Astronomical Data Analysis Software and Systems XII
ASP Conference Series, Vol. 295, 2003
H. E. Payne, R. I. Jedrzejewski, and R. N. Hook, eds.

Systems Integration Testing of OPUS and the new DADS

Lisa E. Sherbert, Lauretta Nagel

Space Telescope Science Institute, Baltimore, MD 21218

Abstract. The Data Archive and Distribution System (DADS) will be entering the IDR (Ingest Distribution Redesign) era soon and more major functions will be shifting from the VMS platforms to various Unix platforms. As the first phase, Distribution, is delivered to testing, interfaces with OPUS and OTFR (On The Fly Reprocessing) will change. We will give a current overview of the OPUS/DADS/OTFR system and identify interface changes that will impact the operators and archive users.

1. Introduction

We illustrate the back-end HST ground systems and the types of data that flow through them, from Goddard Space Flight Center (GSFC) through various pipelines and into the data archive at the Space Telescope Science Institute (STScI). Recent changes to the platforms and software will be highlighted in SMALL CAPS. The OPUS/DADS/OTFR super-system is a mixture of platforms connected by NFS cross-mounted disks, email, and socket connections. The whole super-system is moving towards a Unix-based platform and away from the VMS platforms, making tracking of the various states interesting and the system performance definition difficult. Testing relies on stable environments, attempts to validate the system against its requirements, and then benchmarks the system performance. Here we document the current testing environment.

2. Data Flow Overview

PACOR-A at Goddard Space Flight Center now flows the HST science and engineering data to STScI. STScI picks up the engineering data and runs it through the engineering data pipeline system (EDPS), from whence it is archived. The system now runs on Unix/Solaris instead of VMS.

STScI receives the science telemetry data in the form of POD files and runs them through the data receipt pipeline. The data receipt pipeline puts the POD files into the pre-archive science pipeline. The pre-archive science pipeline reforms and calibrates the data and sends the POD files and the raw data to DADS to be archived. The headers of the raw and calibrated data are used to populate the HST archive catalog via the DADS/Ingest software. The DADS/Ingest subsystem also archives the files directly to storage media and catalogs the file locations. The storage media used in this case is Magneto-Optical (MO) platters stored in jukeboxes. The software that runs the jukeboxes

© Copyright 2003 Astronomical Society of the Pacific. All rights reserved.

lives on a Solaris Unix machine and it interfaces with both DADS/Ingest and DADS/Distribution.

Non-HST and non-FUSE data are input into MAST with CDs as the storage media. The MAST software manages user distribution requests for this data. The StarView and MAST interfaces are used to peruse the science catalog and to request HST and FUSE data. The DADS/Distribution software manages these user requests.

The Distribution software is moving off the VMS operating system onto Unix. It will now communicate with StarView and MAST via port connections instead of via email files. Distribution will make OTFR requests through IPC connections: putting a message on a shared file system which both pieces of software have access to.

The changes that will impact the end users are:

- Users can now receive data on CD or DVD.
- StarView can now generate requests for non-HST/FUSE data from MAST.
- Users may browse the archive anonymously (MAST and HST/FUSE).
- Users may be more specific about exactly what files they want returned to them by specifying the extension of interest.

3. Physical Block Overview

3.1. OPUS

Subsystem PACOR-A

- Operating System: Tru64 Unix
- Input: telemetry from HST
- Output: POD data
- Function: get POD files from GSFC to STScI
- Interfaces: PACOR-A

Subsystem SCIENCE

- Operating System: Tru64 Unix (with a VAX/VMS leftover)
- Input: POD data
- Output: Raw and calibrated data in form of FITS files (GEIS for WF2)
- Function: Change pod files into calibrated observation sets
- Interfaces: STSDAS/PyRAF software; 4 databases; Holding Tank pipeline, which converts WF2 (and FOC) to VMS GEIS

Subsystem EDPS

- Operating System: Solaris Unix
- Input: gzipped FOF files and PASS data, such as mission schedules
- Output: Jitter, astrometry, and engineering data
- Function: Processes engineering telemetry into observation logs (jitter files), instrument save files, breathing/thermal files, and astrometry (FGS science data); and processes PASS products.
- Interfaces: PASS, GSFC (CCS system), 3 databases

Subsystem OTFR

- Operating System: Tru64 Unix

- Input: specific output file requests
- Output: Raw and calibrated data, email about request success
- Function: Replaced OTFC - Reprocesses archived pod files through the SCIENCE pipeline. Automatic for ACS, NIC, STIS, WF2
- Interfaces: OPUS Science pipeline, DADS/Distribution

3.2. DADS

Subsystem INGEST

- OS VAX/VMS
- Input: Receives data over NFS-mounted disks.
- Output: Success/Failure response files over NFS-mounted disks.
- Function: Checks that it is a valid dataset, repairs keywords; Strips out information from the headers for the archive catalog; Writes data to permanent media and logs the location of the files in the housekeeping catalog.
- Interfaces: Sends OSTDB requests to catalog the header keywords in the database; Sends IPC messages to the InstallationServer (NSA); operator command line interface

Subsystem DISTRIBUTION

- OS: Solaris 2.8
- Input: Starview and MAST requests via port connections; AutoBulk requests from a directory.
- Output: Deploys data to portable media: Tape, CD, DVD; Ships data to user: NET, HOST
- Function: Validates the request (is the data proprietary or restricted?), then retrieves from permanent media and ships the data.
- Interfaces: Submits an IPC message to AcquireServer (NSA); Transfers the data via NFS-mounted disks from NSA and OTFR; Requests OTFR processing thru IPC messages

Subsystem NSA

- OS: Solaris 2.7
- Input: IPC messages pick up requests for installation/acquisition to MO platters in the jukebox
- Output: JDBC to update the database as platter/data relations; Java client/server messages help monitor the request status; Transfers files back and forth through NFS-mounted disks
- Function: Store the data on permanent media and retrieve it as requested.
- Interfaces: COTS product called Bakbone to run the jukebox; Database; Ingest, Distribution

Subsystem Archive Database (dadsops, mastops)

- OS: N/A
- Input: OSTDB JDBC requests to store header keywords and indexing information
- Output: OSTDB JDBC requests to provide StarView and MAST web pages with the above; Provides Distribution with platter/file correlations; Provides the Notification Tool (PROMPT) and AutoBulk Distribution with receipt dates and catalog information

- Function: Track location and science information of the files archived.
- Interfaces: Ingest, NSA, Distribution, AutoBulk Distribution, StarView, MAST, OTFR, PROMPT.

Subsystem MAST
- OS: Solaris 2.7
- Input: Accepts requests from the MAST web page
- Output: Transmits the data to the desired location.
- Function: Validates the request, then retrieves from permanent media and ships the data.
- Interfaces: the Solaris machine running the CD jukebox; COTS product called KPAR to run the jukebox; Database

Subsystem MAST CDs
- OS: Solaris 2.7
- Input: pick up requests for installation/acquisition to CDs in the jukebox
- Output: OSTDB to update the database as platter/CD relations; Transfers files back and forth through NFS-mounted disks
- Function: Store the data on permanent media and retrieve it as requested.
- Interfaces: COTS product called KPAR to run the jukebox; Database; MAST

Subsystem Observer Interfaces: StarView (Java GUI Application) AND MAST (Web Page)
- Input: User ID, Archive Password, Address (electronic or snail)
- Output: submit a request to DADS VIA PORT CONNECTION OR pull the data over from MAST as a tar file.
- Function: allow browsing of the MAST (including DADS) databases; allow user to indicate which datasets to retrieve and any special parameters:
  - Calibrated
  - UnCalibrated
  - DataQuality
  - JitterFiles
  - BestReferenceFiles
  - UsedReferenceFiles
  - Media Shipment: Tapes, CDs, DVDs
  - ANONYMOUS ARCHIVE USERIDS
  - EXTENSIONSONLY:<INSERT HERE> (will work even with OTFR)

4. Conclusion

We have presented an overview of the flow of HST data into OPUS science processing and engineering processing systems and from there into the Data Archive and Distribution System. The types of operating systems and recent changes to the software have been highlighted.

Acknowledgments. The authors would like to thank the members of the Data Systems Branch (and those matrixed into it) for their help in describing parts of this complicated system.