Gentle folks,

I'm trying to read a dataset from a h5 file by chunks. Let's say that in the file I have dataset called 'temperature' that contains 3D matrix (nx x ny x nz). Normally I use H5D_READ to read the entire dataset/cube at once. Since the dimensions of the cube may be huge (I easily get out of memory), I wonder is it possible to read h5 datasets chunk by chunk (slice by slice for example)? Something like using ASSOC to read large binary files.

I'm lost in the list of h5-related IDL routines. Any help will be appreciated!

Thanks!

Nikola

Le mardi 7 juin 2016 11:33:10 UTC+2, Nikola Vitas a écrit :
> Gentle folks,
>
> I'm trying to read a dataset from a h5 file by chunks. Let's say that in the file I have dataset called 'temperature' that contains 3D matrix (nx x ny x nz). Normally I use H5D_READ to read the entire dataset/cube at once. Since the dimensions of the cube may be huge (I easily get out of memory), I wonder is it possible to read h5 datasets chunk by chunk (slice by slice for example)? Something like using ASSOC to read large binary files.
>
> I'm lost in the list of h5-related IDL routines. Any help will be appreciated!
>
> Thanks!
>
> Nikola

The recipe with IDL implementation of HDF5 library might be the following:
 - open your file:   fileId = H5F_OPEN(...)
 - open your 3D dataset:   dsId = H5D_OPEN(fileId, ...)
 - get the corresponding dataspace: dId = H5D_GET_SPACE(dsId)
 - define the memory space to hold each readout chunk:
     mId = H5S_CREATE_SIMPLE(dims)
     (dims is the 3-vector containing sizes of the 3D slice):
Inside the reading loop:
 - define an individual chunk:    H5S_SELECT_HYPERSLAB, dId, start, dims, /RESET
     (start is the 3-vector containing position of the 3D slice)

- read the data subset:   data = H5D_READ(dsId, FILE_SPACE=dId, MEMORY_SPACE=mId)
Loop as far as you like.
When finished, close all the opened Ids.
I guess that the reading performance will depend on the way in which the file was originally
written.
Cheers,
alx

---

## Subject: Re: Reading h5 dataset by chunks
Posted by Michael Galloy on Thu, 09 Jun 2016 15:25:40 GMT
View Forum Message <> Reply to Message

On 6/7/16 6:53 am, alx wrote:
> Le mardi 7 juin 2016 11:33:10 UTC+2, Nikola Vitas a écrit :
>> Gentle folks,
>>
>> I'm trying to read a dataset from a h5 file by chunks. Let's say that in the file I have dataset
called 'temperature' that contains 3D matrix (nx x ny x nz). Normally I use H5D_READ to read the
entire dataset/cube at once. Since the dimensions of the cube may be huge (I easily get out of
memory), I wonder is it possible to read h5 datasets chunk by chunk (slice by slice for example)?
Something like using ASSOC to read large binary files.
>>
>> I'm lost in the list of h5-related IDL routines. Any help will be appreciated!
>>
>> Thanks!
>>
>> Nikola
>
> The recipe with IDL implementation of HDF5 library might be the following:
>  - open your file:   fileId = H5F_OPEN(...)
>  - open your 3D dataset:   dsId = H5D_OPEN(fileId, ...)
>  - get the corresponding dataspace: dId = H5D_GET_SPACE(dsId)
>  - define the memory space to hold each readout chunk:
>      mId = H5S_CREATE_SIMPLE(dims)
>      (dims is the 3-vector containing sizes of the 3D slice):
> Inside the reading loop:
>  - define an individual chunk:    H5S_SELECT_HYPERSLAB, dId, start, dims, /RESET
>      (start is the 3-vector containing position of the 3D slice)
>  - read the data subset:   data = H5D_READ(dsId, FILE_SPACE=dId, MEMORY_SPACE=mId)
> Loop as far as you like.
> When finished, close all the opened Ids.
> I guess that the reading performance will depend on the way in which the file was originally
written.
> Cheers,
> alx
>

Yes, I believe those are all the steps/routines you need. Check out
MG_H5_GETDATA for an example of doing this (or just use it, if that
suits your purposes):

https://github.com/mgalloy/mglib/blob/master/src/hdf5/mg_h5_ getdata.pro

Mike
--
Michael Galloy
www.michaelgalloy.com
Modern IDL: A Guide to IDL Programming (http://modernidl.idldev.com)

---

Subject: Re: Reading h5 dataset by chunks
Posted by Nikola on Fri, 17 Feb 2017 09:38:56 GMT
View Forum Message <> Reply to Message

Hi Mike,

Sorry that I haven't give you any feedback earlier. I got distracted with another project at the time
and only now I have tried your routine. It's excellent and it does exactly what I need. Many thanks!

Best, Nikola