Subject: Fast I/O of large files
Posted by eharold on Sat, 28 May 1994 23:57:42 GMT
View Forum Message <> Reply to Message

Here's the problem:

I have astronomically large FITS data files representing the data
taken from a 340 by 256 CCD at 200 time steps and 14 wavelengths.
A day's worth of data is stored in 200 different data files, each of which
represents one time step at 340 by 256 by 14 points. They're arranged such
that with ASSOC I see 14 arrays of 340 by 256 integers each.

I need to perform fourier analysis on this
data at each point in x-y-z space through time. That is what I really
need is data files for each of the 320 by 240 by 14 points that contains the
appropriate (integer) number at all 200 time steps; i.e. I'd like my
files separated by spatial position. Instead they're separated by time.

To do the fourier analysis I therefore need to open 200 files
to read just one number out of each and form an array that I can then analyze
in time. Once I've finished one point I do it again, i.e. open the same 200
files to read the next time series. I've tried variations of this scheme.
It does help if I read out all the z values at a step. However the total
size of the files is on the order of 400 MB so I can't read all of them in
at once. Some of them have to stay on disk.

Almost all the time in my code is being taken up in I/O.
Commenting out all the actual analysis yields negligible speed gains.
Thus I really need to find the fastest means of doing this I/O I can.
Here's what the actual I/O looks like now:

```
 for step=0,timesteps-1 do begin
  ; Associate arrays with all the files at each time step
  filename = "feb15bs5434.f"+strtrim(string(step),2)
  openr, u, filename
  temp = assoc(u, intarr(xdim,ydim), offset)

  ; fill the velocity vectors for this time step
  ; WARNING: This next line is a real time killer and needs to be fixed
  ; See 17-38 of the User's Guide
  for k=0,num_vels-1 do vels(step,k) = temp(i,j,k)
  close,u
 endfor  ; time
```

This collects the needed data at one point in x-y space over 200
time steps and 14 heights. After this I perform the needed analysis,
write my output and loop to the next x-y position.
Can anyone see a better way of doing this?

As is I can't seem to get the code to run in less than a year's time
for a single day of data. Obviously this needs to be improved.
Any help or suggestions you could provide would be greatly appreciated.

--

Elliotte Rusty Harold National Solar Observatory
eharold@sunspot.noao.edu Sunspot NM 88349
