Hello,

I have learned a couple of items worth sharing regarding our continuing
transition from UNIX/WIN to MAC.
Specifically, we have been wrestling with ASCII text data files
generated from a CTD device embedded in a Windows OS environ.   Of
course the issue that surfaces is the carriage return/line feed (CRLF)
that is used in one envronment and not the others (and I can't remember
which does what... it doesn't matter really). Our files have an
complicated (messy) header followed by columnar data. Our interest is
in the header.

My hope is that by sharing what we've learned, someone will point us to
an even better solution.

We have come up with three solutions...

(1) Load the Win/DOS text format files onto a UNIX machine and use the
DOS2UNIX command:

unix> dos2unix -ascii infilename outfilename

This format is very clean for the MAC, too.

(2) Load the Win/DOS text format file into a Mac editior, like BBEdit
(www.barebones.com) and save in a Mac format.

This method is easy, but we have a bazillion of these files and I'm a
pretty poor show at piece work (too much daydreaming!)

(3) Use IDL to read the file and handle the extra control characters
internally.
We are using this method now because the enduser doesn't want to mess
around with exchanging files, etc.   Since the header is relatively
small, there is little performance loss.   In a nutshell we introduced a
test
for the contents of the most recent line read. If the line is emtpy
(meaning there was an extra linefeed or some-such-thing) then read the
next line. Here's a snippet...

   .
   .

.

```
ReadF,U, Input, Format='(A)'
If N_elements(Input) EQ 0 Then ReadF,U, Input, Format='(A)'    ;check
for blanks
```

.

.

.

So, it works, even if a bit brutish.   The most important thing, from
our standing, is that the enduser doesn't have to care a hoot about the
format of the file.

Ben

--
Ben Tupper
Bigelow Laboratory for Ocean Science
West Boothbay Harbor, Maine
btupper@bigelow.org
    note: email address new as of 25JULY2000

Subject: Re: More WIN/UNIX -> MAC transitions
Posted by promashkin on Tue, 22 Aug 2000 07:00:00 GMT
View Forum Message <> Reply to Message

I think HP UX 9000 is about 2.5 yr old. I am not the only user; it might
be part of the reason its slower. Mac is a newer PowerMac G4-400.
Cheers,
Pavel

Andy Loughe wrote:
>
> Care to share any specs (even ages) of the machines... that sort of
> thing is usually helpful when making such performance claims!
>

Subject: Re: More WIN/UNIX -> MAC transitions
Posted by Andy Loughe on Tue, 22 Aug 2000 07:00:00 GMT
View Forum Message <> Reply to Message

Pavel Romashkin wrote:
>
> Along the same subject line, I ran some data processing code yesterday

---

> on HP Unix machine. It took 3.9 hours to do the same thing that my
> desktop Mac finished in 72 min. Impressive. The upside of the Unix is
> that I can let it run as long as it wants while I am doing more fun
> things in my Mac IDL version :-)
> Cheers,
> Pavel

Care to share any specs (even ages) of the machines... that sort of
thing is usually helpful when making such performance claims!

--
Andrew Loughe ==
NOAA/OAR/FSL/AD   R/FS5 | email: loughe@fsl.noaa.gov
325 Broadway | wwweb: www-ad.fsl.noaa.gov/users/loughe
Boulder, CO 80305-3328 | phone: 303-497-6211   fax: 303-497-6301

Subject: Re: More WIN/UNIX -> MAC transitions
Posted by promashkin on Tue, 22 Aug 2000 07:00:00 GMT
View Forum Message <> Reply to Message

Along the same subject line, I ran some data processing code yesterday
on HP Unix machine. It took 3.9 hours to do the same thing that my
desktop Mac finished in 72 min. Impressive. The upside of the Unix is
that I can let it run as long as it wants while I am doing more fun
things in my Mac IDL version :-)
Cheers,
Pavel

Subject: Re: More WIN/UNIX -> MAC transitions
Posted by Martin Schultz on Wed, 23 Aug 2000 07:00:00 GMT
View Forum Message <> Reply to Message

Hi Ben,

   I am not in a mood to try (and I don't have a Mac anyway) but
there may be a way to write a generic ASCII file reader. The main
trick is to read binary into a BYTE array and test for CR (13B)
and LF (10B) characters, then reform them as you go along.
Something like:

```
;; find CR and LF characters
wcr = Where(barray EQ 13B, crcnt)
wlf = Where(barray EQ 10B, lfcnt)
;; convert all CRs to LFs unless you have both (DOS)
```

```
IF crcnt EQ lfcnt THEN BEGIN
   ;; complement of wcr (will be easier in IDL 5.4 ;-)
   wncr = Where(barray NE 13B, cnt)
   IF cnt GT 0 THEN barray = barray
ENDIF ELSE IF crcnt GT 0 THEN BEGIN
   barray = 10B
ENDELSE

;; convert byte array to string
;; (i leave this up to you as part of the EPA exam ;-)
```

Now, of course, you can get more and more sophisticated and
implement buffered reading etc. or make it even an ASCII file
converter (easy, once you know that all line ends are marked with
LF only).

Here is a quick summary of what I learned about the line-end
characters on the various OS's:
Unix :   LF only (10B)
DOS/Win: CR/LF   (13B, 10B)
Mac: CR only (13B)

Cheers,
Martin

Ben Tupper wrote:
>
> Hello,
>
> I have learned a couple of items worth sharing regarding our continuing
> transition from UNIX/WIN to MAC.
> Specifically, we have been wrestling with ASCII text data files
> generated from a CTD device embedded in a Windows OS environ.   Of
> course the issue that surfaces is the carriage return/line feed (CRLF)
> that is used in one envronment and not the others (and I can't remember
> which does what... it doesn't matter really). Our files have an
> complicated (messy) header followed by columnar data. Our interest is
> in the header.
>
> My hope is that by sharing what we've learned, someone will point us to
> an even better solution.
>
> We have come up with three solutions...
>
> (1) Load the Win/DOS text format files onto a UNIX machine and use the
> DOS2UNIX command:
>

```
> unix> dos2unix -ascii infilename outfilename
>
> This format is very clean for the MAC, too.
>
> (2) Load the Win/DOS text format file into a Mac editior, like BBEdit
> (www.barebones.com) and save in a Mac format.
>
> This method is easy, but we have a bazillion of these files and I'm a
> pretty poor show at piece work (too much daydreaming!)
>
> (3) Use IDL to read the file and handle the extra control characters
> internally.
> We are using this method now because the enduser doesn't want to mess
> around with exchanging files, etc.   Since the header is relatively
> small, there is little performance loss.   In a nutshell we introduced a
> test
> for the contents of the most recent line read.  If the line is emtpy
> (meaning there was an extra linefeed or some-such-thing) then read the
> next line.  Here's a snippet...
>
>    .
>    .
>    .
>
> ReadF,U, Input, Format='(A)'
> If N_elements(Input) EQ 0 Then ReadF,U, Input, Format='(A)'    ;check
> for blanks
>    .
>    .
>    .
>
> So, it works, even if a bit brutish.   The most important thing, from
> our standing, is that the enduser doesn't have to care a hoot about the
> format of the file.
>
> Ben
>
> --
> Ben Tupper
> Bigelow Laboratory for Ocean Science
> West Boothbay Harbor, Maine
> btupper@bigelow.org
>     note: email address new as of 25JULY2000


--
 [[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[ [[[[[[[
[[ Dr. Martin Schultz   Max-Planck-Institut fuer Meteorologie
[[
```

```
[[              Bundesstr. 55, 20146 Hamburg
[[
[[              phone: +49 40 41173-308
[[
[[              fax:   +49 40 41173-298
[[
[[ martin.schultz@dkrz.de
[[
 [[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[[ [[[[[[[
```

Subject: Re: More WIN/UNIX -> MAC transitions
Posted by Mark Hadfield on Wed, 23 Aug 2000 21:13:12 GMT
View Forum Message <> Reply to Message

If I understand you correctly the problem is to read Windows-format text
files into IDL on another platform (Macintosh).

It is claimed in the "What's New" that IDL 5.4 on any platform will be able
to read text files using any of the three (Unix,/Windows/Mac) line
termination conventions.

Mark Hadfield
m.hadfield@niwa.cri.nz http://katipo.niwa.cri.nz/~hadfield/
National Institute for Water and Atmospheric Research
PO Box 14-901, Wellington, New Zealand

Subject: Re: More WIN/UNIX -> MAC transitions
Posted by Ben Tupper on Thu, 24 Aug 2000 07:00:00 GMT
View Forum Message <> Reply to Message

Hello,

I wonder if that capability is via READ_ASCII() or using the more general READF
procedure?   I haven't explored using READ_ASCII() since I have needed to
extract tidbits out of the header for this particular data set.   Of course,
today, I'm on a Wintel machine.   It makes me wish my fingers were platform
independent.

Ben

Mark Hadfield wrote:

> If I understand you correctly the problem is to read Windows-format text
> files into IDL on another platform (Macintosh).

>
> It is claimed in the "What's New" that IDL 5.4 on any platform will be able
> to read text files using any of the three (Unix,/Windows/Mac) line
> termination conventions.
>
> ---
> Mark Hadfield
> m.hadfield@niwa.cri.nz http://katipo.niwa.cri.nz/~hadfield/
> National Institute for Water and Atmospheric Research
> PO Box 14-901, Wellington, New Zealand

--
Ben Tupper
248 Lower Round Pond Road
POB 106
Bristol, ME 04539

Tel: (207) 563-1048
Email: PemaquidRiver@tidewater.net

Subject: Re: More WIN/UNIX -> MAC transitions
Posted by Mark Hadfield on Fri, 25 Aug 2000 07:00:00 GMT
View Forum Message <> Reply to Message

"Ben Tupper" <pemaquidriver@tidewater.net> wrote in message
news:39A511E9.20F7C838@tidewater.net...

(referring to the capability to read text files platform-independently on
IDL 5.4)

> I wonder if that capability> is via READ_ASCII() or using the more general
READF
> procedure?

READF.

> I haven't explored using READ_ASCII() since I have needed to
> extract tidbits out of the header for this particular data set.

They claim READ_ASCII is much faster in 5.4 too. I can't confirm that
because I seldom use it. Perhaps I should...

P.S. I hope I am not violating the 5.4 beta confidentiality agreement with
any of this. I got a "What's New in 5.4" sheet with my maintenance renewal
in the mail the other day so I presume new features in 5.4 are more-or-less
public.

Mark Hadfield
m.hadfield@niwa.cri.nz http://katipo.niwa.cri.nz/~hadfield/
National Institute for Water and Atmospheric Research
PO Box 14-901, Wellington, New Zealand