An End-to-End Solution for Archiving, Monitoring, Retrieval, and Post-Processing Archive Files for SOFIA

Lan Lin[1], Li Sun[1], Elizabeth Moore[1], Fran Nelbach Jr.[2] and Ralph Shuping[3]

[1]Universities Space Research Association
NASA/Ames Research Center, MS 211-3
Moffett Field, CA 94035, USA

[2]Nelbach Systems
Mountain View,
CA 94035, USA

[3]Space Science Institute
4750 walnut St,
Boulder, CO 80301, USA

Email: llin@sofia.usra.edu

The SOFIA Mission

The Stratospheric Observatory for Infrared Astronomy (SOFIA) is an airborne astronomical observatory comprised of a 2.5 meter infrared telescope mounted in the aft section of a Boeing 747SP aircraft that will fly at operational altitudes between 37,000 (11.3 km) and 45,000 feet (13.7 km), above 99% of atmospheric water vapor. The telescope collects radiation in the wavelength from 0.3 µm to 1.6 mm. A range of facility instruments will be used with the SOFIA to collect data.

The Data

The Mission Control and Communication System (**MCCS**) consists of the Telescope Assembly Subsystem Proxy (**TA**), Flight Management Subsystem, Water Vapor Monitor Proxy, Cavity Door Drive Systems, multiple imagers and other subsystems. During each mission, the MCCS generates a large amount of house keeping data at various frequencies as high as 50 Hz into local disks. The TA Tracker Subsystem alone logs more than 2500 data items at 20 HZ each, creating 400 kilobytes of data each second. It is imperative to archive the housekeeping data from the observatory subsystems post-flight in order to assess observatory performance and to support observatory troubleshooting and improvements. In addition, access to the data must be straightforward so that observatory staff can conduct data analysis tasks quickly and easily.

The Challenge

The MCCS subsystems log data as fast as the data is generated by writing archive files to the local disks for each subsystem. In addition to supporting high speed logging, the MCCS must also guarantee minimal data loss if the local disk fails during the flight. Meanwhile, users must be able to access the saved data from any workstation on the aircraft. At the end of the flight, data must be unloaded from the aircraft to the Data Center while the workstations that the subsystems run on remain on the plane. It is critical that data logging does not compete with the bandwidth shared by command transport and processing between the users and the TA in order to point the telescope at the right target at the right time: The round trip command/response time on the aircraft must be less than 20 milliseconds.

The Solution

1.Use portable RAID to store a backup copy of each archived file
2.Create an Archive Manager as an integrated tool to perform four major functions:
  1) Control the archiving process
  2) Monitor the archiving status
  3) Make backups at a separate location using RAID
  4) Manage a data manifest (DM) for end of flight processing and future data retrieval

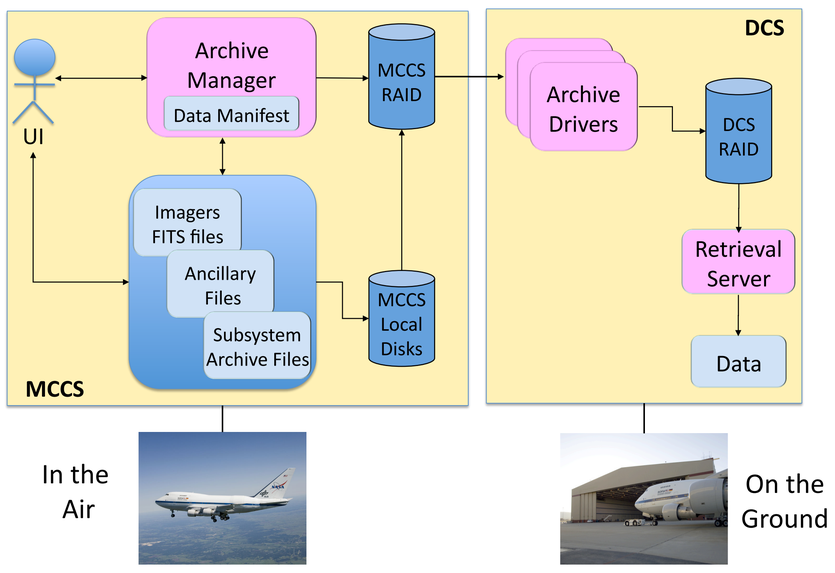
How It Works

The MCCS RAID serves as a portable storage that can be moved from the plane into the Data Center. It also makes all archived data available for in-flight data processing. The Archive Manager must carry out the above four functions without impeding the response time of the system as a whole. The performance requirement is achieved by the Archive Manager using a backup design that moves large blocks of data onto the MCCS RAID at a lower frequency instead of small chunks at high frequency. In addition, the Archive Manager itself serves as a subsystem data source that distributes the overall archive information throughout the system. This ensures that the mission director, telescope operator, and observers all have the same mission view during the flight. The Archive Manager was implemented using C++, CORBA and Java as part of the now-retired Mission Control Subsystem for SOFIA. Nonetheless, it should be possible to deploy the Archive Manager into the MCCS software environment with some updates. The Archive Manager automatically prepares the housekeeping data files for ingestion into the SOFIA Data Cycle System (**DCS**) Archive.

After Landing

At the end of a flight, the archive data files are transferred off the aircraft and ingested into the DCS archive database where they are made available to SOFIA Science and Mission Operations Staff via a simple web page interface. Once downloaded to the users desktop, the files can be opened and analyzed using the Archive Post-Processing Tool which provides a number of data analysis functions. The Archive Manager, DCS Archive, and Archive Post-Processing Tool provide and integrated solution for capturing, archiving, retrieving, and analyzing the huge volumes of housekeeping data produced by the SOFIA observatory subsystems.

SOFIA Archive Management Block Diagram



Individual Software Subsystem Archive Control Process

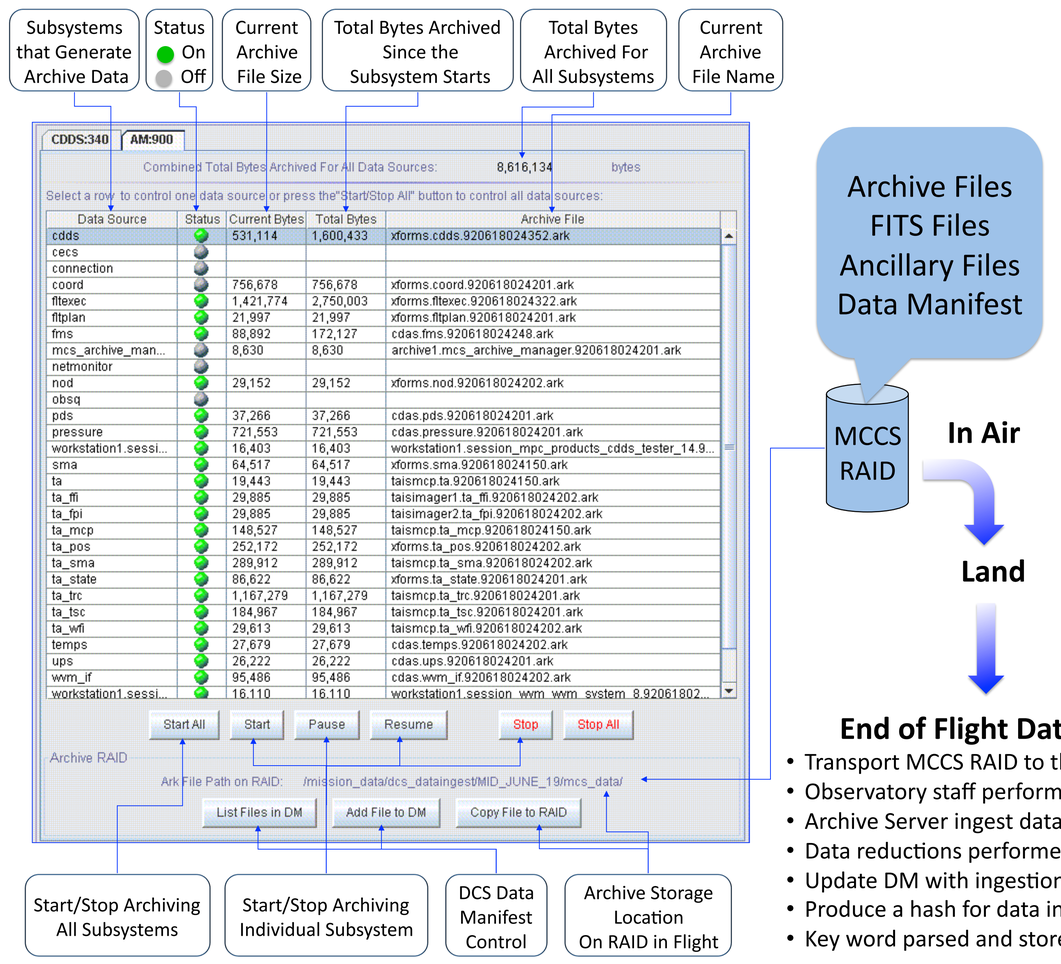
- Subsystem comes up, opens archive file
- Determine archiving state (on/off) and maximum file size from it's configuration file
- Collect data and process commands
- Write the new data into the archive file on local disk
- Provide UI and Archive Manager with the archive status
- Close the archive file when reach configured max size or received the stop command
- Notify the Archive Manager archive available, pass along file name and creation time

Archive Manager Processing

The Archive Manager carries out various tasks according to what the triggers are:

- Receive UI command to initiate archiving
  - Determine set of active subsystem
  - For every subsystem
    - Notify subsystem to start archiving
    - Subscribe to subsystem's archive status
- Receive UI commands to configure archiving, pass along commands to subsystem
  - Start, pause, resume, stop, change sample rate to every nth sample
    - Receive archive file complete status from subsystem
  - Move archive files to RAID
  - Write Data Manifest entry
- Receive UI command to initiate end-of-flight processing
  - For every subsystem
    - Notify subsystem to stop archiving
    - Receive archive file complete status from the subsystem
    - Move archive files to RAID
    - Write Data Manifest entry
  - Notify UI Data Manifest entries complete

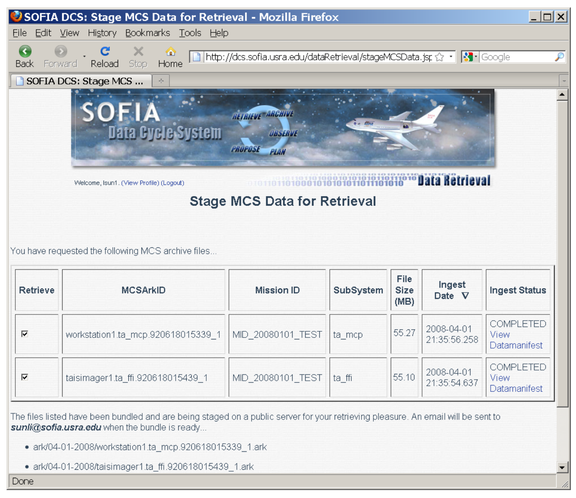
Archive Manager in Action



End of Flight Data Ingestion

- Transport MCCS RAID to the Data Center
- Observatory staff perform data ingestion
- Archive Server ingest data according to DM
- Data reductions performed
- Update DM with ingestion status of each file
- Produce a hash for data integrity check
- Key word parsed and stored

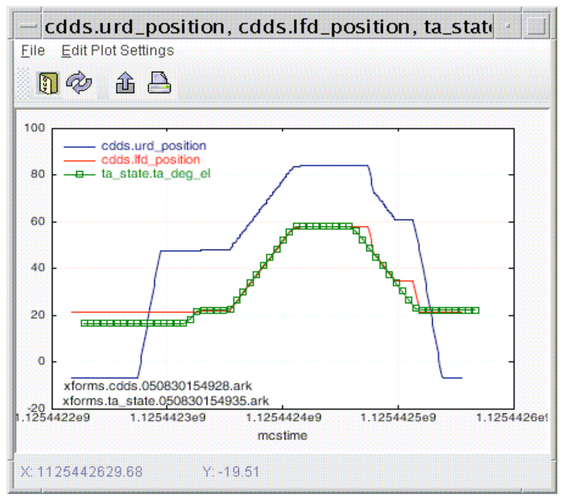
Data Staging



Data Retrieval

- From the web, enter search criteria
- Find files and request them
- The Retrieval Server creates a compressed file, stores into a download site and notifies the user when data is ready
- User downloads the data and its hash
- Check the file integrity against its hash

Post Flight Processing



Example Plot from Archive Files

- From the web, download data
- Use any one of the sets of the post-processing tools to process data
- This plot shows the relationship of Cavity Door position and the Telescope elevation using the Archive Post-Processing Tool . (Data collected during a simulated mission)

Conclusion

SOFIA presents a unique data logging and transfer problems due to the nature of the airborne observatory. Archive Manager, Data Manifest and the MCCS RAID work together to preserve data from the moment they are created in the air to the final storage on the ground. The web-based SOFIA Data Cycle System provides access to archived data for staff and researchers, both in the Science and Mission Operations Centers and remotely.

ADASS XIX, Astronomical Data Analysis Software and Systems, October 4-8, 2009, Sapporo, Japan