EGU2011-8288: Recent Development in NetCDF Libraries and Formats

Edward Hartnett, Unidata/UCAR, Boulder, Colorado, USA, EGU, April, 2011

NetCDF

NetCDF (network Common Data Form) is a set of software libraries and machine-independent data formats that support the creation, access, and sharing of array-oriented scientific data.
For twenty years, the netCDF format and libraries have been developed and supported by Unidata.

In 2008, the netCDF-4.0 release introduced new performance features, and an enhanced data model. The 4.1.1 release in 2010 added remote access features. Backward code and data compatibility has been maitained.

Getting Started

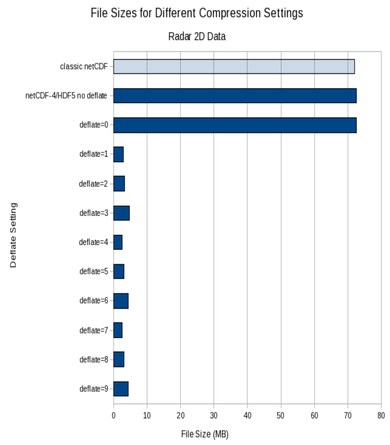
On many systems netCDF is available from package repositories (ex. yum install netcdf).

Extensive documentation, training materials, and example code available at the Unidata website: www.unidata.ucar.edu.

Many third-party tools support netCDF data.

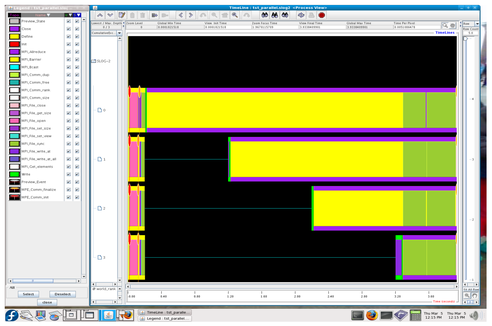
Data Compression

Built-in data compression allows data variables to be compressed/uncompressed on the fly with the gzip algorithm. Uncompressing the data happens automatically, and is transparent to the reader.



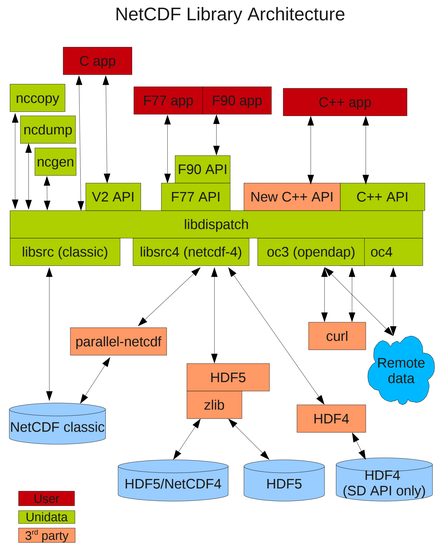
Parallel I/O

For high performance computing, parallel I/O can yield significant performance benefits. NetCDF supports parallel I/O of netCDF/HDF5 files with the parallel I/O features of the HDF5 library. Parallel I/O to classic and 64-bit offset format is provided with the help of the parallel-netcdf library from Argonne/NorthWest University.



Software Architecture

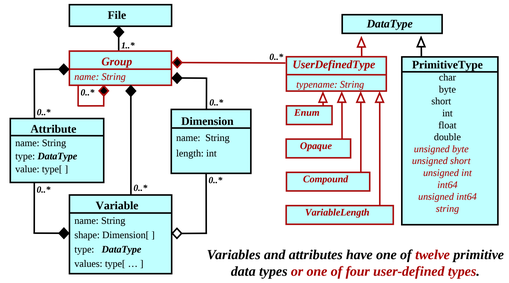
The architecture of the C/Fortran/C++ libraries has changed to support the use of other libraries. NetCDF can still be built to only use the classic netCDF library.



Enhanced Data Model

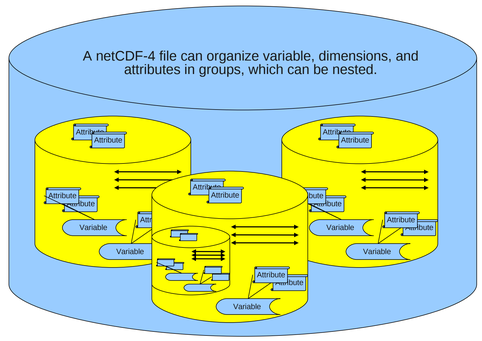
The enhanced data model allows for more complex representations of data.

A file has a top-level unnamed group. Each group may contain one or more named subgroups, user-defined types, variables, dimensions, and attributes. Variables also have attributes. Variables may share dimensions, indicating a common grid. One or more dimensions may be of unlimited length.



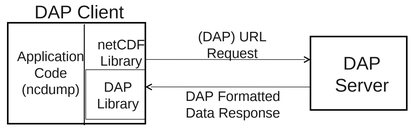
Variables and attributes have one of twelve primitive data types or one of four user-defined types.

The enhanced data model is a super-set of the classic netCDF data model.



Remote Data Access

The built-in opedap client allows netCDF applications to access data stored on remote data servers, as if it were local.



C/Fortran/C++ NetCDF-4.12

The most recent release of the netCDF C/Fortran/C++ libraries is version 4.1.2, available at the Unidata web site. It contains the following features:

- Builds shared libraries by default.
- Speedup in opening large files reading and writing many objects.
- Better handling of ncgen character datalist constants.
- Greatly reduced memory use, better parallel I/O performance.
- nccopy supports compression, chunking, shuffling.
- opendap compatibility improvements.
- Refactored to allows further extensions.
- Source at http://svn.unidata.ucar.edu/repos/netcdf

NetCDF-Java 4.2.23

The most recent release of netCDF-Java is version 4.2.23. Some of the features in the 4.2.x series include:

- FMRC aggregations dynamically adjust to changes in model output.
- FMRC "Best" dataset can be parameterized by forecast time.
- TDS admin can choose which FMRC datasets to expose.
- Caching of Grid dataset info.

Commitment to Backward Compatibility

Because preserving access to archived data for future generations is sacrosanct:

- NetCDF-4 provides both read and write access to all earlier forms of netCDF data.
- Existing C, Fortran, and Java netCDF programs will continue to work after recompiling and relinking.
- Future versions of netCDF will continue to support both data access compatibility and API compatibility.

Unidata

Mission: To provide the data services, tools, and cyberinfrastructure leadership that advance Earth system science, enhance educational opportunities, and broaden participation.

Unidata, funded primarily by the National Science Foundation, is one of eight programs in the University Corporation for Atmospheric Research (UCAR) Office of Programs (UOP). UOP units create, conduct, and coordinate projects that strengthen education and research in the atmospheric, oceanic and earth sciences.

Unidata is a diverse community of over 160 institutions vested in the common goal of sharing data, and tools to access and visualize that data. For 20 years Unidata has been providing data, tools, and support to enhance Earth-system education and research. In an era of increasing data complexity, accessibility, and multidisciplinary integration, Unidata provides a rich set of services and tools.

Unidata: www.unidata.ucar.edu