
Subject: Re: CDF vs HDF

Posted by [Paul Van Delst\[1\]](#) on Tue, 30 Mar 2004 21:03:01 GMT

[View Forum Message](#) <> [Reply to Message](#)

Michael Wallace wrote:

> This isn't an IDL question exactly, but I figured that there are several
> here who have to deal with these file types. Is there any real
> difference between CDF and HDF? I have looked at some web pages that
> describe the differences, but it doesn't mean much to me since I've
> never delved into either one. The only reason I'm even looking at these
> file types is that these seem to be the types preferred by the data
> center we will be working with later on for long-term archival of our
> data. If it means anything, my core data sets are nothing more than
> really big arrays. Later on, other data like PNG images may be included.
>
> Unless there's some big difference between the two, I'm going to go with
> the one that's easier to work with in IDL. So, that brings me to the
> question, is one easier to work with than the other? Or do I flip a
> coin to determine which one to use?

Oh my goodness... use netCDF. CDF is different from netCDF (I don't know exactly how) but IDL has all the netCDF API stuff built in (it may have the CDF stuff too...I dunno). And netCDF is waaaay easier to use than HDF (IMO). HDF may be more "powerful" but my datasets are like me, nice and simple, and HDF is generally overkill.

netCDF has APIs in just about every flavour of language you could ever want too (Fortran90, C, C++, Java, perl, R, matlab, python, etc etc).

Check out <http://www.unidata.ucar.edu/packages/netcdf/> for more info.

cheers,

paulv

Subject: Re: CDF vs HDF

Posted by [mmiller3](#) on Tue, 30 Mar 2004 21:25:43 GMT

[View Forum Message](#) <> [Reply to Message](#)

>>>> > "Michael" == Michael Wallace <mwallace.removethis@swri.edu.invalid> writes:

> This isn't an IDL question exactly, but I figured that
> there are several here who have to deal with these file
> types. Is there any real difference between CDF and HDF?
> I have looked at some web pages that describe the
> differences, but it doesn't mean much to me since I've
> never delved into either one. The only reason I'm even
> looking at these file types is that these seem to be the

> types preferred by the data center we will be working with
> later on for long-term archival of our data. If it means
> anything, my core data sets are nothing more than really
> big arrays. Later on, other data like PNG images may be
> included.

> Unless there's some big difference between the two, I'm
> going to go with the one that's easier to work with in IDL.
> So, that brings me to the question, is one easier to work
> with than the other? Or do I flip a coin to determine
> which one to use?

CDF is pretty much superseded by netCDF, as far as I know, so you might consider netCDF instead. In that case, the biggest difference that I know of is that HDF 5 is capable of handling arbitrary sized data sets. NetCDF cannot handle files larger than 2 Gbytes due to an internal 32 bit integer. This has caused me all sorts of headaches. I've had to write various wrappers that split data sets into multiple netCDF files that are smaller than 2 Gbytes each. HDF 5 handles that sort of thing transparently. Not only can it handle arbitrarily large data sets, but it will split large data sets across multiple files so that OS limitations on individual file size are not surpassed.

That said, we've got such a large base of netCDF-based codes, that we have not yet bitten the bullet and switched to HDF. If we were to start a new, independent project, we'd certainly choose HDF 5 instead.

For a summary and comparison to other formats, see http://hdf.ncsa.uiuc.edu/HDF5/RD100-2002/All_About_HDF5.pdf . That paper compares HDF 5 to the HDF 4, netCDF, PDB, TIFF, FITS and OpenDX formats.

Mike

--

Michael A. Miller mmiller3@iupui.edu
Imaging Sciences, Department of Radiology, IU School of Medicine

Subject: Re: CDF vs HDF

Posted by [K. Bowman](#) on Tue, 30 Mar 2004 22:53:25 GMT

[View Forum Message](#) <> [Reply to Message](#)

In article <87ptaum3hk.fsf@lumen.indyrad.iupui.edu>,
mmiller3@iupui.edu (Michael A. Miller) wrote:

- > NetCDF cannot handle files larger
- > than 2 Gbytes due to an internal 32 bit integer.

This is not strictly true.

From the UNIDATA netCDF FAQ:

- > Is it possible to create netCDF files larger than 2 Gbytes?
- >
- > It is possible to write netCDF files that exceed 2 Gbytes on platforms that
- > have "Large File Support" (LFS). Such files would be platform-independent to
- > other LFS platforms, but if you open such a file on an older platform
- > without LFS, you would expect a "file too large" error.
- >
- > There are significant restrictions on the structure of large netCDF files
- > that result from the 32-bit relative offsets that are part of the netCDF
- > file format. If you don't use the unlimited dimension, only one variable can
- > exceed 2 Gbytes in size, but it can be as large as the underlying file
- > system permits. It must be the last variable in the dataset, and the offset
- > to the beginning of this variable must be less than about 2 Gbytes. If you
- > use the unlimited dimension, any number of record variables may exceed 2
- > Gbytes in size, as long as the offset of the start of each record variable
- > within a record is less than about 2 Gbytes. For examples of both these
- > forms of large netCDF files, see the Large File Support section in the
- > User's Guide.
- >
- > To enable LFS for writing large netCDF files requires that the libraries be
- > built with specific combinations of platform-specific compile flags on some
- > systems. For examples, see the Installation instructions.

OSes with LFS include IRIX, Solaris, AIX, ...

I have not tried this from IDL. We have a mixture of 32- and 64-bit systems, and I need files to be portable.

Ken Bowman

Subject: Re: CDF vs HDF

Posted by [mmiller3](#) on Wed, 31 Mar 2004 14:30:31 GMT

[View Forum Message](#) <> [Reply to Message](#)

>>>> > "Kenneth" == Kenneth Bowman <k-bowman@null.tamu.edu> writes:

- > In article <87ptaum3hk.fsf@lumen.indyrad.iupui.edu>,
- > mmiller3@iupui.edu (Michael A. Miller) wrote:

>> NetCDF cannot handle files larger than 2 Gbytes due to an
>> internal 32 bit integer.

> This is not strictly true.

> From the UNIDATA netCDF FAQ:

[...]

Thanks for pointing that out, Ken!

Mike

Subject: Re: CDF vs HDF

Posted by [K. Bowman](#) on Wed, 31 Mar 2004 16:20:52 GMT

[View Forum Message](#) <> [Reply to Message](#)

In article <87oeqdk1k.fsf@lumen.indyrad.iupui.edu>,
mmiller3@iupui.edu (Michael A. Miller) wrote:

>>> NetCDF cannot handle files larger than 2 Gbytes due to an
>>> internal 32 bit integer.

>

>> This is not strictly true.

>

> Thanks for pointing that out, Ken!

>

> Mike

If you try this out from IDL, let me know whether it works!

Ken

Subject: Re: CDF vs HDF

Posted by [david.b.han](#) on Thu, 01 Apr 2004 12:56:11 GMT

[View Forum Message](#) <> [Reply to Message](#)

mmiller3@iupui.edu (Michael A. Miller) wrote in message
news:<87ptaum3hk.fsf@lumen.indyrad.iupui.edu>...

>>>> >> "Michael" == Michael Wallace <mwallace.removethis@swri.edu.invalid> writes:

>

>> This isn't an IDL question exactly, but I figured that

>> there are several here who have to deal with these file

>> types. Is there any real difference between CDF and HDF?

>> I have looked at some web pages that describe the

>> differences, but it doesn't mean much to me since I've

>> never delved into either one. The only reason I'm even
>> looking at these file types is that these seem to be the
>> types preferred by the data center we will be working with
>> later on for long-term archival of our data. If it means
>> anything, my core data sets are nothing more than really
>> big arrays. Later on, other data like PNG images may be
>> included.
>
>> Unless there's some big difference between the two, I'm
>> going to go with the one that's easier to work with in IDL.
>> So, that brings me to the question, is one easier to work
>> with than the other? Or do I flip a coin to determine
>> which one to use?
>
> CDF is pretty much superseded by netCDF, as far as I know, so you
> might consider netCDF instead.

This is not true. CDF is a superset of netCDF and it's much easier to use from the programmer's and user's perspective. Both IDL and MatLab has a built-in support (APIs) for CDF, and there's a suite of IDL routines with you can manipulate CDF files (read, write, plots, etc.) at <http://cdaweb.gsfc.nasa.gov/cdaweb/>.

There's also a data format translator at <http://translators.gsfc.nasa.gov> that allow users to convert one or more local and/or remote files between CDF, netCDF, and FITS.

> In that case, the biggest
> difference that I know of is that HDF 5 is capable of handling
> arbitrary sized data sets. NetCDF cannot handle files larger
> than 2 Gbytes due to an internal 32 bit integer. This has caused
> me all sorts of headaches. I've had to write various wrappers
> that split data sets into multiple netCDF files that are smaller
> than 2 Gbytes each. HDF 5 handles that sort of thing
> transparently. Not only can it handle arbitrarily large data
> sets, but it will split large data sets across multiple files so
> that OS limitations on individual file size are not surpassed.
>
> That said, we've got such a large base of netCDF-based codes,
> that we have not yet bitten the bullet and switched to HDF. If
> we were to start a new, independent project, we'd certainly
> choose HDF 5 instead.
>
> For a summary and comparison to other formats, see

> http://hdf.ncsa.uiuc.edu/HDF5/RD100-2002/All_About_HDF5.pdf .
> That paper compares HDF 5 to the HDF 4, netCDF, PDB, TIFF, FITS
> and OpenDX formats.
>
> Mike

Subject: Re: CDF vs HDF

Posted by [david.b.han](#) on Thu, 01 Apr 2004 13:10:10 GMT

[View Forum Message](#) <> [Reply to Message](#)

Michael Wallace <mwallace.removethis@swri.edu.invalid> wrote in message
news:<106jn2q11rcsp51@corp.supernews.com>...

> This isn't an IDL question exactly, but I figured that there are several
> here who have to deal with these file types. Is there any real
> difference between CDF and HDF? I have looked at some web pages that
> describe the differences, but it doesn't mean much to me since I've
> never delved into either one. The only reason I'm even looking at these
> file types is that these seem to be the types preferred by the data
> center we will be working with later on for long-term archival of our
> data. If it means anything, my core data sets are nothing more than
> really big arrays. Later on, other data like PNG images may be included.
>
> Unless there's some big difference between the two, I'm going to go with
> the one that's easier to work with in IDL. So, that brings me to the
> question, is one easier to work with than the other? Or do I flip a
> coin to determine which one to use?
>
> -Mike

Both CDF and netCDF are much easier to learn and use than HDF and CDF
is a superset of netCDF, not the other way around. Both IDL and
MatLab have a built-in support for CDF, and there's a suite of IDL
routines with which you can manipulate CDF files (e.g. read, write,
plot, etc.) at <http://spdf.gsfc.nasa.gov/CDAWlib.html>.

There's a data format translator at <http://translators.gsfc.nasa.gov>
that allows users to translate files between CDF, netCDF, and FITS.

Subject: Re: CDF vs HDF

Posted by [giglio](#) on Thu, 01 Apr 2004 18:58:40 GMT

[View Forum Message](#) <> [Reply to Message](#)

In article <87ptaum3hk.fsf@lumen.indyrad.iupui.edu>,
mmiller3@iupui.edu (Michael A. Miller) writes:

- > For a summary and comparison to other formats, see
- > http://hdf.ncsa.uiuc.edu/HDF5/RD100-2002/All_About_HDF5.pdf .
- > That paper compares HDF 5 to the HDF 4, netCDF, PDB, TIFF, FITS
- > and OpenDX formats.

This summary is a good starting point, but be aware that it contains some errors (e.g., "FITS files are not portable between 32 bit and 64 bit systems."). Part of the problem is that the authors sometimes confuse inherent limitations of the physical file format with limitations of software libraries used to manipulate the files. The authors also lapse into writing hype now and then: "We were the first to implement a scientific data format and library that works in parallel computing environments using MPI I/O." (This being said, I do think that HDF5 is extremely flexible, and the library interface is much more logical than that of, say, HDF4.)

Louis
