
Subject: READ, adn get data into an array from LARGE SIZE FILES

Posted by [lucsmm](#) on Tue, 07 Jul 2015 16:43:47 GMT

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

Hello All

I have a big problem

I need to open ,read and extract some useful data from big files, I have about 20 files ranging from 189MB to 22GB in size.

There is a header (first 4 lines or so)

Not all the lines are the same size, want to re-structure the file into an array with only a few data point from each line.

Like from the first data line I just want the 1000

From the following line I want to 212

From the following line I want the 0.80000E+01

So I will write

1000 212 0.80000E+01

Then I want

3000 122 0.80000E+01

3000 211 0.75687E+01

3000 115 0.75687E+01

SKIP 5000 *****

2015 155 0.17684E+01

SKIP 5000 *****

2011 115 0.51101E+00

Or something like that

This is an example kind of format the data files are written :

```
3 7 9 8 9 8 9 8 9 8 9 0 4 0 0 0 0 0 0 0
1 2 3 7 8 9 16 17 18 19 20 21 22 23 24 25 26 27 28 7 8 10 11 16 17 18 19
20 21 22
23 24 25 26 27 28 7 8 12 13 16 17 18 19 20 21 22 23 24 25 26 27 28 7 8 10
11 16 17 18
19 20 21 22 23 24 25 26 27 28 7 8 14 15 16 17 18 19 20 21 22 23 24 25 26 27
28
2475 1000 2
3000 1 40 3 212 0 0
```

0.40643E+01 -0.93584E+01 0.48473E+01 0.20720E-01 0.90154E+00 -0.43220E+00
 0.80000E+01 0.10000E+01 0.00000E+00
 3000 2 43 25 3 122 6 0
 0.42219E+01 -0.25000E+01 0.15593E+01 0.20720E-01 0.90154E+00 -0.43220E+00
 0.80000E+01 0.10000E+01 0.25422E-01
 3000 3 44 31 3 211 0 5
 0.42171E+01 -0.24650E+01 0.15412E+01 -0.83941E-01 0.85475E+00 -0.51220E+00
 0.75687E+01 0.10000E+01 0.25555E-01
 5000 4 117 174 3 115 5 5
 0.40564E+01 -0.82822E+00 0.56041E+00 -0.83941E-01 0.85475E+00 -0.51220E+00
 0.75687E+01 0.10000E+01 0.31955E-01
 2015 4 2 1 3 115 5 932
 0.41191E+01 -0.75830E+00 0.55086E+00 0.90458E+00 0.38728E+00 -0.17820E+00
 0.99789E-03 0.10000E+01 0.32428E-01
 5000 7 0 0 3 115 5 1
 0.43406E+01 -0.74618E+00 0.41210E+00 0.67909E+00 -0.11221E+00 -0.72543E+00
 0.17684E+01 0.10000E+01 0.33286E-01
 2011 7 2 3 3 115 5 1343
 0.43580E+01 -0.75818E+00 0.39485E+00 0.93607E+00 0.33833E+00 -0.96485E-01
 0.99819E-03 0.10000E+01 0.33643E-01
 5000 10 78000 3 3 115 5 1
 0.43578E+01 -0.72648E+00 0.41716E+00 0.38784E+00 0.47850E+00 0.78779E+00
 0.51101E+00 0.10000E+01 0.33772E-01
 2013 10 2 5 3 115 5 1049
 0.43576E+01 -0.72315E+00 0.41485E+00 0.46633E+00 0.88357E+00 -0.43014E-01
 0.98065E-03 0.10000E+01 0.33864E-01
 5000 6 78000 5 3 115 5 1
 0.43406E+01 -0.74618E+00 0.41210E+00 0.90620E+00 0.13522E+00 -0.40064E+00
 0.15462E+00 0.10000E+01 0.33286E-01
 2011 6 2 7 3 115 5 781
 0.43403E+01 -0.74631E+00 0.41167E+00 -0.52533E+00 0.65367E+00 0.54475E+00
 0.99241E-03 0.10000E+01 0.33303E-01
 5000 8 78000 3 3 115 5 1
 0.43402E+01 -0.74649E+00 0.41166E+00 -0.23393E+00 -0.18276E+00 0.95492E+00
 0.58210E-02 0.10000E+01 0.33293E-01
 2011 8 2 9 3 115 5 361
 0.43402E+01 -0.74649E+00 0.41167E+00 -0.16978E+00 0.60617E+00 0.77700E+00
 0.96494E-03 0.10000E+01 0.33293E-01
 5000 7 78000 3 3 115 5 2
 0.41530E+01 -0.76214E+00 0.54112E+00 0.15077E+00 -0.36478E+00 -0.91881E+00
 0.67048E-01 0.10000E+01 0.32533E-01
 2014 7 2 11 3 115 5 897
 0.41530E+01 -0.76206E+00 0.54111E+00 -0.75633E+00 -0.29606E+00 0.58337E+00
 0.99668E-03 0.10000E+01 0.32544E-01
 5000 7 78000 4 3 115 5 1
 0.41502E+01 -0.76024E+00 0.54159E+00 0.83551E+00 -0.10533E+00 0.53929E+00
 0.11567E-01 0.10000E+01 0.32522E-01
 2011 7 2 12 3 115 5 492

0.41502E+01 -0.76025E+00 0.54159E+00 -0.80438E+00 -0.52532E+00 -0.27751E+00
0.98903E-03 0.10000E+01 0.32523E-01

Any idea how to handle this? Please Help?

Subject: Re: READ, adn get data into an array from LARGE SIZE FILES

Posted by [Jeremy Bailin](#) on Wed, 08 Jul 2015 02:12:14 GMT

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

On Tuesday, July 7, 2015 at 12:43:49 PM UTC-4, luc...@gmail.com wrote:

> Hello All
> I have a big problem
> I need to open ,read and extract some useful data from big files, I have about 20 files ranging
from 189MB to 22GB in size.
> There is a header (first 4 lines or so)
>
> Not all the lines are the same size, want to re-structure the file into an array with only a few data
point from each line.
> Like from the first data line I just want the 1000
> From the following line I want to 212
> From the following line I want the 0.80000E+01
>
> So I will write
> 1000 212 0.80000E+01
>
> Then I want
> 3000 122 0.80000E+01
>
> 3000 211 0.75687E+01
>
> 3000 115 0.75687E+01
>
> SKIP 5000 *****
>
> 2015 155 0.17684E+01
>
> SKIP 5000 *****
>
> 2011 115 0.51101E+00
>
>
> Or something like that
>
> This is an example kind of format the data files are written :
>
> 3 7 9 8 9 8 9 8 9 8 9 0 4 0 0 0 0 0 0
> 1 2 3 7 8 9 16 17 18 19 20 21 22 23 24 25 26 27 28 7 8 10 11 16 17 18

```

19 20 21 22
> 23 24 25 26 27 28 7 8 12 13 16 17 18 19 20 21 22 23 24 25 26 27 28 7 8
10 11 16 17 18
> 19 20 21 22 23 24 25 26 27 28 7 8 14 15 16 17 18 19 20 21 22 23 24 25 26
27 28
> 2475 1000 2
> 3000 1 40 3 212 0 0
> 0.40643E+01 -0.93584E+01 0.48473E+01 0.20720E-01 0.90154E+00 -0.43220E+00
0.80000E+01 0.10000E+01 0.00000E+00
> 3000 2 43 25 3 122 6 0
> 0.42219E+01 -0.25000E+01 0.15593E+01 0.20720E-01 0.90154E+00 -0.43220E+00
0.80000E+01 0.10000E+01 0.25422E-01
> 3000 3 44 31 3 211 0 5
> 0.42171E+01 -0.24650E+01 0.15412E+01 -0.83941E-01 0.85475E+00 -0.51220E+00
0.75687E+01 0.10000E+01 0.25555E-01
> 5000 4 117 174 3 115 5 5
> 0.40564E+01 -0.82822E+00 0.56041E+00 -0.83941E-01 0.85475E+00 -0.51220E+00
0.75687E+01 0.10000E+01 0.31955E-01
> 2015 4 2 1 3 115 5 932
> 0.41191E+01 -0.75830E+00 0.55086E+00 0.90458E+00 0.38728E+00 -0.17820E+00
0.99789E-03 0.10000E+01 0.32428E-01
> 5000 7 0 0 3 115 5 1
> 0.43406E+01 -0.74618E+00 0.41210E+00 0.67909E+00 -0.11221E+00 -0.72543E+00
0.17684E+01 0.10000E+01 0.33286E-01
> 2011 7 2 3 3 115 5 1343
> 0.43580E+01 -0.75818E+00 0.39485E+00 0.93607E+00 0.33833E+00 -0.96485E-01
0.99819E-03 0.10000E+01 0.33643E-01
> 5000 10 78000 3 3 115 5 1
> 0.43578E+01 -0.72648E+00 0.41716E+00 0.38784E+00 0.47850E+00 0.78779E+00
0.51101E+00 0.10000E+01 0.33772E-01
> 2013 10 2 5 3 115 5 1049
> 0.43576E+01 -0.72315E+00 0.41485E+00 0.46633E+00 0.88357E+00 -0.43014E-01
0.98065E-03 0.10000E+01 0.33864E-01
> 5000 6 78000 5 3 115 5 1
> 0.43406E+01 -0.74618E+00 0.41210E+00 0.90620E+00 0.13522E+00 -0.40064E+00
0.15462E+00 0.10000E+01 0.33286E-01
> 2011 6 2 7 3 115 5 781
> 0.43403E+01 -0.74631E+00 0.41167E+00 -0.52533E+00 0.65367E+00 0.54475E+00
0.99241E-03 0.10000E+01 0.33303E-01
> 5000 8 78000 3 3 115 5 1
> 0.43402E+01 -0.74649E+00 0.41166E+00 -0.23393E+00 -0.18276E+00 0.95492E+00
0.58210E-02 0.10000E+01 0.33293E-01
> 2011 8 2 9 3 115 5 361
> 0.43402E+01 -0.74649E+00 0.41167E+00 -0.16978E+00 0.60617E+00 0.77700E+00
0.96494E-03 0.10000E+01 0.33293E-01
> 5000 7 78000 3 3 115 5 2
> 0.41530E+01 -0.76214E+00 0.54112E+00 0.15077E+00 -0.36478E+00 -0.91881E+00
0.67048E-01 0.10000E+01 0.32533E-01

```

```

> 2014 7 2 11 3 115 5 897
> 0.41530E+01 -0.76206E+00 0.54111E+00 -0.75633E+00 -0.29606E+00 0.58337E+00
0.99668E-03 0.10000E+01 0.32544E-01
> 5000 7 78000 4 3 115 5 1
> 0.41502E+01 -0.76024E+00 0.54159E+00 0.83551E+00 -0.10533E+00 0.53929E+00
0.11567E-01 0.10000E+01 0.32522E-01
> 2011 7 2 12 3 115 5 492
> 0.41502E+01 -0.76025E+00 0.54159E+00 -0.80438E+00 -0.52532E+00 -0.27751E+00
0.98903E-03 0.10000E+01 0.32523E-01
>
> Any idea how to handle this? Please Help?

```

Write it in C.

Seriously, processing a text file like this that is 22 GB in IDL is not worth it.

-Jeremy.

Subject: Re: READ, adn get data into an array from LARGE SIZE FILES
 Posted by [Craig Markwardt](#) on Wed, 08 Jul 2015 18:12:37 GMT
[View Forum Message](#) <> [Reply to Message](#)

On Tuesday, July 7, 2015 at 12:43:49 PM UTC-4, luc...@gmail.com wrote:

```

> Hello All
> I have a big problem
> I need to open ,read and extract some useful data from big files, I have about 20 files ranging
from 189MB to 22GB in size.
> There is a header (first 4 lines or so )
>
> Not all the lines are the same size, want to re-structure the file into an array with only a few data
point from each line.
> Like from the first data line I just want the 1000
> From the following line I want to 212
> From the following line I want the 0.80000E+01

```

This is a parsing problem.

You will need to come up with a READF statement that can read each different kind of line from the file. I can't tell from your example, but if there are 3 different kinds of line, then you need three different READF statements. You will need to learn about the FORMAT keyword.

Then you will need to put those statements into a loop. If you know you will always get one line of format type 1, one line of format type 2, and 5000 lines of type 3, then you will need to put two READF statements according to those first types followed by a loop that reads the third type 5000 types. And so on. This is how we parse files.

CM
