Subject: Re: Reading HDF5 Compound Datasets in IDL
Posted by James Kuyper on Thu, 19 May 2005 15:43:55 GMT
View Forum Message <> Reply to Message

Justin Bronn wrote:
> Hello All,
>
> I'm stumped. Here's the situation: I have an HDF5 dataset that I
want
> to read, and I cannot figure out from the IDL documentation how to
> selectively read only parts of the dataset. Read on for the gory
> details...
>
> The HDF5 dataset is relatively simple; there are 4 groups, each
> containing a compound data type (the HDF5 compound data type is
> analagous to the IDL struct). There can be N-elements of this
compound
> data type (again, like an array of IDL structures). The compound
data
> type contains 4 different fields: a filename, a time-stamp, and three
> floating point arrays. Below is a representation of this HDF5 file
> with the equivalent IDL datatypes.
>
> /--+
> |- Group_1
> | |
> | |- compound (Can have N number of elements
> | | | => IDL Structure Array)
> | | |
> | | |--- filename (String, 256 Characters long
> | | | => IDL String)
> | | |--- time (64-bit Float Value
> | | | => IDL Double)
> | | |--- data1 (32-bit Float Array, 4096 Elements
> | | | => IDL Float Array)
> | | |--- data2 (32-bit Float Array, 4096 Elements
> | | | => IDL Float Array)
> | | |--- data3 (32-bit Float Array, 4096 Elements
> | | | => IDL Float Array)
> ...
> |
> |- Group_4 (Same as Group_1)
>
> Now if I want to read an entire one of these compound data types into
> IDL, here's what I can do:
>
> ;; Opening up the necessary HDF5 file IDs
> h5fid = h5f_open('data.h5')

> h5gid = h5g_open(h5fid, 'Group_1')
> h5did = h5d_open(h5gid, 'compound')
>
> ;; Reading the data
> data = h5d_read(h5did)
>
> ;; Cleaning up
> h5d_close, h5did & h5g_close, h5gid & h5f_close, h5fid
>
> When I do a 'help' on the data read in, I get exactly what I
expected:
>
> IDL> help, data
> DATA STRUCT = -> <Anonymous> Array[263]
> IDL> help, data, /ST
> ** Structure <8225794>, 5 tags, length=49172, data length=49172,
> refs=1:
> FILENAME STRING '/path/to/ascii_file'
> TIME DOUBLE 2452305.5
> DATA1 FLOAT Array[4096]
> DATA2 FLOAT Array[4096]
> DATA3 FLOAT Array[4096]
> IDL>
>
> Now here's the problem: I do not want to have to read the ENTIRE
> compound data type. When the number of compound elements gets large
> (say N=3000), the read operation takes a _long_ time since the entire
> compound data type is read into memory. I want to selectively read
only
> _portions_ of the compound data type, like the 'time' element, to
> determine what I only really need, and then read out that selection.

When using the HDF5 library from C, the H5Dread() function takes a memory datatype argument, and automatically converts from the file datatype to the memory datatype. If both datatypes are compound types, and the memory datatype's members are a subset of the members of the file's datatype, then it extracts only those members. I looked over the documentation for IDLs HDF5 library, but I couldn't find anything that seemed to be a wrapper for this functionality.

Warning: for both C and IDL, this is based entirely upon reading documentation; I have no actual experience with HDF5 programming in either language.