Moving from HDF4 to HDF5/netCDF-4

Elena Pourmal, Kent Yang, Joe Lee

epourmal@hdfgroup.org , myang6@hdfgroup.org, hyoklee@hdfgroup.org

The HDF Group

Outline

- Difference between HDF4 and HDF5 Data model and capabilities

- Moving data and applications from HDF4 to HDF5

- Taking advantage of HDF5 when converting data

- Creating compatibility with netCDF-4 when migrating data from HDF4 to HDF5

EOSDIS

HDF4 and HDF5 Data Models

HDF4 Objects

- A **scientific dataset** (SD), a multidimensional array **with dimension scales**

- An **8-bit raster image** (DFR8), a 2-dimensional array of 8-bit pixels

- A **24-bit raster image** (DF24), a 2-dimensional array of 24-bit pixels

- A **general raster image** (GR), a 2-dimensional array of multi-component pixels

- An 8-bit color lookup table or **palette** (DFP), a 256 by 3 array of 8 bit integers

- A table (Vdata), a **sequence of records**

- An **annotation** (AN), a stream of text that can be attached to any object

- A **group** (Vgroup), a structure for grouping objects

HDF5 Objects

- **A dataset, a multidimensional array of records; no dimension scales (HL library)**

- **HDF5 dataset with attributes**

- **HDF5 dataset with attributes**

- **HDF5 dataset with attributes**

- **HDF5 dataset with attributes**

- **HDF5 dataset with attributes**

- **HDF5 one-dim dataset of the records**

- **Attributes**, scale down version of HDF5 dataset

- **A group, a structure for grouping objects**

HDF4 and HDF5 Capabilities

HDF4

- 2GB limit on file size
- Limit on number of objects (~20000)
- One unlimited dimension; dataset cannot be compressed
- Compression doesn't require chunking storage
- Limited number of compression methods
- Limited number of supported datatypes
- Numerical data is always in BE format

HDF5

- No limit on file size
- No limit on number of objects
- Up to 32 unlimited dimensions; dataset can be compressed
- Compression requires chunking storage
- Custom compression methods supported
- Datatypes of any complexity
- User-defined endianess

Moving data and applications from HDF4 and HDF5

- Moving Data
  - H4h5tools conversion toolkit
    - Mapping Spec https://support.hdfgroup.org/HDF5/doc/ADGuide/H4toH5Mapping.pdf
    - Library
    - Command-line tools h4toh5 and h5toh4

- Moving Applications
  - Software has to be rewritten if using HDF library
  - HDF-EOS2 and netCDF based applications require minimum rework

EOSDIS

Taking advantages of HDF5 and avoiding pitfalls

- Data endianess
- Chunked storage for compression and data extensibility
  - Contiguous vs. chunked storage
  - Chunk sizes
- Compression methods in HDF4 and HDF5
- Using strings in HDF5
  - HDF4 fixed character arrays vs HDF5 strings
- Working with dimension scales

EOSDIS

Creating compatibility with netCDF-4

- HDF5 files can be read by netCDF-4 library and tools unless they use features that are not supported by netCDF-4
- See unsupported HDF5 features in netCDF-4 http://www.unidata.ucar.edu/software/netcdf/docs/faq.html#fv15
- To assure maximum interoperability do not use
  - Hierarchical HDF5 structure (nested groups)
  - HDF5 user-defined datatypes
  - HDF5 compound datatypes
- http://www.unidata.ucar.edu/software/netcdf/docs/interoperability_hdf5.html

EOSDIS

Creating compatibility with netCDF-4

Tools	NetCDF-3 Format	NetCDF-3 Format following CF Conventions	NetCDF-4 Format following netCDF-4 generic model	NetCDF-4 Format following netCDF-4 classic model	NetCDF-4 Format following netCDF-4 classic model and CF
HDF-EOS5 augmentation tool	No	No	Yes. Note: Users have flexibility to specify dimension scales. Tested with NASA Aura files.	No	No
HDF-EOS5 to netCDF-4 converter	No	No	Yes. Note: Users have no control. The converter tries to map the HDF-EOS5 dimension information provided by the file to netCDF-4 enhanced model.	No	No

EOSDIS

DM_PPT_NP_v02

This work was supported by NASA/GSFC under Raytheon Co. contract number NNG15HZ39C

Raytheon

DM_PPT_NP_v02