HDF5 Dimension Scales in NetCDF4
March 24, 2005

With HDF5-1.8, HDF defines a standard for storing Dimension Scales, along with a standard API for managing dimensions [1]. HDF5 Dimension Scales are a conceptual superset of the features of netCDF-4 dimensions and coordinate variables, i.e., netCDF-4 dimensions and coordinate variables map to specific cases of the HDF5 model.

The netCDF-4 implements the semantics dimensions and coordinate variables, and does not require the HDF standard attributes or API. However, HDF-aware applications will not be able to interpret netCDF dimensions at all, without the netCDF-4 library or equivalent.

Ideally, the netCDF-4 library should, as a side-effect, follow the HDF5 Dimension Scale specification. This would have minimal impact on netCDF-4 applications, but would make the dimensions and coordinates interpretable by any HDF-aware application. The main impact is a few "extraneous" attributes on the HDF5 objects, which have no meaning to netCDF-4.

Fortunately, the Dimension Scale API (H5DS) makes it quite easy to implement the HDF5 standard. The main issue is how to implement it in the netCDF-4 code.

The Mapping

Basically, the HDF5 Dimension Scale specification defines a standard for:

- Storing a "label" for a dimension (which has no data)
- Distinguishing a Dataset to be a Dimension Scale
- Associating Dimension Scales with dimensions of the Dataspace of a Dataset.

In the netCDF-4 model, each Variable has one or more dimensions, each of which has a name. In some cases, a Variable may be defined that has the same name as a dimension. In this case, that Variable is a Coordinate Variable. The following table summarize the proposed approach.

NetCDF-4	HDF5	Changes to netCDF-4
Variable has Dimensions	Set HDF5 Label for each dimension to the name of the ncdim (after H5Dcreate)	Add code to var_create_dataset()
Coordinate Variable	A coordinate variable should be designate to be an HDF5 Dimension Scale (after H5Dcreate of the variable)	No existing code. Create a new, optional API function, e.g., nc_is_coord_var(…) Calls H5DSmake_scale(…) Recommended as a future convention.
Variable has Dimension with Coordinate Variable	Attach the HDF5 Dimension to the dimension of the Dataspace (H5DSattach)	Add code to var_create_dataset(). Requires coordinate variables be identified somehow.

Issues

The big issue is how to handle coordinate variables. HDF5 wants to have them explicitly marked. This isn't difficult, but does require netCDF-4 to recognize them and call HDF5 to mark them.

I suggest that a new, optional, function be added to netCDF-4. This function is something like nc_is_coord_var(), and it's primary effect is to make it an HDF5 Dimension Scale (if possible).

A well-behaved program will work something like:

- Define dimensions
- Define coordinate vars
- Define vars using dimensions and coordinate vars

In the event that a coordinate variable is defined after a Variable that used the corresponding dimension, nc_is_coordvar() might be designed to go back and attach the new dataset to all the Variables that used its dimension. This requires calling H5DSattach() for the dimension of every Dataset that used it.

Summary

This note proposes how to implement HDF5 Dimensions for an NetCDF-4 file.

If implemented, any HDF-aware application will be able to discover the netcdf dimension names and coordinate variables for the HDF5 Datasets that represent the netcdf Variables in the file.

References

1. HDF5 Dimension Scale Specification and Design: Version1.0 (March 1, 2005), http://hdf.ncsa.uiuc.edu/RFC/H5DimScales/.