

Re: Huge data

Source:

<http://www.tech-archive.net/Archive/DotNet/microsoft.public.dotnet.framework.adonet/2004-10/0406.html>

From: Sahil Malik (contactmethrumblog_at_nospam.com)

Date: 10/06/04

Date: Wed, 6 Oct 2004 15:35:00 -0400

I think I have a solution to this problem.

I faced a similar situation in believe it or not a financial system, so I can understand the pain.

What I did was, I inherited from DataSet and DataTable to create LargeDataset and LargeDatatable, and I would simply page parts of the dataset as requested by the callee from the disk to and from temp files – sort of like writing your own virtual memory paging scheme. Albeit my implementation was fairly simple compared to windows Vmem handling the idea being you have to reduce page faults, and my data access was fairly linear in nature so I could simply cut up the rows into smaller configurable chunks and page them as needed.

The good news is, since the inherited dataset is nothing but a dataset (inherited), you can use it to bind, or basically anything a dataset might do. But it's quite a lot of work overriding every behavior, and well the class(es) I ended up writing were multiple pages long.

But at the end of the day – it all worked !! :)

– Sahil Malik

You can reach me thru my blog at

<http://www.dotnetjunkies.com/weblog/sahilmalik>

"Jon Skeet [C# MVP]" <skeet@pobox.com> wrote in message
news:MPG.1bce13b78bff9c7598b5f7@msnews.microsoft.com...

> jy <jy@discussions.microsoft.com> wrote:

>> > Well, you **could** use DataReader, but I wouldn't recommend that. Just
>>> use a select statement that selects a single row, or a page-worth of
>>> rows.

>>

>> I haven't much used dataReaders. Can you tell me how to select a group
of

>> rows and sequentially call the next group when needed?

>

> Doing that would be a bad idea, IMO. You really don't want to use

> *DataReader here.*
>
>>>> *2– won't it be too slow?*
>>>
>>> *Have you tried it? We don't know much about your database – you haven't told us much information about the situation.*
>>>
>> *Well, the database is made of over 1000 Tables. You have around 50 users.*
>> *It's an Oracle 9i DataBase. I haven't tried it but I suspect that having to connect everytime to the database to get a record or a group of records will take more time than retrieving all the data and work with it offline, no?*
>
> *Not necessarily – and if working with it offline just isn't an option, as it seems it isn't, given the exception you've got, surely it's better to have something that works than something that doesn't.*
>
> *Note that connecting every time doesn't necessarily mean making a new underlying database connection each time – connection pooling takes care of that for you.*
>
>>>> *3– How will I search the tables for specific rows?*
>>>>
>>>> *Using a select statement.*
>>>>
>>>> *Yes, but by using select statement, I will lose the feature of DataViewManager to filter locally the rows and displaying the rows I need at any time. Plus, I will have to send a query to the database everytime.*
>>>>
>>>> *And do you have evidence (rather than just guesses) that that will be too slow?*
>>>>
>>>> *Don't forget that you could always fetch, say, 100 records and then allow further filtering within those.*
>>>>
>>>>> *4– Will it be possible to update / delete / insert rows?*
>>>>>
>>>>> *Yes. For update/delete, you use a row you've already fetched. For insert, you don't need any existing data, do you?*
>>>>>
>>>>> *Yes but it won't be working like the rest of the application where I could update many rows before send the commit to the database.*
>>>>>
>>>>> *So keep a table with all the data you're updating and then send it in one go – that still doesn't mean you need to have the whole table in*

- > *memory at a time, unless you're updating every row in the table.*
- >
- > > *In a way, I know that all these arguements are minor problems, but I need to*
- > > *know if it's possible in this case to make it work like the rest of the*
- > > *application (DataViewManager filter, multiple update before commit, ...), or*
- > > *if it will have to be different.*
- >
- > *If you can't fit the whole of the data into memory, as it looks like*
- > *you can't, then it will clearly have to be different.*
- >
- > --
- > *Jon Skeet - <skeet@pobox.com>*
- > *<http://www.pobox.com/~skeet>*
- > *If replying to the group, please do not mail me too*