



Accessing HDF Data in the Cloud via OPeNDAP Web Service

Kent Yang

Software Engineer/NASA EED-3 contractor

myang6@hdfgroup.org

2024 ESIP Summer Meeting



Topics Overview

- Accessing HDF* Data in the Cloud via dmrpp**
- Direct IO Performance Improvement
- Work in progress to access NASA HDF4 and HDF-EOS****2 files

*Hierarchical Data Format

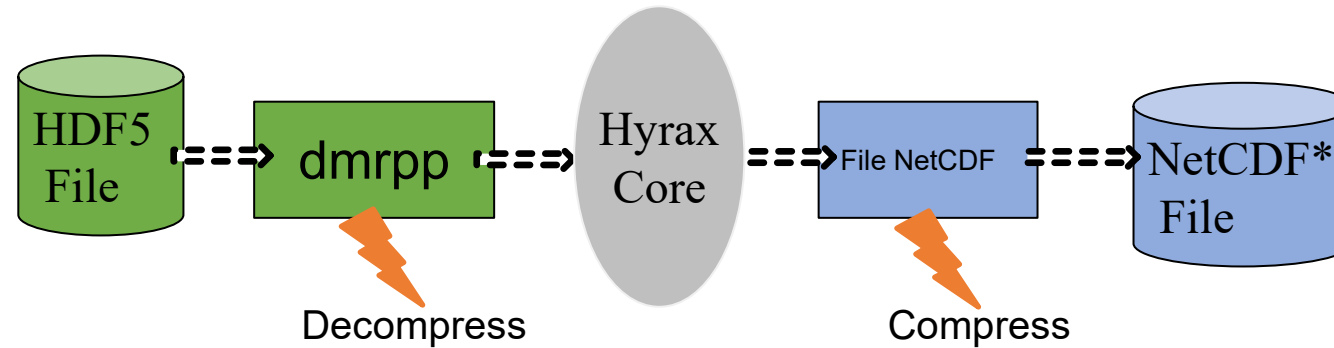
** Dataset Metadata Response Plus Plus

*** Input Output

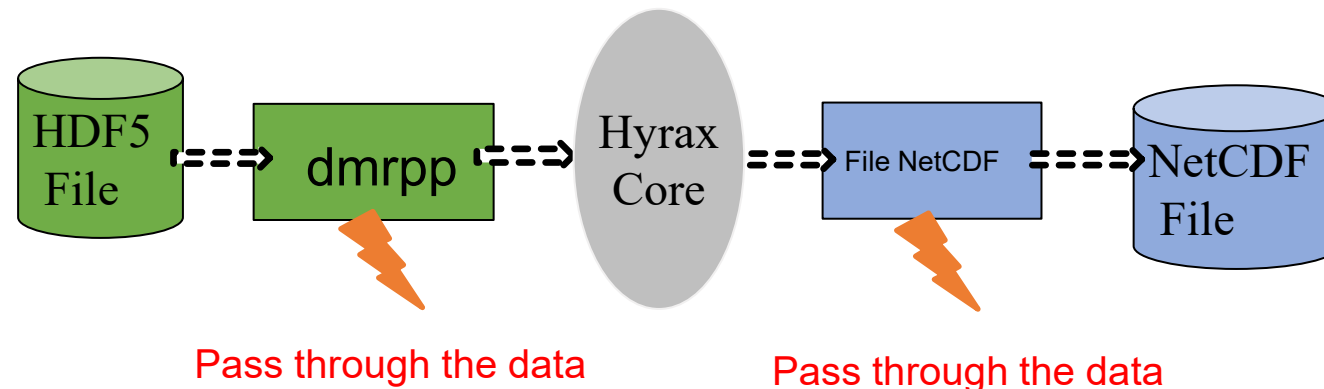
**** Earth Observing System

Direct IO Performance Improvement Concept

General Approach



Approach with Direct IO



* Network Common Data Form

Hyrax Server Response Time Speed-up With Direct IO

Product Sample File	File Size (MB)	Response Time without Direct IO (Seconds)	Response Time with Direct IO (Seconds)	Speed-up in Response Time by using Direct IO
GHR SST*	9	2.8	0.2	14 X
TROPOMI**	292	26.6	1.8	15 X
SSM I***	1.4	0.5	0.3	1.7 X

*: Group for High Resolution Sea Surface Temperature

** : TROPospheric Monitoring Instrument

***: Special Sensor Microwave Imager

Big Files With Direct IO

Product Sample File	File Size (GB)	Response Time with Direct IO (Second)	Server Response Message Without using Direct IO
Daymet	3.5	45	The maximum response time limit(165 seconds) is exceeded.
MODIS* Derived	4	65	Insufficient memory
CH4 Level 4	11	Direct IO feature is not used because it doesn't contain any compressed variable.	The maximum response time limit(165 seconds) is exceeded.

* Moderate Resolution Imaging Spectroradiometer

Facts for the Direct IO Feature

- Hyrax will use direct IO automatically for those cases when end users request to obtain the whole array of the selected variable(s) and those variable(s) are compressed.
- This process is entirely transparent to the end users.
- Direct IO doesn't work for some old dmrpp files if they don't contain the key information needed for using the Direct IO feature. These dmrpp files need to be regenerated to take advantage of the Direct IO feature.

Direct IO Performance Improvement Summary

- Can greatly reduce server computation time
- Can greatly reduce server memory usage
- **Use HDF5 direct chunk IO API*s**
- **The feature is in the current Hyrax release**

Accessing HDF4 and HDF-EOS2 via dmrpp

- Map HDF4 to DMR*
- Access HDF4 via dmrpp
 - Not only handle data stored in chunking and contiguous layouts
 - Also need to handle data stored in linked blocks
 - Handle HDF-EOS2/HDF4 geolocation data
 - Not stored as HDF4 variables
 - Need to calculate them based on the metadata information
 - Save the data in a proper way

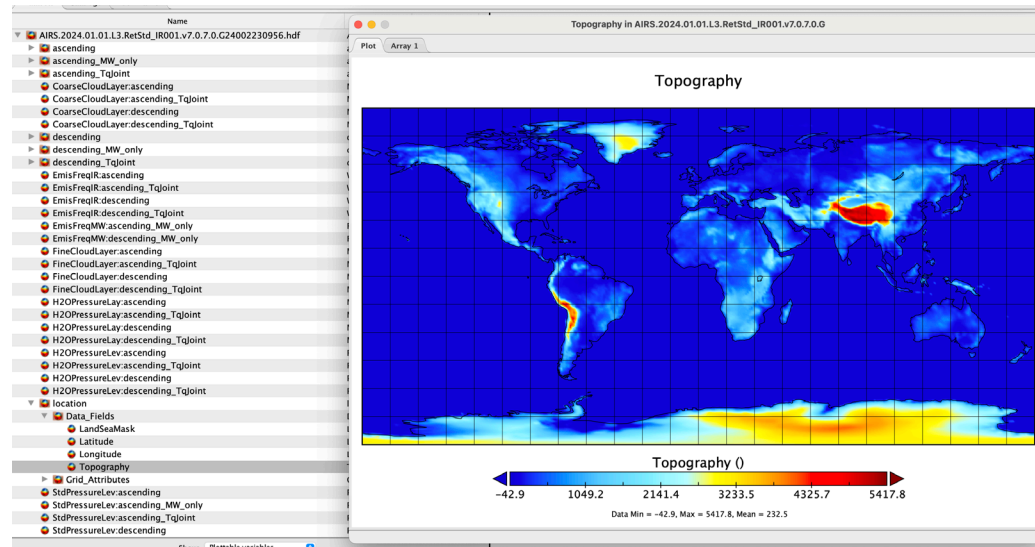
Current Status

- We can successfully map and access the sample NASA HDF4 and HDF-EOS2 products via dmrpp.
- We are still working on a better way to store HDF-EOS2/HDF4 geolocation data.

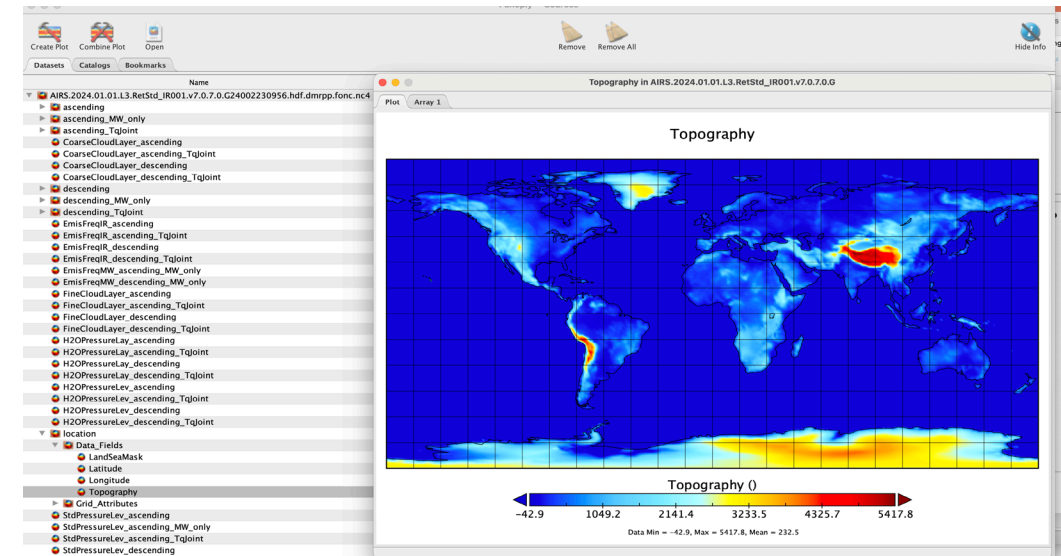
AIRS* Local HDF4 file via dmrpp's netCDF-4 file

- Panoply screenshots of variable Topography
 - The identical plots show the dmrpp module can successfully access this HDF4 file

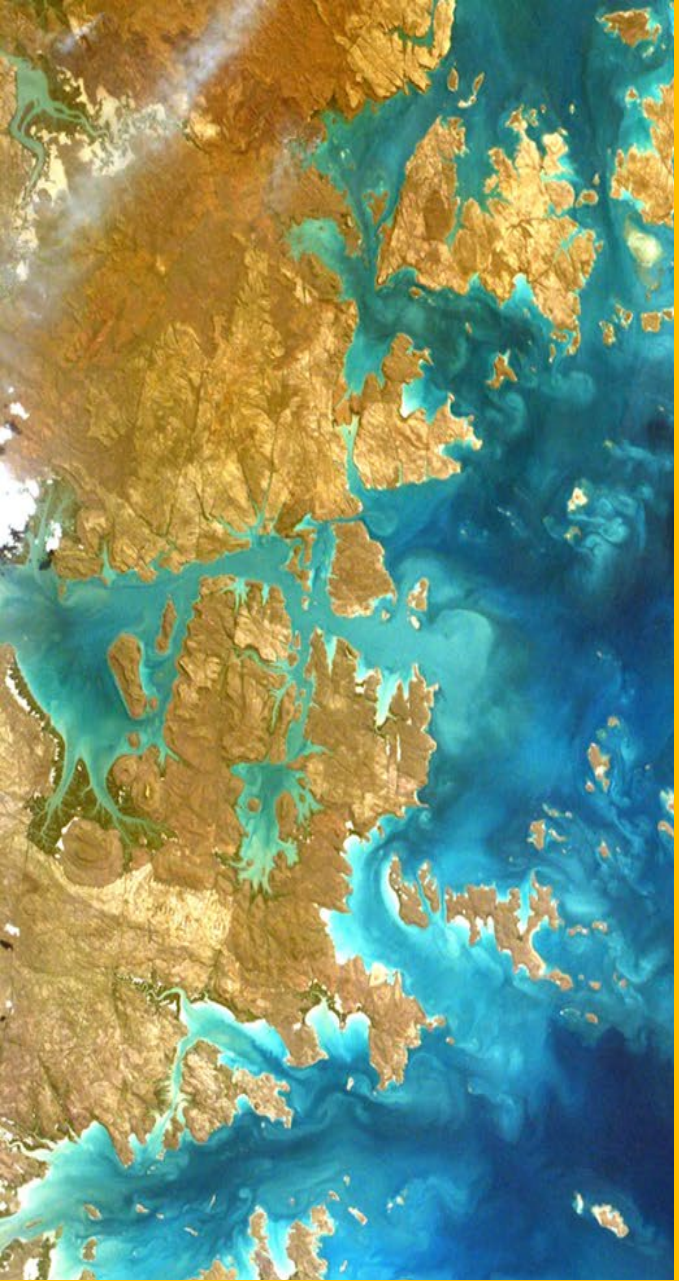
Local HDF4 file



netCDF-4 file via dmrpp



Thank you!



This work was supported by NASA/GSFC under
Raytheon Company contract number
80GSFC21CA001