Subject: Re: Structure of arrays or arrays of structures?
Posted by Craig Markwardt on Fri, 23 Jan 2009 00:15:06 GMT
View Forum Message <> Reply to Message

On Jan 22, 3:52 pm, Brian Larsen <balar...@gmail.com> wrote:
> All,
>
> this may be philosophy but I have fought with both today and I am
> wondering if there are pros and cons to the different
> implementations.
>
> My example is that when I read in text files I build a structure of
> arrays:
> IDL> help, data
> DATA STRUCT = -> <Anonymous> Array[1]
> IDL> help, data, /str
> ** Structure <1a07a08>, 5 tags, length=212480, data length=212480,
> refs=1:
> JD DOUBLE Array[5312]
> PAS0 DOUBLE Array[5312]
> PAS90 DOUBLE Array[5312]
> MEP0E3 DOUBLE Array[5312]
> MEP90E3 DOUBLE Array[5312]
>
> and when I read cdf data (at least from ACE) I get arrays of
> structures
> IDL> help, data
> DATA STRUCT = -> <Anonymous> Array[36451]
> IDL> help, data, /str
> ** Structure <19cf008>, 45 tags, length=200, data length=196, refs=1:
> DNUM DOUBLE 0.0000000
> YEAR LONG 2007
> DAY LONG 165
> HR LONG 0
> MIN LONG 0
> SEC FLOAT 8.98560
>
> Are there memory issues with one way or the other? Other things I
> haven't thought about?
>
> I prefer the feel of the structure of arrays since I like typing
> tmp = data.jd[0:10]
> more than I like typing
> tmp = data[0:10].jd

My thoughts:

Both methods will be slower than pure arrays.

I believe the array-of-structures approach will be the slowest, because every time you need to fetch a field (as your second example), IDL has to assemble the array from structure parts. That's been my experience anyway.

However, the array-of-structures approach is natural if you can think of your data as a database, and you can add or remove more entries at a later time. You can do this easily by filtering the array of structures and/or appending, without destroying the existing entries. For a structure of arrays you basically need to tear apart the structure and re-build it if you want to change any of the contents.

The structure-of-arrays approach is natural when you will be passing it as an _EXTRA argument, since for that you need all of your data bundled into a single structure.

Craig
