
Subject: Re: How to read first and last line of an ascii file FAST

Posted by [Heinz Stege](#) on Sun, 29 Aug 2010 16:32:35 GMT

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

On Sun, 29 Aug 2010 08:32:15 -0700 (PDT), ahs wrote:

> Hello group,
>
> My data is stored in a bunch of ascii files, each 5-10MB big. Each
> line represent a sample with a timestamp. I would like to analyse all
> my files by reading the timestamp from the first and last line of
> every file. Getting the timestamp from the last line is the issue
> here. Right now I'm doing it like
>
> readf, fid, firstline
> nlines = FILE_LINES(fnames[i])
> skip_lun, fid, nlines-2, /LINES
> readf, fid, lastline
>
> skip_lun is really slow when the files are big and numerous. Does
> anyone have a tip of I could do this faster? Is there a way to
> directly point to the end of the file minus one line?
>
> regards,
> Agnar

Hi Agnar,

I would expect, that the file_lines function is taking some time too.
You can try the following:

Get the file size in bytes by the fstat function. Then use point_lun
to set the file pointer to n bytes before the file end. n should be
greater than the length of the last line. Read the n bytes into a byte
array and search for the last occurrence of CR and/or LF in that array.
If you don't find a CR/LF increase n and repeat the steps from
point_lun to here. Finally convert the bytes after CR/LF to string
type.

HTH, Heinz

Subject: Re: How to read first and last line of an ascii file FAST

Posted by [Jeremy Bailin](#) on Mon, 30 Aug 2010 01:27:15 GMT

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

On Aug 29, 11:32 am, ahs <agna...@gmail.com> wrote:

> Hello group,

>
> My data is stored in a bunch of ascii files, each 5-10MB big. Each
> line represent a sample with a timestamp. I would like to analyse all
> my files by reading the timestamp from the first and last line of
> every file. Getting the timestamp from the last line is the issue
> here. Right now I'm doing it like
>
> readf, fid, firstline
> nlines = FILE_LINES(fnames[i])
> skip_lun, fid, nlines-2, /LINES
> readf, fid, lastline
>
> skip_lun is really slow when the files are big and numerous. Does
> anyone have a tip of I could do this faster? Is there a way to
> directly point to the end of the file minus one line?
>
> regards,
> Agnar

What OS? On unix and Mac, I'd just spawn "tail -1 filename". No idea if there's a Windows equivalent.

-Jeremy.

Subject: Re: How to read first and last line of an ascii file FAST

Posted by [R.Bauer](#) on Tue, 31 Aug 2010 16:08:14 GMT

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

Am 29.08.2010 18:32, schrieb Heinz Stege:

> On Sun, 29 Aug 2010 08:32:15 -0700 (PDT), ahs wrote:
>
>> Hello group,
>>
>> My data is stored in a bunch of ascii files, each 5-10MB big. Each
>> line represent a sample with a timestamp. I would like to analyse all
>> my files by reading the timestamp from the first and last line of
>> every file. Getting the timestamp from the last line is the issue
>> here. Right now I'm doing it like
>>
>> readf, fid, firstline
>> nlines = FILE_LINES(fnames[i])
>> skip_lun, fid, nlines-2, /LINES
>> readf, fid, lastline
>>

did you tried reading the file in one string array of len file_lines

```
IDL> f = "example.txt"
IDL> n = file_lines(f)
IDL> data = make_array(n,/string)
IDL> openr, lun, /get_lun, f
IDL> readf, lun, data
IDL> free_lun, lun
IDL> print, data[n-1]
```

that needs 0.05 seconds to get that information

Reimar

```
>> skip_lun is really slow when the files are big and numerous. Does
>> anyone have a tip of I could do this faster? Is there a way to
>> directly point to the end of the file minus one line?
>>
>> regards,
>> Agnar
```

Subject: Re: How to read first and last line of an ascii file FAST
Posted by [agnarhs](#) on Fri, 03 Sep 2010 09:56:10 GMT
[View Forum Message](#) <> [Reply to Message](#)

```
On 29 Aug, 18:32, Heinz Stege <public.215....@arcor.de> wrote:
> On Sun, 29 Aug 2010 08:32:15 -0700 (PDT), ahs wrote:
>> Hello group,
>
>> My data is stored in a bunch of ascii files, each 5-10MB big. Each
>> line represent a sample with a timestamp. I would like to analyse all
>> my files by reading the timestamp from the first and last line of
>> every file. Getting the timestamp from the last line is the issue
>> here. Right now I'm doing it like
>
>> readf, fid, firstline
>> nlines = FILE_LINES(fnames[i])
>> skip_lun, fid, nlines-2, /LINES
>> readf, fid, lastline
>
>> skip_lun is really slow when the files are big and numerous. Does
>> anyone have a tip of I could do this faster? Is there a way to
>> directly point to the end of the file minus one line?
>
>> regards,
>> Agnar
>
```

> HiAgnar,
>
> I would expect, that the file_lines function is taking some time too.
> You can try the following:
>
> Get the file size in bytes by the fstat function. Then use point_lun
> to set the file pointer to n bytes before the file end. n should be
> greater than the length of the last line. Read the n bytes into a byte
> array and search for the last occurrence of CR and/or LF in that array.
> If you don't find a CR/LF increase n and repeat the steps from
> point_lun to here. Finally convert the bytes after CR/LF to string
> type.
>
> HTH, Heinz

Hi Heinz,

That was a good idea. All the lines in the ascii files are equal with
LINE_B bytes, and doing the following solved my problem.

```
A = fstat(fid)
readf, fid, firstline
point_lun, fid, A.size-LINE_B
readf, fid, lastline
```

Agnar
