Available online at www.sciencedirect.com
ScienceDirect
Fusion Engineering and Design 83 (2008) 438–441
Fusion Engineering and Design
www.elsevier.com/locate/fusengdes

Data processing for 'SUBARU' telescope using GRID

Y. Shirasaki [a,*], M. Tanaka [a], S. Kawanomoto [a], S. Honda [a], M. Ohishi [a], Y. Mizumoto [a], N. Yasuda [b], Y. Masunaga [c], Y. Ishihara [d], J. Tsutsumi [d], H. Nakamoto [e], Y. Kobayashi [e], M. Sakamoto [e]

[a] *National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan*
[b] *University of Tokyo, 5-1-5 Kashiwa-no-Ha, Kashiwa, Chiba 277-8582, Japan*
[c] *Aoyama Gakuin University, 5-10-1 Fuchinobe, Sagamihara, Kanagawa 229-8558, Japan*
[d] *Fujitsu Ltd., 4-1-1 Kamikodanaka, Nakahara-ku, Kawasaki 211-8588, Japan*
[e] *Systems Engineering Consultants Co. Ltd., 22-4 Sakuraoka-cho, Shibuya-ku, Tokyo 150-0031, Japan*

Available online 24 October 2007

Abstract

The amount of astronomical data is rapidly growing as the progress of telescope technology and the detection technique of recent years. As a result, the way of traditional analysis is becoming insufficient for utilizing the large amount of data efficiently. The SuprimeCam, which is one of the instruments equipped with the Subaru telescope, has been generating 7 TB of public data since its start of operation. It is almost impossible to transfer all the data to the local machine to analyze them. It is, therefore, desirable to have an environment where the data is analyzed where it is stored. In addition, it is not easy to use the Subaru data for an researcher who is not familiar with the Subaru data reduction. To overcome these difficulties, we developed the server side data analysis system and applied GRID technology to construct the system that carry out multiple jobs on multiple servers. Integrating this system to the Japanese Virtual Observatory, a user can easily utilize the GRID system through the web browser interface.

Keywords: GRID; Astronomy; Virtual observatory; Web

1. Introduction

Thanks to the recent technological innovation of electronic devices, a large scale of high-quality data is continuously generated. The way of traditional analysis is becoming insufficient for using the large amount of data effectively and efficiently, and for getting the maximum scientific results. Although many astronomers recognize the importance of research that uses the multi-wavelength data, such research actually needs considerable effort. One reason is that, for each data set, one needs to learn how to reduce and analyze the data, and even needs to know where the analysis tools are available and how to install them (see top panel of Fig. 1). To overcome such situation and maximize the scientific return from a big project like Subaru and ALMA, it is important to construct an environment where user can access to the science-ready data with very few effort (see bottom panel of Fig. 1). National Astronomical Observatory of Japan (NAOJ) started its VO project (Japanese Virtual Observatory, JVO[1]) in 2002. The objectives of the JVO project are to provide a seamless access to the distributed data service of the world, and to provide user-friendly analysis environment. To realize the interoperability among the astronomical databases of the world, it is crucial to define a world standard of the access protocol for the databases. The International Virtual Observatory Alliance (IVOA[2]) was formed in 2002 for that purpose. The IVOA now comprises 16 VO projects from Armenia, Australia, Canada, China, Europe, France, Germany, Hungary, India, Italy, Japan, Korea, Russia, Spain, the United Kingdom, and the United States. This paper describes our recent progress on the data analysis environment of the JVO.

* Corresponding author. Tel.: +81 422 34 3840.
E-mail address: yuji.shirasaki@nao.ac.jp (Y. Shirasaki).

[1] http://jvo.nao.ac.jp/.
[2] http://www.ivoa.net/.

0920-3796/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.fusengdes.2007.09.004

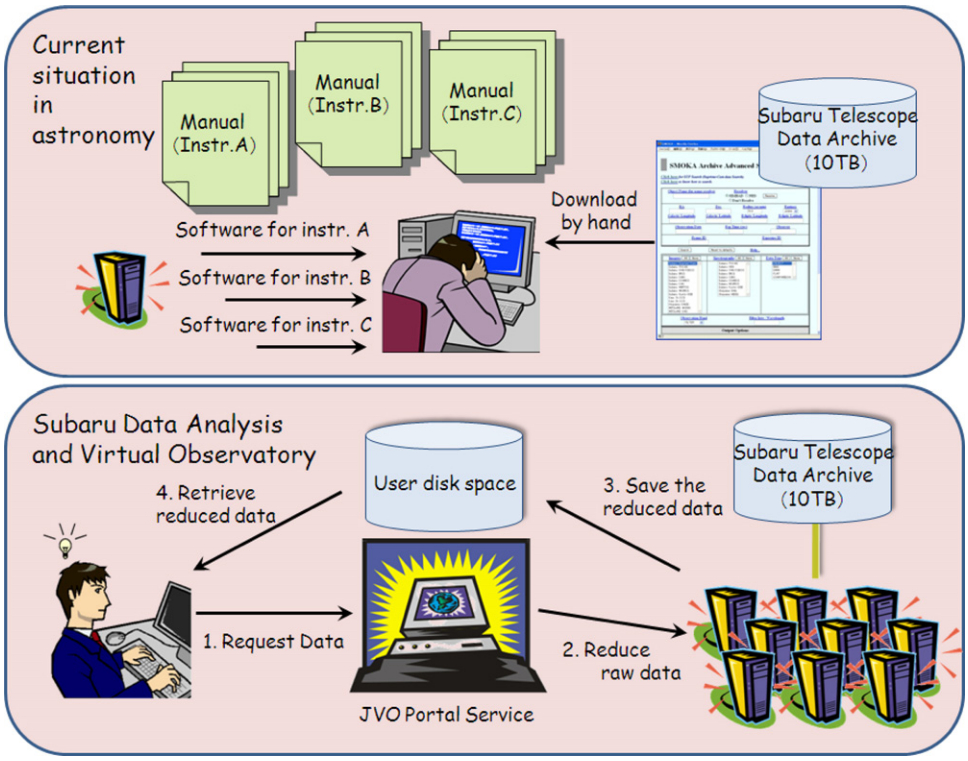

Fig. 1. Current situation of astronomical study (top) vs. astronomy of the future (bottom).

2. Subaru data archive

Subaru[3] is an optical-infrared 8.2 m telescope operated by National Astronomical Observatory of Japan (NAOJ) at Mt. Mauna Kea, Hawaii. Subaru has seven open use instruments: CIAO, COMICS, FOCUS, IRCS, SuprimeCam, HDS and MOIRCS. Using these instruments, observation can be made for wavelengths from optical (300 nm) to infrared (20 μm) with spectrum resolution up to 10^5 (HDS).

As of October 2006, 8 TB of data is archived in the public area. More than 70% of the data are from the SuprimeCam, which is a mosaic of ten 2048 × 4096 CCDs and covers a $34' \times 27'$ field of view with a pixel scale of $0.20''$ (Fig. 2). More than 90% of all the data requests are for the SuprimeCam, so our current priority issue is how to improve the usability of the SuprimeCam data. There is a plan to upgrade the SuprimeCam to the Hyper SuprimeCam (HSC) in 2011. The HSC will have 10 times larger detection area than the SuprimeCam, and is expected to generate data in 10 times higher rate (Fig. 3). The data are registered in the Subaru Telescope Archive System (STARTS) in Hawaii as soon as the data are acquired by the instruments, so an observer retrieves his data from STARS during and/or after the observation. The data of STARS is mirrored to the Mitaka Advanced STARS (MASTARS) in Japan, so the observer can retrieve the data also from the MASTARS when he returns to Japan. STARTS and MASTARS are not public data archive. To use the system, you need to get an account on the Subaru computing system for STARS or an account on the Mitaka computing system for MASTARS. The data that has passed 18 months of a proprietary period becomes publicly available through the SMOKA[4] and the JVO system. The SMOKA system provides various query modes for the Subaru archive. The JVO system provides IVOA standard access interface to the Subaru archive.

3. Data reduction of SuprimeCam

The amount of data, especially of SuprimeCam, is very large, so it is important to provide a way to analyze the data without moving the data to a users' machine. One of the ways to do so is to login to the Subaru or Mitaka computing system and analyze the data on the machine. It is, however, not practical to use the visualization tool from a remote machine, especially when accessing through a slow network, and is also not easy for a novice user to analyze the Subaru data with command line user interface. Another solution is to provide a web service, through which one can access to the data analysis software and visualize the data in a compact graphical format. Recently a lot of open source framework are available for making such a service, and it has been realized that interactivity of the web-based service can be improved by using Ajax technique as demonstrated by the google map service. To reduce the data of Subaru SuprimeCam, a lot of computing resources are required. To make one final mosaic image from raw data, flat-field frames must be prepared for each CCD (Fig. 4). The flat-field frame is derived from hundreds of observation frames by calculating their median values

[3] http://subarutelescope.org/.

[4] http://smoka.nao.ac.jp/.