Subject: Re: reading and writing very slow
Posted by R.Bauer on Fri, 11 Feb 2011 14:09:32 GMT
View Forum Message <> Reply to Message

Am 11.02.2011 14:12, schrieb geoff:
> Hi
>
> I have some 1-2 GB text files (lots of them!), each containing weather
> data for many thousands of stations for 1 year (per file). I need to
> get the data out of the year files and into files which have all the
> data for 1 weather station. It's easy but slow.
>
> I am reading each year file line by line and appending that line to
> the filename of the station (which is one of the fields on the line).
> (openw.../append, close). Does opening and closing files so many
> times have such an overhead? Is there a quicker way?

no

but reading line by line has.

Reimar

>
> thanks
>
>
>

Subject: Re: reading and writing very slow
Posted by oxfordenergyservices on Fri, 11 Feb 2011 15:16:41 GMT
View Forum Message <> Reply to Message

On Feb 11, 2:09 pm, Reimar Bauer <R.Ba...@fz-juelich.de> wrote:
> Am 11.02.2011 14:12, schrieb geoff:
>
>> Hi
>
>> I have some 1-2 GB text files (lots of them!), each containing weather
>> data for many thousands of stations for 1 year (per file). I need to
>> get the data out of the year files and into files which have all the
>> data for 1 weather station. It's easy but slow.
>
>> I am reading each year file line by line and appending that line to
>> the filename of the station (which is one of the fields on the line).
>> (openw.../append, close). Does opening and closing files so many

>> times have such an overhead? Is there a quicker way?
>
> no
>
> but reading line by line has.
>
only way i know how. variable length ascii unfortunately :(

Subject: Re: reading and writing very slow
Posted by ben.bighair on Fri, 11 Feb 2011 15:27:53 GMT
View Forum Message <> Reply to Message

On 2/11/11 10:16 AM, geoff wrote:
> On Feb 11, 2:09 pm, Reimar Bauer<R.Ba...@fz-juelich.de> wrote:
>> Am 11.02.2011 14:12, schrieb geoff:
>>
>>> Hi
>>
>>> I have some 1-2 GB text files (lots of them!), each containing weather
>>> data for many thousands of stations for 1 year (per file). I need to
>>> get the data out of the year files and into files which have all the
>>> data for 1 weather station. It's easy but slow.
>>
>>> I am reading each year file line by line and appending that line to
>>> the filename of the station (which is one of the fields on the line).
>>> (openw.../append, close). Does opening and closing files so many
>>> times have such an overhead? Is there a quicker way?
>>
>> no
>>
>> but reading line by line has.
>>
> only way i know how. variable length ascii unfortunately :(

Hi,

I can't tell from your description if exactly how you are managing the
output process so this might not be all that helpful. If I had to do
this in IDL, I would use something like Mike Galloy's resizeable array
list. See...

http://michaelgalloy.com/2006/04/24/collection-package-mgarr aylist.html

I might use that array list to aggregate all of the data. When the
aggregation is complete I would then dump it all to file at once.

For each input file you can read in all the data at a swipe and then
parse as needed within IDL. That would might be a lot faster than
trying to read in formatted lines one-at-a-time.

Cheers,
Ben

Subject: Re: reading and writing very slow
Posted by oxfordenergyservices on Fri, 11 Feb 2011 17:06:39 GMT
View Forum Message <> Reply to Message

On Feb 11, 3:27 pm, Ben Tupper <ben.bigh...@gmail.com> wrote:
> On 2/11/11 10:16 AM, geoff wrote:
>
>
>
>
>
>
>
>
>
>
>> On Feb 11, 2:09 pm, Reimar Bauer<R.Ba...@fz-juelich.de> wrote:
>>> Am 11.02.2011 14:12, schrieb geoff:
>
>>>> Hi
>
>>>> I have some 1-2 GB text files (lots of them!), each containing weather
>>>> data for many thousands of stations for 1 year (per file). I need to
>>>> get the data out of the year files and into files which have all the
>>>> data for 1 weather station. It's easy but slow.
>
>>>> I am reading each year file line by line and appending that line to
>>>> the filename of the station (which is one of the fields on the line).
>>>> (openw.../append, close). Does opening and closing files so many
>>>> times have such an overhead? Is there a quicker way?
>
>>> no
>
>>> but reading line by line has.
>
>> only way i know how. variable length ascii unfortunately :(
>
> Hi,
>
> I can't tell from your description if exactly how you are managing the
> output process so this might not be all that helpful. If I had to do

> this in IDL, I would use something like Mike Galloy's resizeable array
> list. See...
>
> http://michaelgalloy.com/2006/04/24/collection-package-mgarr aylist.html
>
> I might use that array list to aggregate all of the data. When the
> aggregation is complete I would then dump it all to file at once.
>
> For each input file you can read in all the data at a swipe and then
> parse as needed within IDL. That would might be a lot faster than
> trying to read in formatted lines one-at-a-time.

Thanks for this I'll take a look. Turns out is was my fault.
Processing on one linux machine, filesystem on another.

Subject: Re: reading and writing very slow
Posted by R.Bauer on Mon, 14 Feb 2011 16:29:05 GMT
View Forum Message <> Reply to Message

Am 11.02.2011 16:16, schrieb geoff:
> On Feb 11, 2:09 pm, Reimar Bauer <R.Ba...@fz-juelich.de> wrote:
>> Am 11.02.2011 14:12, schrieb geoff:
>>
>>> Hi
>>
>>> I have some 1-2 GB text files (lots of them!), each containing weather
>>> data for many thousands of stations for 1 year (per file). I need to
>>> get the data out of the year files and into files which have all the
>>> data for 1 weather station. It's easy but slow.
>>
>>> I am reading each year file line by line and appending that line to
>>> the filename of the station (which is one of the fields on the line).
>>> (openw.../append, close). Does opening and closing files so many
>>> times have such an overhead? Is there a quicker way?
>>
>> no
>>
>> but reading line by line has.
>>
> only way i know how. variable length ascii unfortunately :(

If you know the data structure you can design an idl structure and read
directly into that.

for example if that is a piece of your data

a='21.4 4544 5656.234'

then define a structure of

s= {temp:0.0, count:0L, height:0.0}

and use reads

reads, a, s

```
IDL> help,s,/str
** Structure <13a32a8>, 3 tags, length=12, data length=12, refs=2:
   TEMP            FLOAT          21.4000
   COUNT           LONG            4544
   HEIGHT          FLOAT          5656.23
```

for multiple lines use an array of the structure, e.g.

a = ['21.4 4544 5656.234', '22.3 4567 5555.1']

s = replicate({temp:0.0, count:0L, height:0.0}, 2)

```
IDL> help,s[0],/str
** Structure <13a3798>, 3 tags, length=12, data length=12, refs=2:
   TEMP            FLOAT          21.4000
   COUNT           LONG            4544
   HEIGHT          FLOAT          5656.23

IDL> help,s[1],/str
** Structure <13a3798>, 3 tags, length=12, data length=12, refs=2:
   TEMP            FLOAT          22.3000
   COUNT           LONG            4567
   HEIGHT          FLOAT          5555.10
```

You see it does not matter if ascii or binary ;)

cheers
Reimar