Subject: Re: reading/writing large files
Posted by Russell[1] on Fri, 01 Feb 2013 20:23:41 GMT
View Forum Message <> Reply to Message

On Friday, February 1, 2013 1:44:46 PM UTC-5, Craig Markwardt wrote:
> On Friday, February 1, 2013 11:49:05 AM UTC-5, rr...@stsci.edu wrote:
>
>> On Friday, February 1, 2013 11:36:22 AM UTC-5, Craig Markwardt wrote:
>
>> 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?
>
>
>
> Do you really have 5000 distinct name-worthy values? I bet you have vectors of data. FITS binary tables can easily accomodate vectors (or even arrays) in a single table cell. We routinely put whole images into a FITS binary table cell.
>
>
>
> For real programming I usually use the fits_bintable library (FXB*.PRO), which allows one to be very explicit about the table structure. For quick-n-dirty you can use MRDFITS and MWRFITS. (for MWRFITS you need to be very careful about making sure your data has the same data structure and data type, so that you can merge it later without hassle.)
>
>
>
> Craig

Hiya Craig,
Unfortunately I do have ~5000 unique variables. More concretely, I have an MCMC simulation which has ~5000 dimensions --- and I want ~10^6-8 random deviates from the posterior. But, only ~20 of these dimensions are truly interesting (yet don't want to throw away all that extra data).

It sounds like you're suggesting I write a FITS binary table, but each entry in the table is an array. This is possible, and I'm considering; especially since there is a natural way of doing this in my code. But it will have a minor draw back in recovering the file. In this scheme, each array (per cell) would be like a 1000 element array and if I have say 10^6-8 draws from it, I'm looking at reading 10^9-11 floating point numbers. This is becoming a problem, though I'm not sure if I'll ever *NEED* to do this. I just wanted to keep my options open.

Thank you all so much!
R