
Subject: Re: reading/writing large files

Posted by [Russell Ryan](#) on Fri, 01 Feb 2013 16:49:05 GMT

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

On Friday, February 1, 2013 11:36:22 AM UTC-5, Craig Markwardt wrote:

> On Friday, February 1, 2013 10:15:25 AM UTC-5, rr...@stsci.edu wrote:

>

>> Okay gang I've been working on this for a few days and have given up.

>

>>

>

>>

>

>>

>

>> I've got this simulation that outputs an array of floating point numbers (roughly 5000 or so), which I want to put into a file. If the file exists, I want to append to it; if not, I want to create it. I want to do this of order a million times (at least append of order a million times). When the simulation finishes, I want to read these numbers and do some post-processing. I don't want to read the entire file at once because I'm afraid I'll run into memory problems (especially since I can envision doing the appending 10^7 or even 10^8 times). So, instead I'd like to read say all 10^6 (or 10^7 or 10^8) trials of the k-th element of the array and get a single floating-point array of 10^6 elements (or what have you). Basically, I'm envisioning a table with say 5000ish columns but the number of rows is variable, and I want to read the k-th column.

>

>>

>

>

>

> Suddenly all these questions about dealing with "large" files...

>

>

>

> To me, a table with 10^6 , 10^7 or 10^8 rows doesn't seem that large (although 5000 columns is pretty big).

>

>

>

> If you are dealing in astronomy, consider using FITS files. There are lots of tools to deal with FITS files inside or outside of IDL.

>

>

>

> Personally, I would have each simulation write one small table and then at the end, use a merging program such as 'ftmerge' (for FITS) to combine the small tables into one master table. This also allows you to have multiple simulations running on different processors, without fear of stomping on each other.

>

>
>
> Craig

Hi Gang,

Thanks for the ideas. Yeah, I'm familiar with the HDF files and think I'm gonna look at the CDF files. Craig, Yeah I am an astronomer and have been using fits files. It would be awesome if I could use a binary table, but it seems that I can only have 999 columns (for fx*pro). I was looking at the ft*pro library from Landsman, but it's not clear to me that this will work for me either. DO you have any other advice on the fits I/O libraries?

R
