
Subject: Cut down a big file into several files based on its a column.

Posted by [kim20026](#) on Wed, 14 Mar 2007 01:22:27 GMT

[View Forum Message](#) <> [Reply to Message](#)

Good day, everyone! (I wrote 'Good morning', but maybe it is not morning for others. anyway) I would like to slice a big data file into several files based on its a column, but it was not successful yet. Please give me some suggestions.

The contents of my original data file are...

```
90|2000|1|1|95|95|95|95|95|96|95|95|94|93|93|93|94|94|94|95| 95|93|91|
90|94|95|96|95|
90|2000|1|2|96|95|94|96|93|93|76|74|76|81|85|84|76|53|43|40| 39|41|33|
33|32|32|33|35|
90|2000|1|3|35|34|38|35|29|28|30|29|26|23|25|22|22|30|29|24| 23|24|36|
31|34|39|31|34|
.
.
.
```

This is a 2-dim array, I named this array as 'data(*, *)'.

What I want to do now is to slice this file into 6 files based on the information in 2nd column, (data(1,*)). This column includes year info, and the values are from 2000 to 2006.

I tried this way so far.

```
pro hum_year
; 1. Read original file, and designate the initial values of
variables and arrays.
```

```
file='hum.txt'
ndata=file_lines(file)
data=intarr(28, ndata)
```

```
; 2. Close any unit files before processing.
```

```
close,1
close,2
close,3
close,4
close,5
close,6
```

```
; 3. Open files and prepare for writing.
```

```
openw,1,'HumNWS_2000.txt'
openw,2,'HumNWS_2001.txt'
```

```
openw,3,'HumNWS_2002.txt'
openw,4,'HumNWS_2003.txt'
openw,5,'HumNWS_2004.txt'
openw,6,'HumNWS_2005.txt'
```

; 4. Classify data and write them into designated file.

```
for t=0L,ndata-1 do begin
```

```
  case data[1,t] of
    2000:printf,1,format='(28i6)',data[*],t]
    2001:printf,2,format='(28i6)',data[*],t]
    2002:printf,3,format='(28i6)',data[*],t]
    2003:printf,4,format='(28i6)',data[*],t]
    2004:printf,5,format='(28i6)',data[*],t]
    2005:printf,6,format='(28i6)',data[*],t]
  endcase
```

```
endfor
```

; 5. Close files.

```
close,1
close,2
close,3
close,4
close,5
close,6
```

```
end
```

However, compile stops process 2 everytime. I don't understand. Do I have do something to treat the '|'s in original file? Please give me any suggestions and comments. Thank you in advance.

Subject: Re: Cut down a big file into several files based on its a column.

Posted by [Peter Clinch](#) on Fri, 16 Mar 2007 10:24:39 GMT

[View Forum Message](#) <> [Reply to Message](#)

DirtyHarry wrote:

> Thanks, Maarten... but... T.T

>

> Is there any way to do this in Windows version of IDL?

Perl is usually better for this sort of string handling IME, and Perl is freely available for Windows and is a very handy thing to have to pre-process data files.

It'll probably be quicker to hack on in IDL for this immediate problem, but I'd look into Perl for future jobs, and get the O'Reilly Learning Perl "Llama book" by Schwarz and Phoenix to go with it.

Pete.

--

Peter Clinch Medical Physics IT Officer
Tel 44 1382 660111 ext. 33637 Univ. of Dundee, Ninewells Hospital
Fax 44 1382 640177 Dundee DD1 9SY Scotland UK
net p.j.clinch@dundee.ac.uk http://www.dundee.ac.uk/~pjclinch/

Subject: Re: Cut down a big file into several files based on its a column.

Posted by [Maarten\[1\]](#) on Fri, 16 Mar 2007 15:34:54 GMT

[View Forum Message](#) <> [Reply to Message](#)

On Mar 16, 9:46 am, "DirtyHarry" <kim20...@gmail.com> wrote:

> Thanks, Maarten... but... T.T

>

> Is there any way to do this in Windows version of IDL?

This was my first try. It doesn't quite work, because I can't get the format statment right (how do you insert linebreaks?)

;;; failed attempt.

file='hum.txt'

ndata=file_lines(file)

data=intarr(28, ndata)

for year=2000,2005 do begin

 openw, unit, string(format="('HumNWS_',I04,'.txt')", year), /get_lun

 idx = where(data[1,*] eq year, cnt)

 if cnt gt 0 then begin

 fmt=string(format="(' ',I04, '(28i6)')", cnt)

 ; it would ne neat if IDL actually allowed to add newlines between
the blocks here.

 printf, unit, format=fmt, data[*,idx]

 endif

 free_lun, unit

endfor

end

;;; Second attempt: use explicit for loop.

file='hum.txt'

ndata=file_lines(file)

data=intarr(28, ndata)

```
for year=2000,2005 do begin
  openw, unit, string(format="('HumNWS_',I04,'.txt')", year), /get_lun
```

```
  idx = where(data[1,*] eq year, cnt)
  fmt='(28i6)'
  if cnt gt 0 then $
    for ii=0,cnt-1 do $
      printf, unit, format=fmt, data[:,idx[ii]]
```

```
  free_lun, unit
endfor
end
```

HTMH,

Maarten

Subject: Re: Cut down a big file into several files based on its a column.
Posted by [Maarten\[1\]](#) on Fri, 16 Mar 2007 15:54:35 GMT
[View Forum Message](#) <> [Reply to Message](#)

On Mar 14, 1:22 am, "DirtyHarry" <kim20...@gmail.com> wrote:

> However, compile stops process 2 everytime. I don't understand. Do I
> have do something to treat the '|'s in original file? Please give me
> any suggestions and comments. Thank you in advance.

This threw me off a bit: you're not reading your data at all.

Now, the delimiter is a but screwed up. Read the help on read_ascii
and ascii_template. You'll end up with your data in a structure, and
then you can use where as you see fit.

Maarten
