Astronomical Data Analysis Software and Systems X
ASP Conference Series, Vol. 238, 2001
F. R. Harnden Jr., F. A. Primini, and H. E. Payne, eds.

The Submillimeter Array Data-Handling System

Jun-Hui Zhao, Takahiro Tsutsumi

Harvard-Smithsonian Center for Astrophysics, 60 Garden St., Cambridge MA 02138

Abstract. We report on the basic design and current status of the data-handling system for the Submillimeter Array (SMA). Components of this system currently under development include the data storage format, archive, and off-line data reduction software.

1. Introduction

The Submillimeter Array (SMA) is under construction at Mauna Kea (Moran 1998). The SMA's first fringes from observations of celestial sources were obtained with two antennae on September 29, 1999. A year later, the first phase closures were successfully achieved on Uranus. A synthesis image of this planet at 230 GHz was made from the observations using the SMA's first three elements. As the SMA correlator comes on-line, the maximum data production rate will approach 2.75 MB per sampling. For a typical integration time of 10 seconds, the daily data production rate of the SMA would be 20 GB/day. In this paper, we present the design of the data-handling system and report the status of the software development in support of data reduction and analysis for SMA users.

2. Software Design and Development

Figure 1 shows an overview of the architecture of the SMA on-line data-handling software. Communication between the data-handling computer *Smadata* (a Sun Ultra 60 running Solaris) and the real-time system (the SMA correlator *Crates* and a control computer *Hal9000*) is accomplished with remote procedure calls (RPC) via a local network (100 Meg/sec Ethernet). *Smadata* is a central host of the data-handling server (smadata-svc), performing the post-correlator data processes such as data formatting, on-line correction, and flagging. In addition, this data computer also hosts the servers for data archiving, database management, data replication and HTTP.

RPC server and Data Format The RPC server *smadata_svc*, developed in C, provides several data services to process the data received from real-time computers *Crates* and *Hal9000.* The cross-correlation data from the SMA correlator and ancillary data are organized and stored in a number of FITS tables following the FITS-IDI standard (Diamond et al. 1997;Flatters 1998). During

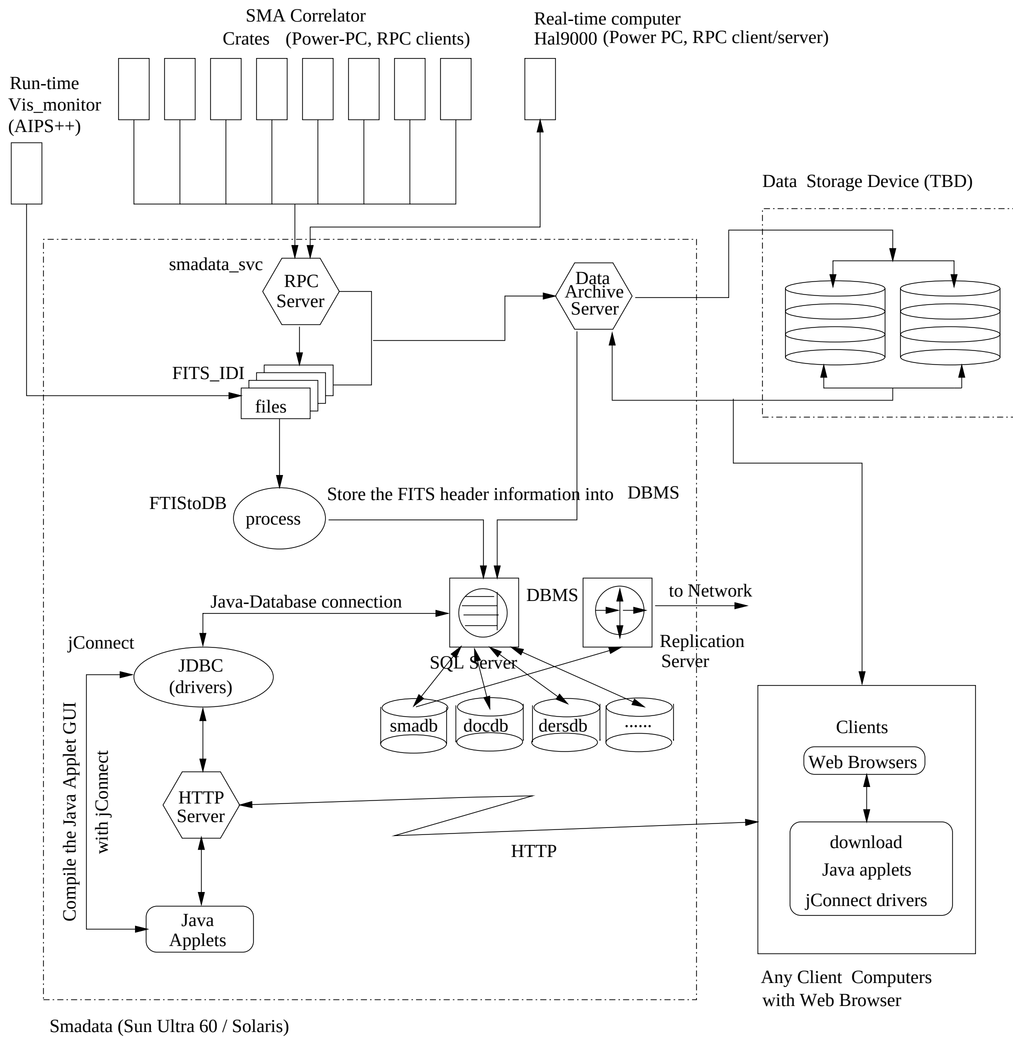

Figure 1. The SMA Data-Handling Software Architecture. The host computer is Smadata, a Sun Ultra 60 running Solaris. The RDBMS is Sybase. The JDBC utilizes jConnect from Sybase. This configuration is for the primary site currently located on Mauna Kea. Eventually, this system will be moved to the SMA headquarters in Hilo. There already exists a dedicated network link (45 MB/s) between Mauna Kea and Hilo.

an observing run, a visibility data monitor (Vis_monitor, under development in AIPS++) will provide a handy, run-time imaging facility for data quality control. At the end of each observing run, a single portable FITS-IDI is produced. The SMA FITS-IDI can be directly read into the AIPS environment and is ready for off-line data analysis.

On-line Archive and Storage A Sybase SQL Server relational database management (RDBM) system is being used at the SMA for various types of data management. Its relatively low cost (compared with other commercial packages such as Oracle) is suitable to the size of a project like the SMA. With standard ANSI SQL (Structure Query Language), the software functions supported by Sybase also meet our requirements for archiving documentation and data management.

With this commercial software, we are also developing an on-line archive system to handle SMA interferometer data. The FITS-IDI files will be stored in mass storage. The header information of the FITS tables in each FITS-IDI file along with the file location is archived in the SMA astronomical database (SMADB), which is managed by the Sybase server. The preliminary design of this system is illustrated in Figure 1. At the termination of each observing run, the RPC server *smadata_svc* triggers a process, FITStoDB, which extracts all the header data from the FITS-IDI files and converts them to the database in Sybase.

The SMA Astronomical Database (SMADB) The database model is based on the data structure of the FITS-IDI file. Ten relational Sybase tables are needed to model the SMADB. Table 1 (RUN_LOG) contains the general information for each observing run. The information about the correlator that generates the visibility data is included in Table 2 (CORR). The mandatory keywords for each FITS-IDI file are stored in Table 3 (FITS_KEY). The general information on FITS tables in each FITS-IDI file is stored in Table 4 (TAB_NM). The parameters for frequency setup, source coordinates and velocities are stored in Tables 5 (FREQ), Table 6 (SOUR), and Table 7 (VELO). The information regarding the array geometry is saved in Table 8 (ARR_GEO). The information on the visibility data can be found in Table 9 (VIS), and byte-size and location of each FITS-IDI file are stored in Table 10 (DFILE).

Data Replication A primary data archive system is located at the Mauna Kea site Eventually, it will be shifted to the SMA Hilo base facility. Most SMA users are located at two remote institutes, CfA in Cambridge (Massachusetts) and ASIAA at Nankang (Taipei). Due to the large volume of SMA visibility data, users and applications at these sites would suffer unacceptable delays in receiving complete data sets and would also generate a large amount of network traffic if they could only access data from the primary site. To avoid this problem, the current design includes replication of the data on the local systems.

User Interface A JDBC driver, Sybase's jConnect, has been installed in the Server host computer Smadata. The basic configuration for the SMA On-Line Archive System is illustrated in Figure 1. JDBC provides standard Java API codes that allow us to develop a specific Java Applet GUI (Graphical User Interface) to communicate with SMADB via the SQL server. The data computer also hosts an HTTP server. This Server provides a port for outside clients to download Java Applets and therefore to establish a connection with the database server. As soon as the client/server connection is established, the data transaction can proceed via the network.

3. Hardware for Data Storage

As the SMA becomes fully operational (with all 8 antennae and a full set of MIT/SAO correlators), data storage will become a major issue for the on-line data archive system described in the previous sections. We will inevitably need a high capacity mass storage system.

We continue to investigate hardware devices for data storage, including a DLT library or DVD-R juke-box. However, we have a temporary solution for keeping the visibility data on-line during the construction and testing phase. The current storage hardware system is implemented with several multipack disks attached to the data server *Smadata* (Ultra 60) while either the DLT library or DVD-R juke box is being considered.

4. Off-line Data Reduction Software

Three primary interferometric data reduction environments (AIPS++, AIPS, and Miriad) are chosen by the SMA staffs for off-line data reduction. Utility codes for calibrations are under development in support of the SMA specifications.

Acknowledgments. This paper is based on SMA Technical Memo 138. We thank the SMA staff for their many helpful comments and discussions in the course of the software development.

References

Diamond, P. J., Benson, J., Cotton, W. D., Wells, D. C., Romney, J. D. and Hunt, G. 1997, VLBA correlator Memo No. 108 (NRAO)

Flatters, Chris, 1998, AIPS Memo No. 102 (NRAO)

Moran, J. M. 1998, in Advanced Technology MMW, Radio, and Terahertz Telescopes, Ed. Thomas G. Phillips, Proc. SPIE Vol. 3357, 208

Zhao, J.-H., Mailhot, P., and Tsutsumi, T., 2000, SMA Technical Memo No. 138 (SAO)