Subject: Re: convert very large string to numeric
Posted by mvukovic on Tue, 26 Aug 2003 16:42:07 GMT
View Forum Message <> Reply to Message

Paul van Delst <paul.vandelst@noaa.gov> wrote in message
news:<3F4A7ADE.AF8396AD@noaa.gov>...
> Mirko Vukovic wrote:
>>
>> Hello,
>>
>> I have a large two column matrix stored as a string,
>
> Forgive my denseness, but what do you mean exactly when you say you "have a large two
> column matrix stored as a string"? By stored do you mean on disk as an ASCII file, or in a
> variable as an actual character variable?
>
> If the latter, my next question is: how did it get that way? (It's not a facetious
> question...I'm fishing for more details)
>
> paulv

Hmmm. It seems that my exposition was lacking in crucial details.

The data is comming from an E&M simulation program (Maxwell 2D,
student version). The really gory details are as follows:

- From Maxwell I generate the text file with the data.
- With an editor, and insert some XML tags. The file now has a
snippet that looks as follows, and whose contents I need to get into
IDL

<Data-Set>
239843420958.0 23049823048.023984032
3240.83240 0239483.2094
20348.3204 20394803.24
.
.
.
39458.7435 348324.497324
</Data-Set>

- I use IDL's XML reader (properly customized via inheritance) to read
the data.
- Now, inside this reader, the data is in a very large character
string (character buffer). The string contains the verbatim contents
of that particular part of the file. Thus it includes line-feeds,
carriage returns, spaces, tabs, numerals, everything:

239843420958.0 23049823048.023984032
3240.83240 0239483.2094
20348.3204 20394803.24
.
.
.
39458.7435 348324.497324

I have to convert this very long string to a 2*N matrix.

If you look at my original post, the way I do it is first ``flatten
the string'' by removing all line-feeds and carriage returns, and
replacing them with spaces (I do this by converting it to BYTE, doing
a WHERE and replacing. Now my string corresponds to a very loong line
of text. Before, it had line breaks.

239843420958.0 23049823048.023984032 3240.83240 0239483.2094
20348.3204 20394803.24 . . . 39458.7435 348324.497324

At this point, I need to pluck out individual groups of numbers (which
are separated by spaces), and convert them to floats or doubles, and
store them into a vector. I use PARSELINE.

Finally I REFORM the vector to desired dimensions. And that part
takes some time that I was hoping to shorten.

How much time? Oh, 3-5 sec per data set. So far, since yesterday I
have spent a total of about 2 minutes waiting for PARSELINE.
Composing the original post, reading the replies, and writing this,
took another 15min. :-)

Hope this explains my problem better. Thanks for all replies. I'm
off now to check Mr. Bauer's suggestions.

You may wonder why use XML. Well, It strated out as a challenge.
But, after I did it for the first time, I was really impressed that I
could add some intelligent information to my data files, and my file
reader would be able to read them, or skip them, or whatever. So for
now, I continue to use them.

Mirko