

The Distributed Analysis System Hierarchy (*DASH*) for the SUBARU Telescope

Y. Mizumoto, Y. Chikada, G. Kosugi¹, E. Nishihara², T. Takata¹ and M. Yoshida²

National Astronomical Observatory, Osawa, Mitaka, Tokyo 181, Japan

Y. Ishihara and H. Yanaka

Fujitsu Limited, Nakase, Mihama, Chiba 261, Japan

Y. Morita and H. Nakamoto

SEC Co. Ltd. 22-14 Sakuragaoka, Shibuya, Tokyo 150, Japan

Abstract. We are developing a data reduction and analysis system *DASH* (Distributed Analysis Software Hierarchy) for efficient data processing for the SUBARU telescope. We adopted CORBA as a distributed object environment and Java for the user interface in the prototype of *DASH*. Moreover, we introduced a data reduction procedure cube (*PROCube*) as a kind of visual procedure script.

1. Introduction

The purpose of *DASH* is efficient data processing for the SUBARU telescope which will produce up to 30TB of data year. This data production rate requires high computational power and huge mass storage capacity. *DASH* is, therefore, designed as an observatory system for astronomical data reduction and analysis, which cooperates with the SUBARU observation software system and the data archival system. *DASH* is, of necessity, designed for a distributed heterogeneous computer environment; considering role sharing and joint work. We adopt CORBA (Common Object Request Broker Architecture) for a distributed environment platform and Java for the user interface objects. As the first step of the development, which is a three year project, we made a prototype of *DASH* for trial of the new computer technology such as CORBA and Java.

¹Subaru Telescope, National Astronomical Observatory of Japan, Hilo, HI 96720, USA

²Okayama Astrophysical Observatory, Kamogata-cho, Okayama 719-02, JAPAN

2. Design Concept and Requirement of *DASH*

The requirements of the data reduction system for SUBARU telescope are the following;

1. *DASH* has an enriched function of data management. The data management cooperates with SUBARU data archives and database (Ichikawa 1995, Yoshida 1997). It can collect all kind of data necessary for data reduction from the SUBARU data archives using Observation Dataset, which is produced by the SUBARU observation control systems (Kosugi 1996). The Observation Dataset is a set of information about acquired object frames, related calibration frames such as a bias frame of CCD, and others. The data management also keeps track of temporary files that are created during data reduction process.
2. *DASH* can be used on a distributed heterogeneous computer environment. In particular, the console or operation section of *DASH* must work on many kinds of computers (workstations and PCs).
3. It is easy to assemble a suitable processing pipeline with the steps chosen through a series of data processing trials.
4. *DASH* aims at open architecture. We intend to use applications from widely used astronomical data reduction packages such as IRAF and Eclipse within *DASH*.

According to the above requirements, we chose the following line;

1. We adopt CORBA as a distributed object environment (Takata 1996).
2. We adopt Java for the operation section or user interface of *DASH*. The user interface is defined on the character basis. A Graphical User Interface (GUI) is built on it.
3. We use object oriented methodology. We acquire new experience in object oriented analysis and design of software development.

3. *DASH* System Architecture Model

We analyzed the data reduction system of astronomical CCD images and made a restaurant model shown in Figure 1. The model has 3 tier structure. The first tier is “Chef” which corresponds to the user interface or console section. The second tier is “Kitchen” and “Books”. The third tier is composed of “Ware House” which is the data server or data archives, “Cooks” which is analysis task or CPU server, and “Books” which is the database. These components are on the network computers. Each tier is linked through the Object Request Broker.

4. PROCube

PROCube is an object that unifies the whole process of the data reduction of an astronomical object frame (CCD image). It handles processing flow together

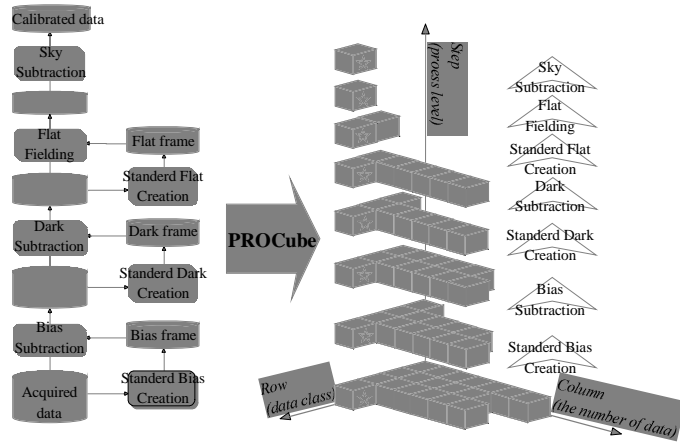
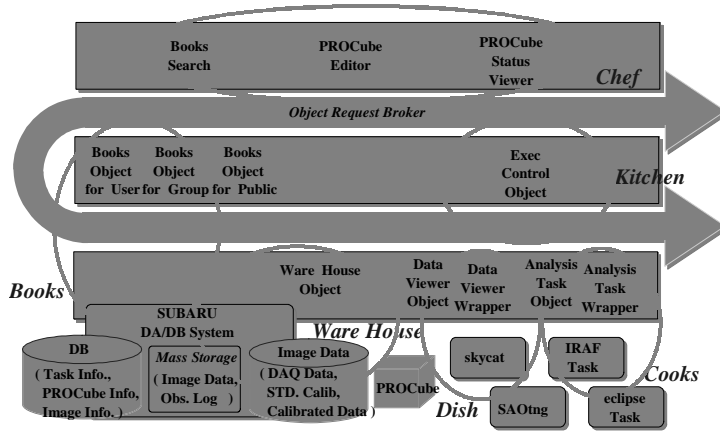


Figure 2. PROCube

with image data. Figure 2 shows a *PROCube* like a 3 dimensional flow chart. The X-axis is the kind of image frame, the Y-axis is the number of image frames, and the Z-axis is the flow of reduction. The base bricks of the cube stand for raw data files of an object frame and related calibration frames to be used in the reduction process. The apex of the cube corresponds to the resultant file. *PROCube* is a kind of visual procedure script which can be executed by “Kitchen”. A log of its execution is also recorded in the *PROCube*. Abstract *PROCube* that is extracted from the executed *PROCube* works as a procedure script with a variable of an object frame. The abstract *PROCube* can be used for the pipeline.

5. Prototype of DASH

We made a prototype for a feasibility study of *DASH* in a distributed heterogeneous computer system using CORBA (ObjectDirector2.0.1), Java (JDK1.0.2), and Web. Some tasks of IRAF 2.11b and Eclipse 1.3.4, and Skycat 1.1 are used as an analysis engine and an image viewer using a wrapping technique. *PROCube* is also implemented in the prototype.

The prototype is tested and evaluated on a small scale system. We used two workstations (Solaris 2.5.1 J) connected through FibreChannel switch (266Mbps) and two personal computers (Windows 95 J). The four computers composed a distributed computer system using Ethernet (10baseT). This computer system is isolated from the other networks.

6. Results from the Prototype

A data reduction system implemented the *PROCube* is realized on the network computers using CORBA as a platform for the distributed software environment. The prototype works well on a system composed of 2 server workstations and 2 client PCs. From a client PC, 2 server workstations look like one network computer.

The wrapping of an existing software package as a CORBA object is not easy and it takes a long time to study details of the package. Eclipse and Skycat were easier to wrap than IRAF. A series of data reduction processes can be done with a *PROCube* that uses IRAF tasks and Eclipse tasks together. The *PROCube* is able to work as a data reduction pipeline. Parallel processing of the tasks is realized in one *PROCube*. It is confirmed that parallel processing is effective for I/O bound tasks.

Problems or disadvantages of the prototype also become clear. CORBA and Java are still under development. So, the products CORBA and Java may have some unsatisfactory features as well as bugs. Tasks for image data reduction are I/O bound in most cases. It is important in data transfer to harmonize performance of the network with that of disk I/O. NFS through 10Mbps Ethernet is not satisfactory for data transfer of large image files.

Data reduction with a *PROCube* may create a lot of temporary files, which eat a lot of disk space. Yet it is hard to manage lifetimes of temporary files in a *PROCube*.

References

- Ichikawa, S., et al. 1995, in ASP Conf. Ser., Vol. 77, Astronomical Data Analysis Software and Systems IV, ed. R. A. Shaw, H. E. Payne & J. J. E. Hayes (San Francisco: ASP), 173
- Yoshida, M., et al. 1997, in ASP Conf. Ser., Vol. 125, Astronomical Data Analysis Software and Systems VI, ed. Gareth Hunt & H. E. Payne (San Francisco: ASP), 302

- Kosugi, G., et al. 1996, in ASP Conf. Ser., Vol. 101, *Astronomical Data Analysis Software and Systems V*, ed. George H. Jacoby & Jeannette Barnes (San Francisco: ASP), 404
- Takata, T., et al. 1996, in ASP Conf. Ser., Vol. 101, *Astronomical Data Analysis Software and Systems V*, ed. George H. Jacoby & Jeannette Barnes (San Francisco: ASP), 251