Subject: Re: Speeding up multiple file reading
Posted by Maarten[1] on Thu, 02 Feb 2006 15:48:00 GMT

clivecook59@gmail.com wrote:

> I have a program where i need to read in multiple files. Currently i
> read in 6000 binary files using a function i have written. This reads
> three columns of data out of each file. To do this i use a loop that
> calls the function to read a file who's data is then added to an array.
> At the moment it takes around 90 seconds to go through the 6000 files.
> Is there any way that i can read this data not using a loop? Or at
> least are there any tips for speeding this up?

Is this data static (a look-up table, for instance?) In that case you
might want to read the data once, and write it back out to a format
that can more easily hold multiple arrays (say, hdf or netcdf). I think
you'll find that hdf is much faster, especially since you don't have to
append to an array, but create it once at the right size.

That said, if you can avoid resizing an array, and just create one
array of the final size and use that instead, I think you'll see some
improvement.

How large are these files? 90 seconds for 6000 files doesn't sound too
shabby, unless each file holds only 3 points. Part of the problem may
be due to overhead of the filesystem itself, which goes avay by using a
single file instead.

Maarten