NOAO DPP Document PL010

The NOAO Mosaic Pipeline Architecture

Francesco Pierfederici, Francisco Valdes, Chris Smith, Rafael Hiriart, Michelle Miller

National Optical Astronomy Observatory
Data Products Program

December 4, 2003
ASP Conf. Ser., Vol. 314 ADASS XIII, 2003, 476

NOAO Data Products Program, P.O. Box 26732, Tucson, AZ 85732

Table of Contents

List of Figures

Abstract

The NOAO Mosaic Pipeline is a fully distributed and parallel system able to efficiently process and reduce mosaic imaging data in near real time. Basic CCD reduction, removal of instrumental features (e.g. fringes, pupil ghosts and crosstalk), astrometric calibration and zero point photometric calibration are performed. The NOAO Mosaic Pipeline System is composed of a variable number of processing nodes organized in a network. Data can enter the processing network at any node, thus improving the robustness of the whole architecture. While being developed with the NOAO Mosaic Imagers in mind, the system is general enough that it could be easily customized to handle other instruments.

Keywords: Pipelines, Mosaic, Distributed Computing, High Performance Computing, IRAF, OPUS

1 Overview

The main characteristics of the NOAO Mosaic Pipeline System are:

- Science driven development.
- Based on a comprehensive Data Model.
- High degree of modularity and code reuse.
- Dynamic, freely variable number of computing nodes.
- Dynamic, XML based configuration of modules and pipelines.
- Dynamic master-slave hierarchy among the nodes.
- Dynamic routing of data through available nodes.
- Sophisticated predictive load balancing.
- Database based calibration file management.
- High degree of abstraction from the underlying implementation.

2 System Architecture

To the outside world, the NOAO Pipeline System appears as a black box (Fig. 1). Data is submitted to the Pipeline either directly from the telescope or from the NOAO Science Archive. A pipeline operator is able to constantly monitor the health and performance of the system and, if necessary, completely control the processing.

Quality control data is produced at several steps in the Pipeline, covering both basic telemetry and advanced image parameters (e.g. sky uniformity, PSF variations etc.). Monitor GUIs can subscribe to these streams of informations, enabling, for instance, the instrument scientist to monitor the performance of the instrument.

The system produces calibrated data, master calibration frames, catalogues and data quality information that can be delivered to the observer and ingested in the Science Archive.

3 Node Architecture

Processing nodes have a layered architecture, as illustrated in Fig. 2. The Processing Software (e.g. IRAF tasks, scripts, compiled code) does the actual number crunching. The Software is logically organized in modules. Modules are then grouped into standalone pipelines. Pipelines form the full processing system. Any number of instances of each module can be started, to fully exploit the processing power of the host machine.

The Black Board subsystem is responsible for making data flow through the processing modules/pipelines (modules and pipelines can be dynamically chained together at run-time, using an

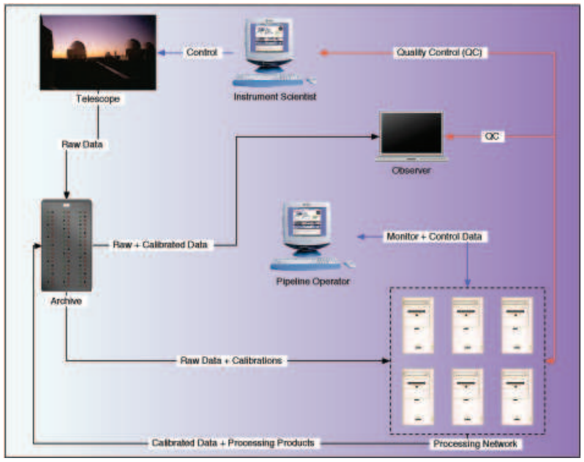

Figure 1: Schematic description of the NOAO Mosaic Pipeline architecture.

XML based configuration system). The Black Board also provides an event handling and message passing framework that individual modules and pipelines use.

The Node Manager is a high performance server, running on each node. It fully controls the operation of the Pipeline System on that node, allowing pipeline operators (via Control GUIs) to:

1. Start/stop/restart the whole Pipeline System or parts of it.
2. Control the processing of each dataset.
3. Monitor the status of the processing network.

The Node Manager also serves as load balancer. This functionality is implemented in a fairly sophisticated algorithm able to predict the load of a given processing node given its current CPU load, number of processors, number of instances of a given pipeline and the number of files in the queue.

The current architecture implements well defined interfaces for inter-machine communication and for communications with Monitor and Control GUIs. Data being processed, software and state are always kept local to each machine (having a private copy of the Black Board). The result is that each node of the processing network is an independent entity. This makes the NOAO Pipeline able to handle the failure of one or more nodes by transparently re-routing data to the available machines.

References

[1] Hiriart, R., Valdes, F., Pierfederici, F., Smith, C. & Miller, M. in this volume, [P1-21].

[2] Miller, M., Valdes, F., Smith, C., Hiriart, R. & Pierfederici, F. in this volume, [P4-31].

[3] Valdes, F., Smith, C., Hiriart, R. Pierfederici, F. & Miller, M. in this volume, [O10-5].

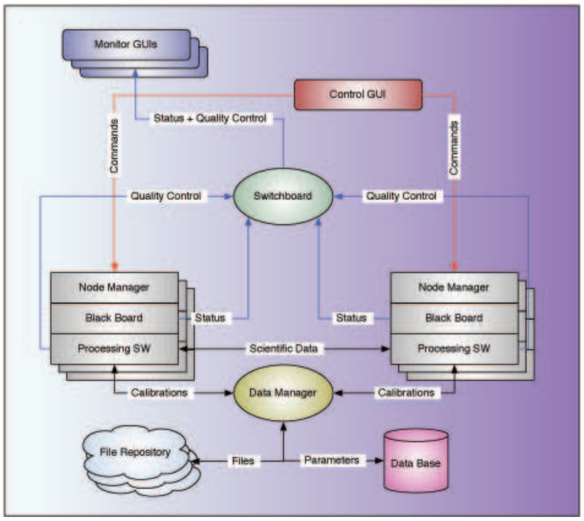

Figure 2: Schematic description of the NOAO architecture of the processing network.