Subject: Re: Very slow IDL vs Matlab (ascii file reading)
Posted by R.Bauer on Tue, 04 Nov 2003 22:13:07 GMT
View Forum Message <> Reply to Message

Marcin Jakubowski wrote:

> Hi all,
> I've tried to read an ASCII file, which is composed of one very long row
> (660*496) double precision numbers, each of them delimited with
> tabulator. In Matlab 6.5 I am using small program
>
> ==========
> fid = fopen('Mydata.dat');
> data = fscanf(fid,'%g',[660,496]);
> fclose(fid)
> ==========
>
> and it takes about one second to read the file. If I try to do similar
> in IDL 6.0
>
> ==========
> data = FltArr(660, 496)
> OpenR, lun, 'Mydata.dat', /Get_Lun
> ReadF, lun, data
> Free_Lun, lun
> ==========
>
> then it takes about 20 minutes (!!!) to read the same file. What causes
> the problem? Unfortunately I need to use the IDL as it is a part of huge
> code written in IDL. Is it any chance to shorten that time?
>
>
> Many thanks in advance,
> Marcin
>
> P.s. I've performed checks on PC and Linux machines and the outcomes are
> similar.

Dear Marcin,

two things
1) you are speaking from double precision but you have defined float only
2) there must be a local problem on your machines.
How much memory is free after the first reading.
You could use help,/memory to get this information. The thing you
described usally happens if swapping is necessary or someone else uses
100 % CPU time. I tried today during the lessons a similiar example
without a problem.

regards

Reimar

--
Forschungszentrum Juelich
email: R.Bauer@fz-juelich.de
http://www.fz-juelich.de/icg/icg-i/
 ========== ======
a IDL library at ForschungsZentrum Juelich
 http://www.fz-juelich.de/icg/icg-i/idl_icglib/idl_lib_intro. html

---

Subject: Re: Very slow IDL vs Matlab (ascii file reading)
Posted by R.G. Stockwell on Tue, 04 Nov 2003 22:56:24 GMT
View Forum Message <> Reply to Message

"Marcin Jakubowski" <ma.jakubowski@fz-juelich.de> wrote in message
news:bo93p8$1a19g3$1@ID-112500.news.uni-berlin.de...
> Hi all,
> I've tried to read an ASCII file, which is composed of one very long row
> (660*496) double precision numbers,and it takes about one second to read
the file. If I try to do similar
> in IDL 6.0
>
> ==========
> data = FltArr(660, 496)
> OpenR, lun, 'Mydata.dat', /Get_Lun
> ReadF, lun, data
> Free_Lun, lun
> ==========
>
> then it takes about 20 minutes (!!!) to read the same file.

I suggest directly typing in the 660 by 496 array, that might be faster than
the 20 minutes.

The above took 1.8 seconds (and that is with casting the doubles into the
float array).
Perhaps your problem is elsewhere?

cheers,
bob

Subject: Re: Very slow IDL vs Matlab (ascii file reading)
Posted by R.Bauer on Wed, 05 Nov 2003 07:11:58 GMT
View Forum Message <> Reply to Message

Marcin Jakubowski wrote:

> Hi all,
> I've tried to read an ASCII file, which is composed of one very long row
> (660*496) double precision numbers, each of them delimited with
> tabulator. In Matlab 6.5 I am using small program
>
> ==========
> fid = fopen('Mydata.dat');
> data = fscanf(fid,'%g',[660,496]);
> fclose(fid)
> ==========
>
> and it takes about one second to read the file. If I try to do similar
> in IDL 6.0
>
> ==========
> data = FltArr(660, 496)
> OpenR, lun, 'Mydata.dat', /Get_Lun
> ReadF, lun, data
> Free_Lun, lun
> ==========
>
> then it takes about 20 minutes (!!!) to read the same file. What causes
> the problem? Unfortunately I need to use the IDL as it is a part of huge
> code written in IDL. Is it any chance to shorten that time?
>
>
> Many thanks in advance,
> Marcin
>
> P.s. I've performed checks on PC and Linux machines and the outcomes are
> similar.

My previous examples does not have all data in one line!!!
(It's always easier to check if a complete example is provided to us)

Here is an example to create the data file.

```
x=make_array(/double,value=999D,660L*496L)
w=17l*660L*496L
openw,lun,/get_l,'Mydata.dat',width=w
printf,lun,x
```

free_lun,lun

By reading this data I got the same problem as described above.
This could be a bug in width. Any ideas?

I would suggest to use a shell command to split the lines after 1000 numbers
in new lines. On linux this could be done by sed.

Reimar

--
Forschungszentrum Juelich
email: R.Bauer@fz-juelich.de
http://www.fz-juelich.de/icg/icg-i/
 ========== ======
a IDL library at ForschungsZentrum Juelich
 http://www.fz-juelich.de/icg/icg-i/idl_icglib/idl_lib_intro. html

---

Subject: Re: Very slow IDL vs Matlab (ascii file reading)
Posted by Marcin Jakubowski on Wed, 05 Nov 2003 09:15:38 GMT
View Forum Message <> Reply to Message

Reimar Bauer wrote:

>
>
> My previous examples does not have all data in one line!!!
> (It's always easier to check if a complete example is provided to us)
>
> Here is an example to create the data file.
>
> x=make_array(/double,value=999D,660L*496L)
> w=17l*660L*496L
> openw,lun,/get_l,'Mydata.dat',width=w
> printf,lun,x
> free_lun,lun
>
> By reading this data I got the same problem as described above.
> This could be a bug in width. Any ideas?
>
> I would suggest to use a shell command to split the lines after 1000 numbers
> in new lines. On linux this could be done by sed.
>
> Reimar
>

Thanks for help, I'll try to wrap the file.

Regards,
Marcin

Subject: Re: Very slow IDL vs Matlab (ascii file reading)
Posted by Richard French on Wed, 05 Nov 2003 12:13:15 GMT
View Forum Message <> Reply to Message

I can reproduce this problem on MacOS with IDL6.0 - but it does not seem like a problem with 'width' to me, since that is related to writing the file - the problem is with reading a large single-line file in free format. It amazes me that it takes many minutes to read this file, and I think it is a definite bug. I'd suggest that you report this to RSI.

Dick French

Subject: Re: Very slow IDL vs Matlab (ascii file reading)
Posted by Richard French on Wed, 05 Nov 2003 12:22:50 GMT
View Forum Message <> Reply to Message

On 11/5/03 7:13 AM, in article BBCE5389.FEA%rfrench@wellesley.edu, "Richard G. French" <rfrench@wellesley.edu> wrote:

> I can reproduce this problem on MacOS with IDL6.0 - but it does not seem
> like a problem with 'width' to me, since that is related to writing the file
> - the problem is with reading a large single-line file in free format. It
> amazes me that it takes many minutes to read this file, and I think it is a
> definite bug. I'd suggest that you report this to RSI.
>
> Dick French
>

I tried to do an end-run by reading the file in as a string, and then using READS to extract the required information from the string:

```
Openr,lun,/GET_LUN,'MyData.dat'
S=''
Readf,lun,s
Free_lun,lun
This gets executed almost instantaneously.
Help,s
S               STRING    = '       999.00000       999.00000       999'...
Print,strlen(s)
5237760
```

```
Data=fltarr(660,496)
Reads,s,data
```

This takes forever, too. So, we've shown that the problem is not an I/O problem, but one related to the actual parsing of the long string. What the heck is happening during all of those CPU cycles? READS needs some serious attention!

Dick French

Subject: Re: Very slow IDL vs Matlab (ascii file reading)
Posted by Richard French on Wed, 05 Nov 2003 12:38:20 GMT
View Forum Message <> Reply to Message

OK, I figured out a work-around using STRSPLIT:

```
; write the data file

x=make_array(/double,value=999.d0,660L*496L)
w=17L*660L*496L
openw,lun,/get_lun,'Mydata.dat',width=w
PRINTF,lun,x
free_lun,lun
; read the file as a single string
s=''
openr,lun,'Mydata.dat',/get_lun
readf,lun,s
IDL> help,s
S               STRING    = '       999.00000       999.00000
999.000'...
free_lun,lun
;convert the long string to desired floating point array
Data=fltarr(660,496)
Reads,strsplit(s,/EXTRACT),data
```

This is fast, and should do what you need without having to split the file itself in to shorter lines. You can do it all in IDL.

Dick French

Subject: Re: Very slow IDL vs Matlab (ascii file reading)
Posted by R.Bauer on Wed, 05 Nov 2003 16:54:33 GMT
View Forum Message <> Reply to Message

Good work !

Reimar

Richard G. French wrote:
> OK, I figured out a work-around using STRSPLIT:
>
> ; write the data file
>
> x=make_array(/double,value=999.d0,660L*496L)
> w=17L*660L*496L
> openw,lun,/get_lun,'Mydata.dat',width=w
> PRINTF,lun,x
> free_lun,lun
> ; read the file as a single string
> s=''
> openr,lun,'Mydata.dat',/get_lun
> readf,lun,s
> IDL> help,s
> S STRING = ' 999.00000 999.00000
> 999.000'...
> free_lun,lun
> ;convert the long string to desired floating point array
> Data=fltarr(660,496)
> Reads,strsplit(s,/EXTRACT),data
>
> This is fast, and should do what you need without having to split the file
> itself in to shorter lines. You can do it all in IDL.
>
>
> Dick French
>

--
Reimar Bauer

Institut fuer Stratosphaerische Chemie (ICG-I)
Forschungszentrum Juelich
email: R.Bauer@fz-juelich.de
---------- -------
a IDL library at ForschungsZentrum Juelich
http://www.fz-juelich.de/icg/icg-i/idl_icglib/idl_lib_intro. html
========== =======

Subject: Re: Very slow IDL vs Matlab (ascii file reading)
Posted by Marcin Jakubowski on Thu, 06 Nov 2003 08:47:52 GMT
View Forum Message <> Reply to Message

Richard G. French wrote:
> OK, I figured out a work-around using STRSPLIT:
> Dick French

It's a very good work-around!

Thank you very much!!!

Marcin

Subject: Re: Very slow IDL vs Matlab (ascii file reading)
Posted by R.Bauer on Thu, 06 Nov 2003 10:08:35 GMT
View Forum Message <> Reply to Message

Marcin Jakubowski wrote:
> Richard G. French wrote:
>
>> OK, I figured out a work-around using STRSPLIT:
>> Dick French
>
>
> It's a very good work-around!
>
> Thank you very much!!!
>
> Marcin
>

Did you have sent a bug report to rsinc ?

Reimar

--
Reimar Bauer

Institut fuer Stratosphaerische Chemie (ICG-I)
Forschungszentrum Juelich
email: R.Bauer@fz-juelich.de
---------- -------
a IDL library at ForschungsZentrum Juelich
http://www.fz-juelich.de/icg/icg-i/idl_icglib/idl_lib_intro. html

========== =======

Subject: Re: Very slow IDL vs Matlab (ascii file reading)
Posted by Marcin Jakubowski on Thu, 06 Nov 2003 10:53:33 GMT
View Forum Message <> Reply to Message

Reimar Bauer wrote:
> Marcin Jakubowski wrote:
> Did you have sent a bug report to rsinc ?
> Reimar
>

Yes, I did.

Marcin