Subject: Appending data to a dataset in an HDF5 file
Posted by Maarten[1] on Wed, 31 Jan 2007 16:40:21 GMT
View Forum Message <> Reply to Message

Hi,

The title says it all, really: I'm trying to create a dataset with an unlimited dimension, so that I can append data later, preferably without reading the previous data. Right now I'm not even getting to that last refinement, just reading the data, then extending the array and finally writing the extended array back to the file fails.

I create the dataspace with

dataspace_id = H5S_CREATE_SIMPLE([10,20,13], max_dim=[-1,20,13])

and I think this creates an unlimited first dimension (if I understand the docs, which may be a tad optimistic). I can write a single data set, but when writing the second (appended) set I'd expect that writing something with dimensions [12,20,13] would be possible, but I get an error:

% H5D_WRITE: Number of elements in the IDL variable must match that of the HDF5 dataspace/datatype

What am I doing wrong here? Do I need to set the dimension itself to [-1,20,13]? The docs don't seem to suggest that.

Help!

Maarten

Subject: Re: Appending data to a dataset in an HDF5 file
Posted by eddie haskell on Thu, 01 Feb 2007 16:51:47 GMT
View Forum Message <> Reply to Message

> The title says it all, really: I'm trying to create a dataset with an
> unlimited dimension, so that I can append data later, preferably
> without reading the previous data.

There are several possible issues as to why it is failing for you:

When you create your dataset you must set the chunk_dimensions keyword to be able to extend the size of a dataset.

You use h5d_extend to actually extend the dataset

If you want to write new data in the extended space without touching the original data you need to use a hyperslab

Example:

```
;; create file
f = h5f_create('test.h5')
;; create datatype
dt = h5t_idl_create(1b)
;; create dataspace with unlimited size
ds = h5s_create_simple(10, max_dim=[-1])
;; create dataset: must set chunk dimensions
d = h5d_create(f, 'test', dt, ds, chunk=10)
;; write original data
h5d_write, d, bindgen(10)

;; extend dataset by two elements
h5d_extend, d, 12
;; get updated matching dataspace
ds2 = h5d_get_space(d)
;; mark dataspace to write to the last two elements only
h5s_select_hyperslab, ds2, 10, 2, /reset
;; create data to write
data = bytarr(12)
;; put new data in the last two elements
data[10:11] = [100b, 200b]
;; write only the new elements to the extended dataset
h5d_write, d, data, file_space=ds2, memory_space=ds2

;; verify new dataset
print, h5d_read(d)

;; close everything
h5f_close, f
```

Hope this helps.
Cheers,
eddie

Subject: Re: Appending data to a dataset in an HDF5 file
Posted by Maarten[1] on Fri, 02 Feb 2007 08:50:32 GMT
View Forum Message <> Reply to Message

On Feb 1, 4:51 pm, eddie haskell <hask...@nospam.edu> wrote:
>> The title says it all, really: I'm trying to create a dataset with an
>> unlimited dimension, so that I can append data later, preferably
>> without reading the previous data.

>
> There are several possible issues as to why it is failing for you:
>
> When you create your dataset you must set the chunk_dimensions keyword
> to be able to extend the size of a dataset.
>
> You use h5d_extend to actually extend the dataset
>
> If you want to write new data in the extended space without touching the
> original data you need to use a hyperslab

Thanks, the example is really useful (you'd hope that something like
this appears
in the manual at some time in the future). For now I accumulate the
data in memory,
and dump at the end of the run. Since it seems that you need a full
array, even when
appending a few bytes, I don't think this is really practical, or good
on the memory use.
I guess it is possible to call the write routine with a different
memory space, but that
part of the documentation is quite bad.

thanks anyway.

Maarten

Subject: Re: Appending data to a dataset in an HDF5 file
Posted by eddie haskell on Fri, 02 Feb 2007 15:54:14 GMT
View Forum Message <> Reply to Message

>> Since it seems that you need a full array, even when
>> appending a few bytes, ...

Not true. In your original code example you were trying to write a full
sized larger array into the dataset, and when I wrote my example I had
mixed the two options in my mind and ended up using a full dataset array
but only writing two elements. My bad.

By defining a new dataspace you can write only the data to any portion
of the dataset you need without wasting memory.

```
;; extend dataset by two elements
h5d_extend, d, 12
;; get updated matching dataspace
ds2 = h5d_get_space(d)
;; mark only needing to write to the last two elements
```

h5s_select_hyperslab, ds2, 10, 2, /reset

;; create data to write
data = [100b, 200b]
;; create dataspace for writing
ds3 = h5s_create_simple(2)
;; write only the new elements to the dataset
h5d_write, d, data, file_space=ds2, memory_space=ds3

Sorry for the confusion.

Cheers,
eddie

Subject: Re: Appending data to a dataset in an HDF5 file
Posted by Maarten[1] on Mon, 05 Feb 2007 10:35:48 GMT
View Forum Message <> Reply to Message

On Feb 2, 3:54 pm, eddie haskell <hask...@nospam.edu> wrote:

> By defining a new dataspace you can write only the data to any portion
> of the dataset you need without wasting memory.
>
> ;; extend dataset by two elements
> h5d_extend, d, 12
> ;; get updated matching dataspace
> ds2 = h5d_get_space(d)
> ;; mark only needing to write to the last two elements
> h5s_select_hyperslab, ds2, 10, 2, /reset
>
> ;; create data to write
> data = [100b, 200b]
> ;; create dataspace for writing
> ds3 = h5s_create_simple(2)
> ;; write only the new elements to the dataset
> h5d_write, d, data, file_space=ds2, memory_space=ds3
>
> Sorry for the confusion.

Nothing to be sorry about, this post is already much easier to
understand than the IDL documentation. When I find the time to add
this to my simple HDF5 writer, I'll include this, and perhaps prepare
a simple article for the Coyote, for others to enjoy.

Thanks,

Maarten

Subject: Re: Appending data to a dataset in an HDF5 file
Posted by David Fanning on Mon, 05 Feb 2007 13:22:09 GMT
View Forum Message <> Reply to Message

Maarten writes:

> When I find the time to add
> this to my simple HDF5 writer, I'll include this, and perhaps prepare
> a simple article for the Coyote, for others to enjoy.

Si, por favor!

Salud,

David
--
David Fanning, Ph.D.
Fanning Software Consulting, Inc.
Coyote's Guide to IDL Programming: http://www.dfanning.com/