Astronomical Data Analysis Software and Systems XV **P.125**
ASP Conference Series, Vol. XXX, 2005
C. Gabriel, C. Arviset, D. Ponz and E. Solano, eds.

The Chandra Source Catalog Automatic Data Processing System

John D. Grier Jr., David Plummer, and Kenny Glotfelty

Smithsonian Astrophysical Observatory

Abstract. The Chandra Automatic Processing System (AP) is a set of adaptable components designed to work in various data processing environments. An instance of this system, The Chandra Standard Data Processing System (SDP), performs the day to day processing of the data by observation (levels 0-2). A new system, The Chandra Source Catalog Automatic Processing System (Level 3 AP), is another instance of AP that generates source catalogs. Level 3 AP is different from SDP in many ways, but due to the adaptability of AP, it was put together almost entirely of existing components. This paper describes these adaptable components and the Level 3 AP system through their specific use.

1. Introduction

The Chandra Source Catalog Automatic Processing System (Level 3 AP) generates source catalogs from the data produced in Standard Data Processing (SDP). This processing is done only on non-proprietary data so instead of being part of SDP, Level 3 AP is a separate instance of the Chandra Automatic Processing System (AP). To accomplish this task, this system uses AP's core components, a Level 3 version of the pipeline processing registry, l3proc (an instance of the Data Set Processing system), the Chandra Automatic Processing Task Interface and pipeline tools built to run on a Linux Beowulf Cluster (Evans, et al. 2006).

2. AP

AP is a general Chandra data processing system that was designed to adapt to different data processing requirements. The following are the current instances of AP:

- Standard Data Processing (SDP)
  Regular automated processing of new data (Plummer, et al. 2001).
- Reprocessing (Repro)
  Bulk processing of archived data due to better calibrations and/or enhanced software.
- Special Automatic Processing (SAP)
  Reprocessing of selected pipelines that require non-standard handling for a given dataset.
- Source Catalog Automatic Processing (Level 3 AP)
  Generation of source catalogs from data produced in SDP.

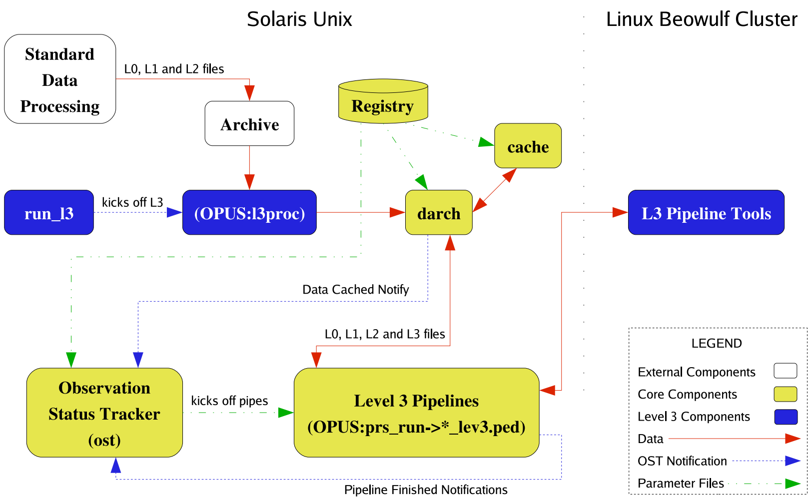

Figure 1. Level 3 Automatic Processing

3. Core Components

All instances of AP include components that handle the basic Chandra pipeline processing:

- cache

  Stores the data produced by the Chandra pipelines.
- darch (Data Archiver/Retriever)

  Archives/Retrieves data to/from cache and/or archive.
- ost (Observation Status Tracker)

  Kicks off a Chandra pipeline (through prs_run) when all the required data products are ready for it.
- prs_run

  An OPUS pipeline that is executed for each Chandra pipeline. It executes applications in stages to retrieve the required data from cache via darch, executes the pipeline tools and stores the results to cache for use by other Chandra pipelines.
- OPUS

  A generic distributed pipeline system developed by the Space Telescope Science Institute[1] (Rose 1998). This system is used by AP to organize and monitor Chandra pipeline applications.

4. Level 3 Pipeline Processing Registry

The core components can be used in all instances of AP without modification because the pipeline processing behavior is not hard-coded into the programs. Instead, this behavior was written into a set of configuration files called the

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

Pipeline Processing Registry. So the registry for Level 3, tells the core components how to run Level 3 AP. Currently, this processing is limited to processing the sources per observation. Future iterations will include processing sources over multiple observations.

5. l3proc

The Data Set Processing system (DSP) is a wrapper for OPUS that provides for easier configuration and execution of OPUS pipelines utilizing a single configuration file. An instance of this system (r3proc), is already being used to prepare and kickoff data for reprocessing III. For Level 3 AP, l3proc is an instance of DSP that maintains the Level 3 processing for a specific obi. This pipeline performs the following:

- Retrieves the required data from the archive using the Chandra Automatic Processing Task Interface (CAPTAIN)
- Stores the retrieved data to cache using CAPTAIN
- Kicks off the Level 3 pipeline processing though prs_run
- Monitors the processing
- When processing complete, transfers data to Level 3 cache

6. Chandra Automatic Processing Task Interface (CAPTAIN)

CAPTAIN is a system that allows for the selection of specific pipelines based on user selection and/or observation information. It determines and retrieves the data required to process the selected pipelines from the archive. SAP currently uses this system to allow the user to graphically select the desired Chandra pipelines for special processing. l3proc uses CAPTAIN to determine and retrieve the data needed to kickoff Level 3 processing for the specific obi. Generally, CAPTAIN provides the following functionality:

- Processes one or more observations
- Based on given observations and registry, determines which pipelines to process
- Retrieves required data products from the archive
- Stores retrieved data products to cache
- Kicks off processing
- Can run in graphical or command line mode

7. L3 Pipeline Tools

Pipeline tools are a set of applications executed by prs_run to perform the processing on the retrieved data. The behavior of these tools is defined by configuration files called Chandra pipeline profiles (*.ped). Level 3 pipeline tools are different than those executed by other instances of AP in that they are built for Linux. While AP is executed on Solaris UNIX, these pipeline tools are executed remotely on a Linux Beowulf Cluster to improve performance. The load for the Beowulf nodes is regulated by OPUS prs_run external pollers. This remote execution functionality was one of the few modifications made to AP to allow for

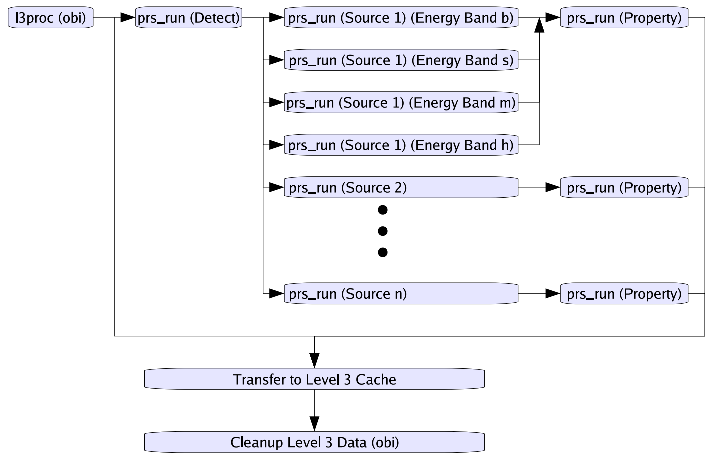

Figure 2. Level 3 Processing Flow

Level 3 processing. It does not effect previous AP instances, but does provide remote execution to future releases.

8. Processing Flow

Once Level 3 processing is initiated for an obi, an l3proc pipeline is instantiated. This pipeline gets the archived data needed for Level 3 processing and kicks off one prs_run detect pipeline. The detect pipeline then kicks off a source/band pipeline for each source and energy band. The number of sources varies per obi. If the obi is ACIS, four band pipelines per source are executed (b - broad, s - soft, m - medium and h - hard). If the obi is HRC, one band per source is executed (b - broad). For example, if the obi is ACIS and has 3 sources, 12 source/band pipelines will be executed. Once all bands for a source are complete, a property pipeline for that source is kicked off. So if the obi has 3 sources, 3 property pipelines will be executed. Once all the property pipelines for the obi have completed, the l3proc pipeline will transfer the results to the Level 3 cache directory.

Acknowledgments. Support of the development of the Chandra Source Catalog Automatic Processing System is provided by the National Space and Aeronautics Administration through the Chandra X-ray Center, which is operated by the Smithsonian Astrophysical Observatory for and on behalf of the National Aeronautics and Space Administration contract NAS-03060

References

Evans, I. et al. 2006, this volume, [P.33]

Plummer, D., et al. 2001, in ASP Conf. Ser., Vol. 238, ADASS X, ed. F. R. Harnden, Jr., F. A. Primini, & H. E. Payne (San Francisco: ASP), 475

Rose, J., 1998, in ASP Conf. Ser., Vol. 145, ADASS VII, ed. R. Albrecht, R. N. Hook, & H. A. Bushouse (San Francisco: ASP), 344