Subject: Re: Reading in text data
Posted by Craig Markwardt on Tue, 08 Aug 2000 07:00:00 GMT
View Forum Message <> Reply to Message

reardonb@my-deja.com writes:

> Hi. I am reading in text data (columns and rows of numbers) and I would
> like to know if there is a more elegant way of doing it. Currently, the
> user must specify how many columns there are. In my case the number of
> columns is manually inserted into the first line of the file like this:
>
> 3
> 0 1 2
> 1 2 3
> 2 3 4
> 3 4 5
> 4 5 6
> 5 6 7
> 6 7 8
> 7 8 9
> 8 9 10
> 9 10 11

You've already had some pretty good responses. You're really asking two questions: (1) What if I don't know how many columns there are? and (2) What if I don't know how many rows there are?

Question 1: how many columns? Answer: count them! If you can read the first line, then with judicious application of STRTRIM and STRCOMPRESS you can do this quite readily:

```
str = '' & readf, unit, str              ;; Read string
str = byte(strcompress(strtrim(str,2)))   ;; Remove spaces, convert to bytes
wh = where(str EQ 32B, n_columns)         ;; Count number of remaining spaces
n_columns = n_columns + 1
```

Then you will have to rewind the file pointer to actually read the data.

Question 2: how many rows? Answer: either count them, or use a dynamic resizing technique.

You've seen some suggestions already for counting rows, which are good. The "wc" trick works only on Unix.

The dynamic resizing technique is to grow your array as needed. I have found that growing the array with each line is too slow and memory-wasting. What I normally do is grow the size of the array by a

factor of two, up to a certain limit, beyond which the arrays grows
linearly. This has the benefit that you do a minimum number of growth
operations for small-to-mid sized arrays.

To use your terminology, it would be something like this:

```
max_rows = 0L
counter = 0L
while NOT EOF(lun) do begin
  .... read data ....
  if count GE max_rows then begin
    if max_rows EQ 0 then max_rows = 128L     ;; Minimum array size
    max_rows = max_rows + (max_rows < 4096L)  ;; Maximum growth is 4k
    newdata = make_array(n_columns, max_rows, value=tp)  ;; Make new array

    if n_elements(data) GT 0 then $
      newdata(0,0) = data                     ;; Copy old data into newdata
    data = 0 & data = temporary(newdata)      ;; Now "data" contains new data
  endif

  data(*,counter) = temporary_data            ;; Insert one row
  counter = counter + 1
endwhile
data = data(*,0:counter-1)                    ;; Trim the array
```

Meditate on that for awhile. :-)

Good luck,
Craig

--
-- --------------
Craig B. Markwardt, Ph.D.   EMAIL: craigmnet@cow.physics.wisc.edu
Astrophysics, IDL, Finance, Derivatives | Remove "net" for better response
-- --------------