Subject: Re: Faster way to search/split a string?
Posted by David Fanning on Wed, 23 Nov 2011 14:40:19 GMT
View Forum Message <> Reply to Message

Rob writes:

> I was hoping that someone might have a cleverer way of approaching
> this problem.
>
> The following command is the bootleneck in my code:
>
> row_of_data=strsplit(all_rows[(where(stregex(all_rows, id, /boolean)
> eq 1))[0]],' ', /extract)
>
> I have a large text file with lots of columns of data (which I don't
> know exactly what the columns are until I've read them in). There are
> then say 10000 rows of this data.
>
> This is all read into one large string array (all_rows) which contains
> each row as a single very long string.
> The first 20 characters of the row contain a unique id which I need to
> search the rows for and then extract the entire matching row. This row
> then needs to be split up into it's columns (space delimited).
>
> Hopefully that all makes sense.
>
> The problem is having to do this 10000 times, (once for for each id)
> is very slow and the time to do all of the other stuff done in the
> code, reading, writing, some maths, etc is being dominated by this one
> line.
>
> Any thoughts or suggestions?

I guess I would try just reading the IDs. Then sort the IDs,
and use Value_Locate to find all the unique IDs at once. Then
you could subset your string array and do the string extraction.
Hard to tell if this would be faster.

> p.s. This needs to be GDL compatible as well which I think most
> solutions would be anyway.

This is like telling most of us "be sure it works in the
Martian atmosphere." How the hell would I know!? :-)

Cheers,

David

--
David Fanning, Ph.D.
Fanning Software Consulting, Inc.
Coyote's Guide to IDL Programming: http://www.idlcoyote.com/
Sepore ma de ni thui. ("Perhaps thou speakest truth.")