple intuitive interface or customized plots for more specific types of analysis.

Many scientists prefer to run more advanced analysis techniques locally. The Arecibo Pulsar data are made available through Web services to support such methods. Scientists on different platforms (such as Windows or Linux) can connect to the database through the Web services and download the specific set of information that interests them.

The Web service interface supports VOTable, a standard XML data format developed by the Virtual Observatory. This allows the data to be easily shared and distributed across the astronomy community.

The Client

The Arecibo Observatory

- World's largest and most sensitive single-dish radio telescope located in Puerto Rico
- Operated by the National Astronomy and Ionosphere Center (NAIC) at Cornell and NSF
- New multibeam feed system is providing a 7-fold surveying rate speedup
- The Pulsar ALFA Survey is searching for neutron stars (pulsars) and other transient radio sources
- The Arecibo Legacy Fast ALFA (ALFALFA) Survey is expected to detect 20,000 extragalactic objects, mostly galaxies, via radiation from atomic hydrogen.

The Collaborative Relationship

CAC is helping Arecibo provide refined data products and Web interfaces that will provide scientific value for years to come, providing synergistic opportunities with other large-scale surveys and with new telescopes such as the Gamma-ray Large Area Space Telescope. The Arecibo data is being made available through CAC's Science Gateway on the National Science Foundation's TeraGrid.

"We anticipate that our results will be of considerable value to astronomers and astrophysicists world-wide. The service-oriented interface to the data will allow users from all over the world to interactively query the multidimensional search space and to allow interactive and efficient exploration of the dataset."

James M. Cordes
Professor of Astronomy, Department of Astronomy
Cornell University