Astronomical Data Analysis Software and Systems XIII
ASP Conference Series, Vol. 314, 2004
F. Ochsenbein, M. Allen, and D. Egret, eds.

The NOAO Pipeline Data Manager

R. Hiriart, F. Valdes, F. Pierfederici, C. Smith and M. Miller

National Optical Astronomy Observatories, Tucson, AZ 85719

Abstract. The Data Manager for NOAO Pipeline system is a set of interrelated components that are being developed to fulfill the pipeline system data needs. It includes: (1) management of calibration files (flat, bias, bad pixel mask and xtalk calibration data.); (2) management of the pipeline stages' configuration parameters; and (3) management of the pipeline processing historic information, for each of the data products generated by the pipeline.

The Data Manager components uses a distributed, CORBA based architecture, providing a flexible and extensible object oriented framework, capable of accommodating the present and future pipeline data requirements. The Data Manager communicates with the pipeline modules, with internal and external databases, and with other NOAO systems such as the NOAO Archive and the NOAO Data Transport System.

1. Introduction

The NOAO Pipeline is a parallel and distributed system that processes observations from the two NOAO 8K-square CCD wide-field mosaic imagers in near real time. It performs basic CCD reduction, instrumental features removal, astrometric calibration, and zero point calibration (Pierfederici 2004; Valdes, Miller, this conference).

The core of system is composed of several processing modules, which are connected to form sub-pipelines. The OPUS[1] system is used to control the process execution environment for each of the modules.

The Data Manager is a subsystem of the NOAO Pipeline that provides several data management services to the system. These services include storage and retrieval of calibration files, management of pipeline module configuration parameters, and recording of processing information.

2. Architecture

The extensibility of the Data Manager has been an important goal in the design, in order to facilitate integration of new services as the pipeline system evolves

[1] `http://www.stsci.edu/resources/software_hardware/opus`

and new requirements arise. Therefore, the Data Manager contains several components implemented as distributed CORBA objects.

Figure 1 presents the overall architecture of the Data Manager. The system is structured in three layers: the *Interface* layer, the *Service* layer, and the *Resource* layer. The Interface layer exposes the Data Manager services to the clients by calling the appropriate component(s) in the Service layer upon receiving a request. The Service layer, on the other hand, uses the components in the Resource Layer to access a number of internal services, such as persistent data in relational databases, and file storage in multiple repositories.

The Interface layer is composed of the following subsystems:

- **DMSocketRouter:** Receives requests from the pipeline modules through a lightweight socket based protocol. Request processing is asynchronous and concurrent, which enables the handling of a large number of connections and achieve good performance.
- **DMOpusBlackboards:** An experimental interface, designed to exploit the benefits of the blackboard architecture and components of OPUS, to integrate the Data Manager with the rest of the system. This feature provides another method of communication with pipeline modules.
- **DMAdministration:** Exposes interfaces for administration GUIs used by system operators.

The subsystems in the Service layer implements the "business logic" of Data Manager operations. The services are based on a rich object-oriented data model that is maintained in a PostgreSQL database. The subsystems of this layer are:

- **DMCalibrations:** Exposes interfaces for the management of calibration libraries.
- **DMParamManager:** Maintains a set of configuration parameters for each pipeline module.
- **DMHistorian:** Allows the system to keep a record of all relevant variables needed to review the pipeline processing history, and to address provenance issues for pipeline-generated data products.
- **DMConfiguration:** It maintains a model with the overall current configuration of the pipeline system. The pipeline modules can be dynamically chained together at run-time, using an XML based configuration system. The XML configuration feeds the DMConfiguration subsystem's data model, and this information is stored as part of the processing provenance of the generated data products with each pipeline run.

The Service layer uses components from the Resource layer, which is composed by the following subsystems:

- **DMFileRepositories** Maintains collections of files stored in multiple repositories, organized in hierarchical directories on the file system. Maintains a catalog of stored files.
- **DMDBManager** Provides access to relational databases. This component is divided in two parts: the interface with the DBMS, that provides persistence services to objects in the upper layers, and maintains a pool of database connections; and a set of stored procedures that are used for processing within the database.
- **DMLogger** Provides logging services for all other components of the Data Manager.

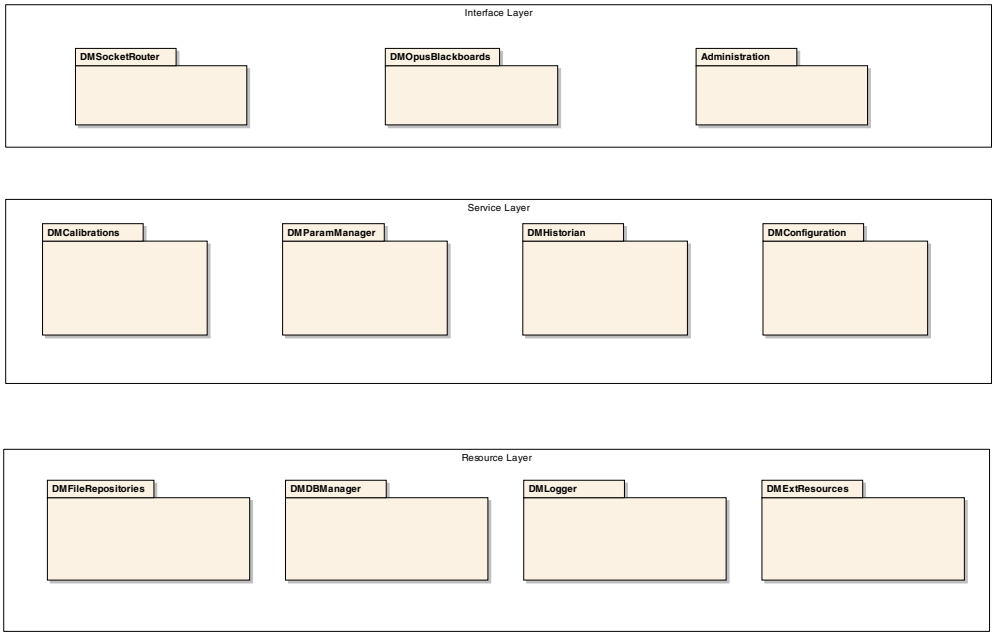

Figure 1. Overall architecture of the Data Manager system.

- **DMExtResources** Accesses external resources such as external catalogs, the NOAO Data Transport system, and the NOAO Science Archive.

3. Conclusions

The NOAO Data Manager is being developed as part of the NOAO Pipeline project, with the goal of providing an extensible, robust, and high performance solution for the data management needs of the pipeline system, as well as to allow the integration of other NOAO systems that depends on the data generated in the pipeline executions, such as the NOAO Archive system.

The NOAO Data Manager is being developed using distributed object middleware and relational databases. These technologies provide the necessary framework over which this application can be successfully constructed.

References

Pierfederici, F., Valdes, F., Smith, C., Hiriart, R., & Miller, M. 2004, this volume, 476.