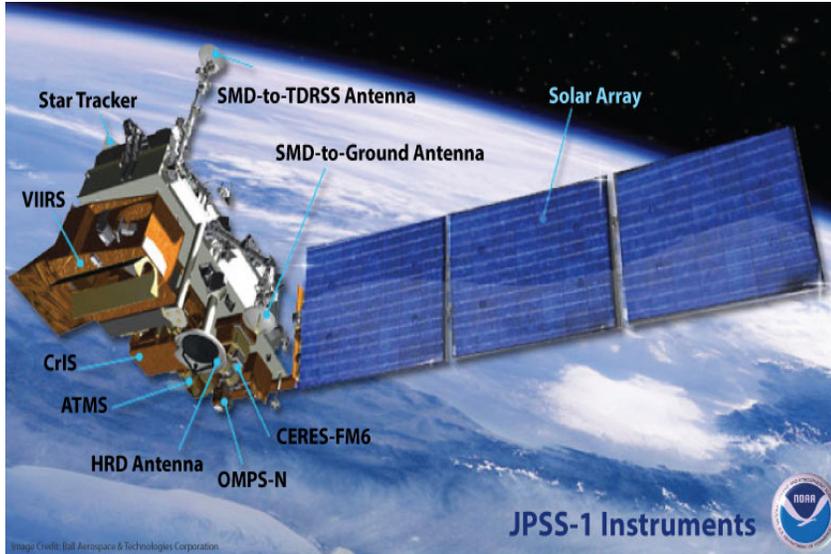# Joint Polar Satellite System - JPSS

About the Project

The problem.

For NOAA and NASA, the data problem with their Joint Polar Satellite System (JPSS) is how to best handle a large volume data stream from the five different instruments on a satellite. JPSS is a new generation of low Earth orbiting satellites that monitor environmental conditions and provide data for long-range weather and climate forecasts.

Here is NOAA's description of one of the instruments from a 2013 news release:

CrIS is the Cross-track Infrared Sounder. CrIS is the first in a series of advanced operational sounders that provides more accurate, detailed atmospheric temperature and moisture observations for weather and climate applications. This high-spectral resolution infrared instrument measures the three-dimensional structure of atmospheric temperatures, water vapor and trace gases. It provides more than 1,000 infrared spectral channels at an improved horizontal spatial resolution and measure temperature profiles with keen vertical resolution to an accuracy approaching 1 Kelvin (the absolute temperature scale). This information helps significantly improve prediction, including both short-term weather "now casting" and longer-term forecasting. It also provides a vital tool for NOAA to continuously take the pulse of the planet and assist in understanding major seasonal and multi-year shifts.

The solution: HDF5 and custom tools.

We know from other projects such as NASA's Earth Science Data and Information System (ESDIS) project that HDF Group software can be used to store large amounts of climate data. ESDIS has in 13 years (as of September 30, 2013) archived in our software 9.8 petabytes of climate data. Each petabyte is a million gigabytes.

How might data be extracted in a timely manner from such a large dataset? A data granule holds the data from a short period of observation by an instrument. Data is stored by granule in HDF5 files. With a custom tool built by The HDF Group, data granules can be aggregated and extracted. This means that only the data for a certain period of time and for a limited location need be retrieved from the data for study. This is much more efficient than having to open the entire file to see any amount of data. Since the archived data is not changed, data granules can be extracted repeatedly, and the data files themselves only need to be downloaded once.

Team Members

Principal Investigator
Elena Pourmal

Project Manager
Joel Plutchak

The HDF Group
Scot Breitenfeld
Allen Byrne
Albert Cheng

Mark Evans
Jordan Henderson
Aleksandar Jelenak
Barbara Jones
Larry Knox
Mike McGreevy

Software

The software that The HDF Group has developed for the JPSS project is described below.

Tool Development

The HDF Group developers created and currently support the following tools for the JPSS project:

- h5edit

With h5edit, users can edit attributes in an HDF5 file.

- h5augjpss

With h5augjpss, users can modify a JPSS product file to make it accessible by netCDF-4 based applications.

- nagg

The nagg tool was created to provide individual users the ability to rearrange product data granules from downloaded files into new files with aggregations or packaging that are better suited as input for a particular application.

- Chunking and Compression Performance (CCP) Tool

This prototype tool allows users to access their data files using different parameters such as chunking sizes, compression methods, access patterns, and chunk cache settings. The tool provides performance statistics to help users to find the optimum parameters to create and sccess their HDF5 files.

Library Development

JPSS data is distributed in HDF5 files containing raw data and indexing metadata that allows fast access to the raw data. The HDF Group continues to develop software libraries and tools to improve access to this data. As part of this effort, The HDF Group has created a library of C and Fortran routines to access and manipulate data referenced by object and region references and to access and manipulate data packed into integer values. We continue to seek feedback from JPSS applications developers and users, as well as from the wider HDF5 community, and will improve this library as requested.

Supported Systems

The HDF Group is maintaining HDF5 software on the following systems used by JPSS:

- Linux 32 and 64-bit
- AIX 5.3 and 6.1
- Windows 32 and 64-bit
- Mac Intel OS X 10.5 and later

Documentation

The latest versions of documentation for software developed for the JPSS project are available for download. See the list below.

- **High-level library for handling HDF5 object and region references**
  - Reference Manual
  - User's Guide (pdf)
  - C examples
  - Fortran examples
- **h5edit**
  - Reference Manual (pdf)

- User's Guide (pdf)
  - Definition of the h5edit Command Language (pdf)
  - H5edit BackupVFD Atomicity Performance Study (pdf) (docx)
  - Previous Study: H5edit Atomicity Performance Study
- **h5augjpss**
  - Reference Manual (pdf)
  - User's Guide (pdf)
  - Design document (pdf)
- **nagg**
  - Reference Manual (pdf)
  - User's Guide (pdf)
  - Tutorial slides
  - Packaged Tutorial demo files and 64-bit Linux nagg 1.4.0 executable

Downloads

The latest versions of the software developed for the JPSS project are available for download. See the list below.

High-level Library for handling HDF5 object and region references Release 1.1.5 July 20, 2016

The library contains C and Fortran APIs to:

- Get information and read data pointed to by region references
- Create an array of region references using paths to datasets and corner coordinates of hyperslabs
- Create a dataset and write data pointed to by region references
- Copy data pointed to by region references
- Retrieve data packed in an integer (quality flags)

The 1.1.5 release is a minor release. It was tested with HDF5-1.8.17 and HDF5-1.10.0-patch1.

Please see the Release Notes for complete details regarding this release.

HL REGION 1.1.5 Source Code

Source Code	Unix (tar)	Windows (zip)

Check-sum (MD5 format) for source tar file

HL REGION 1.1.5 Pre-built Binary Distributions

The pre-built binary distributions in the table below include the HL REGION libraries and include files.

Platform	Compiler
Linux 2.6 CentOS 6 x86_64	gcc, gfortran 4.4.7 (w/Fortran)
Windows (64-bit)	CMake VS 2013 C, gfortran
Windows (64-bit)	CMake VS 2015 C, gfortran

nagg Release 1.6.2 July 22, 2016

nagg is a tool for aggregating JPSS data granules from existing files into new files with a different number of granules per file or different combinations of compatible products than in the original files. The tool was created to provide individual users the ability to rearrange NPP product data granules from downloaded files into new files with aggregations or packaging that are better suited as input for a particular application.

The 1.6.2 release provides an environment variable to override the limit on the total number of granules processed at run time and adds two new command options:

- --onefile command option to aggregate granules from all input files into a single aggregation, creating one output file for packaged output or one output file for each product for unpackaged output

- --nofill command option to suppress creation of fill granules when granules are missing from a time sequence or to match a compatible product

The tool was tested on Linux 64-bit systems. For more information on this release, see the RELEASE.txt file.

Source code and binaries

(For earlier versions, see: All Releases)

---

h5edit Release 1.3.1, November 17, 2014

The h5edit tool is a command-line tool to edit HDF5 files. The current version is limited to operations on attributes only. It supports:

- Creation and deletion of attributes for groups and datasets
- Integer, floating point, and string datatypes for attributes

The h5edit 1.3.1 release contains the following changes:

- The software was ported to Mac OS X 10.9 Mavericks.
- It includes a performance study on the Backup VFD feature that was implemented in the 1.3.0 release. This study has been shown to improve execution speed for incremental atomicity on large sized files. For example, the execution time dropped from 3.8 seconds without the Backup VFD to 0.59 seconds with the Backup VFD (a 600% speed up) for a 500 MB data file running nine successful h5edit commands. A Linux CentOS 5 system was used for the study. More details can be found in the H5edit with Backup Virtual File Driver Atomicity Performance Study.

See the Release Notes for complete details on this release.

Source code (tgz format)

---

h5augjpss Release 1.0.0 August 15, 2011

The h5augjpss tool is designed to modify a JPSS HDF5 product file to be accessible by the netCDF-4 version 4.1.3 library. The tool:

- Adds information stored in the XML product file and in the geo-location product file
- Hides HDF5 objects unknown to the netCDF-4 library
- Restores access to hidden objects

The tool was tested on Linux 32 and 64-bit systems. For more information, see the RELEASE.txt file.

Source code (tgz format)

See the README.txt file in the source code for information on building and running h5augjpss.

Chunking and Compression Performance Tool prototype July 17, 2016

The Chunking and Compression Performance (CPP) tool is a prototype designed to help assess the effect of using various file parameters and access patterns on performance and storage.

The tool was tested on Linux 64-bit systems. For more information, see the README.txt file.

Source code (tgz format)