
Subject: Re: What's better: 1 big HDF file or several samller ones??

Posted by [Robert Stockwell](#) on Wed, 25 Sep 2002 12:23:53 GMT

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

Brian Huether wrote:

- > I have radar data that is broken down based on target orientation (i.e. the
- > data is split into 72 5 degree azimuthal windows). So do I create one HDF
- > file with 72 datasets? The other thing to consider is this: I need to run
- > computational algorithms that will need to access the data so are there
- > speed considerations when saving the data this way? Basically I have to data
- > storage goals: 1) make the data readily shareable, 2)make the data storage
- > appropriate for quick retrieval.
- >
- > I suppose when it comes to the computational stuff, I can use the one big
- > file and then one time I can just read all the info into an array. So the
- > awkwardness of the big file will only be problematic one time.
- >
- > Any thoughts?
- >
- > -brian
- >
- >

Hi Brian,

I tend to vote with "one big file". It is easier to keep track of, to backup (assuming it fits on your media), and to transfer. With many smaller files, it is not always easy to tell that one of them "disappear".

In fact I did have that problem with my satellite data. It had 13,000 files in a directory, and retrieving the directory listing resulted in some 100 random files getting dropped (on a slightly older version of linux, 7.0 maybe). It is very difficult to find an error like that, which would not happen if the data was in giant files.

I store global data maps (x,y,t) in large (1 gig) yearly files as contiguous spatial maps one on top of each other, but I also reproduce the data as a collection of contiguous time series, so depending on what type of data you want (a snapshot of the data at a particular time, or all the data from one site as a time series) the retrieval is blazingly fast.

Cheers,
bob

Subject: Re: What's better: 1 big HDF file or several samller ones??

Posted by [James Kuyper](#) on Wed, 25 Sep 2002 14:06:29 GMT

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

Brian Huether wrote:

- >
- > I have radar data that is broken down based on target orientation (i.e. the
- > data is split into 72 5 degree azimuthal windows). So do I create one HDF
- > file with 72 datasets? The other thing to consider is this: I need to run
- > computational algorithms that will need to access the data so are there
- > speed considerations when saving the data this way? Basically I have to data
- > storage goals: 1) make the data readily shareable, 2) make the data storage
- > appropriate for quick retrieval.
- >
- > I suppose when it comes to the computational stuff, I can use the one big
- > file and then one time I can just read all the info into an array. So the
- > awkwardness of the big file will only be problematic one time.

The basic issues you have to consider are file size limits, and how you intend to use the data. On many systems there's an upper limit on the size of files, either set by the available disk space, by a file addressing limit (typically 2GB). You have to break up your file if that's the case.

How often will the code that reads this (these) file(s) need to access data across multiple azimuth bins? If never, then you should split each azimuth bin into a separate file. If frequently, they should be in the same file if possible, and you should consider the possibility of merging them into a single SDS with one dimension for the azimuth bins. If infrequently, it's a judgement call.
