Astronomical Data Analysis Software and Systems IX
ASP Conference Series, Vol. 216, 2000
N. Manset, C. Veillet, and D. Crabtree, eds.

"On-the-Fly" Calibration System for the ISO Data Archive

P. Osuna[1], C. Arviset, R. Saxton[2], A. Pollock[3], E. Verdugo[1]

ISO Data Centre, ESA, Villafranca del Castillo, Apartado 50727, 28080 Madrid, Spain

Abstract. The Infrared Space Observatory, ISO, has performed around 137000 observations during its almost 28 months of flight. At the end of the mission, all available data were processed and ingested into the ISO Data Archive (IDA)[1] specifically designed at the ISO Data Centre in Villafranca del Castillo near Madrid, Spain. However, data processing and calibration software improve continuously as the behavior of the instruments is better understood, making necessary the existence of a proper "On-the-Fly" Calibration System, by which the astronomical community would be able to request the state-of-the-art ISO products via the IDA User Interface.

1. Introduction

The core of our processing system consists of an ALPHA Cluster made up of six Alpha machines running under OpenVMS. Each of the ALPHA machines has got two different data environments accounting for a final number of twelve processing areas. Each of them corresponds physically to a different disk unit, making I/O operations very efficient while sharing -every two- the CPU time of the same machine. The user sends a request through the ISO Data Archive User Interface to get the latest available data products. The system starts the On-the-Fly Reprocessing by sending the request to the ALPHA cluster where a scheduling system looks for a free area and eventually starts the processing. After the reduction process has finished, it sends the processed data back to the IDA machine together with a log specifying details of the executed jobs. The IDA sends a mail to the user who made the request with information about the availability of the data via FTP or CD-Rom.

[1]INSA, Spain

[2]VEGA Group PLC, UK

[3]CSCL, UK

[1]See Arviset et al. 2000. For the origins of the project, see Saxton et al. 1998.

2. The Reasons to Build the System

An initial "Bulk Reprocessing" of the 137000 observations executed by the ISO satellite was needed to populate the initial ISO Data archive. Such a tool was developed and the processing started on a per-revolution basis, taking four months approximately to complete. Besides the population of the initial archive, this Bulk Reprocessing of the ISO data left the so-called "auxiliary" data on disk, which would then be used for the future On-the-Fly Calibration System. The fact that the first step in the reduction of the data (the extraction and separation of the science data from the housekeeping data directly from the telemetry) was very stable allowed this approach.

The improvement in the knowledge of the instruments was leading to better calibrations once every three to four months, and the need to reprocess the data every time a better one was found. However, it was impractical to reprocess all the observations for the whole mission each time a new calibration is released, due to the huge amount of data to be processed and the fact that calibration improvements are instrument-dependent.

The solution to previous problem was to develop a system that would allow the user to request either data which were already archived (i.e., the data which populated the initial archive with the initial calibrations) or the "latest available" data, i.e., those data which would be processed with the latest available calibration. This solution was implemented through the so-called "OFRP System" (On-the-Fly Reprocessing System). The advantages of this approach are manifold:

- Users can ask for the processing of their products with the last ever calibration;
- Only the products which are known to have improved since the last calibration will be processed. With this approach, we avoid any unnecessary processing of any products which might have not improved with respect to the already existing (archived) ones;
- We make full use of parallel processing power by allowing different requests to be processed at the same time. The six different ALPHA machines with a total of twelve different working areas allow for a processing rate of more than 1000 observations a day.

3. Hardware Environment

As already mentioned, the hardware environment comprises six clustered alpha machines running under OpenVMS. Each of them has got assigned two different data areas, named after colors, plus one area (named after Neptune's daughters) which is used for parallel multipurpose test reprocessing. A total of 45 disks are mounted on the system, each one with a capacity of 9 GB, holding the complete telemetry of the ISO mission plus the auxiliary data needed to start the On-the-

Fly Reprocessing. This cluster is connected with the ISO Data Archive so that it can send the information back whenever the processing is finished.

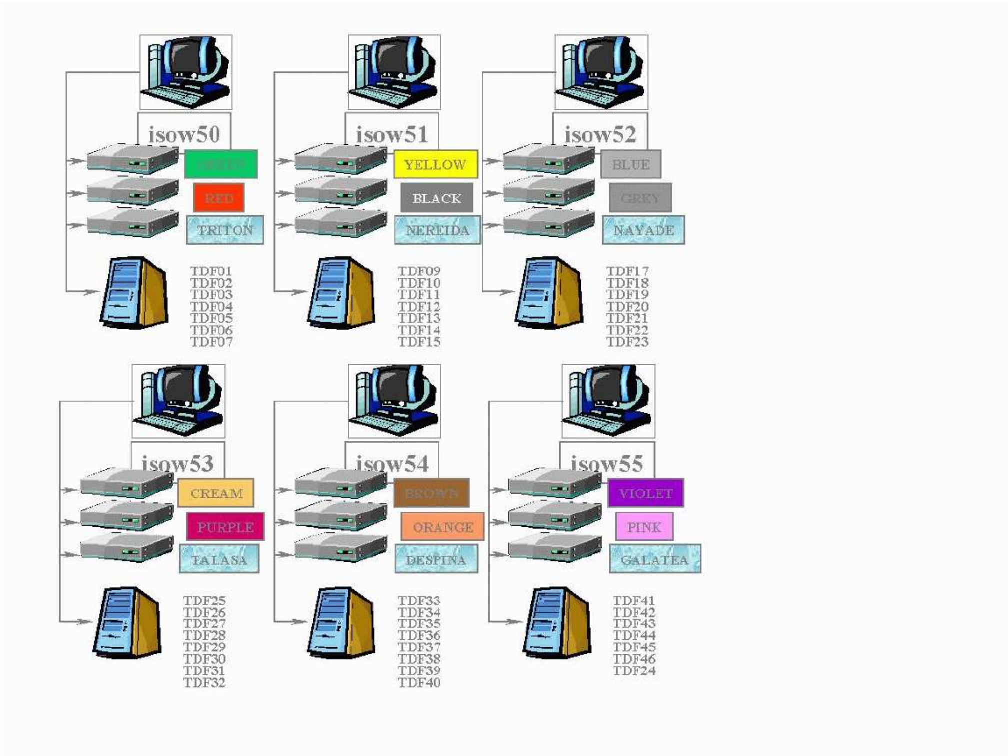

Figure 1. OFRP Hardware Environment.

4. Monitoring the System

The OFRP system was designed to execute with a minimum of human interaction. During the whole of the ISO mission, there were four Off Line Processing Operators caring for the pipeline processing of the data which happened to take place in only two areas and a minimum of disks. The number of operators was reduced to only one operator to care for the whole of the processing of up to one thousand observations a day. This made it necessary to create a monitor tool for the operator to automatically check the status of the processing taking place in each of the areas. Such a monitor tool was created using HTML technology and PERL cgi-bins so that the operator could even remotely check whether the system is behaving properly. An image of the MONITOR tool "Process Overview Page" is shown below at work.

- First column gives the request number of the processing,
- Second column indicates the area in which the processing is taking place,
- Third column indicates the type of processing taking place ("BK" for "Bulk" Reprocessing and "OF" for "On-the-Fly Reprocessing". The sys-

tem has been designed to allow to work in parallel with OFRP and BKRP requests),

- Fourth column indicates revolution or observation being processed,
- Fifth column indicates the scheduling date,
- Sixth column gives the last level of processing required (which the user can select from the ISO Data Archive User Interface),
- Seventh column gives the status of the queue under which the processing is taking place.

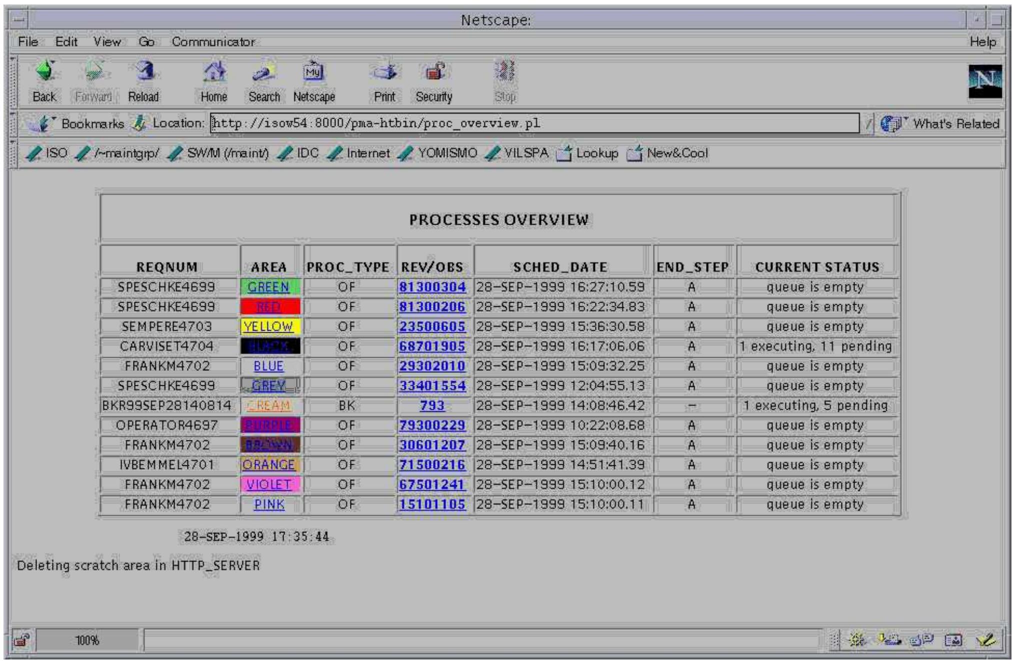

Figure 2. Monitor Tool.

Acknowledgments. The authors would like to thank Antonio de la Fuente, Neil Jenkins and Alex Scohier for their participation in the earlier stages of the project.

References

Arviset, C. et al. 2000, this volume, 191

Saxton, R. D. et al. 1998, in ASP Conf. Ser., Vol. 145, Astronomical Data Analysis Software and Systems VII, ed. R. Albrecht, R. N. Hook, & H. A. Bushouse (San Francisco: ASP), 438