Subject: Re: Huge datasets under Win2000
Posted by david[2] on Wed, 04 Jul 2001 16:31:51 GMT
View Forum Message <> Reply to Message

Joe Means writes:

> I run Win2000 on a PC with 1Gb RAM and lots
> of hard disk space.

I think the critical factor will be what version of
IDL you are using.

Cheers,

David

--
David Fanning, Ph.D.
Fanning Software Consulting
Phone: 970-221-0438 E-Mail: davidf@dfanning.com
Coyote's Guide to IDL Programming: http://www.dfanning.com/
Toll-Free IDL Book Orders: 1-888-461-0155

Subject: Re: Huge datasets under Win2000
Posted by Joe Means on Wed, 04 Jul 2001 17:06:39 GMT
View Forum Message <> Reply to Message

I am using IDL 5.4
Joe

David Fanning wrote:

> Joe Means writes:
>
>> I run Win2000 on a PC with 1Gb RAM and lots
>> of hard disk space.
>
>
> I think the critical factor will be what version of
> IDL you are using.
>
> Cheers,
>
> David
>

--

Joseph E. Means
Assistant Professor, joe.means@orst.edu
Department of Forest Science
Oregon State University
Corvallis, OR 97331-5752
541-750-7351

Subject: Re: Huge datasets under Win2000
Posted by Craig Markwardt on Wed, 04 Jul 2001 17:37:28 GMT
View Forum Message <> Reply to Message

Joe Means <joe.means@orst.edu> writes:

> I have to read in over 13,000,000 observations [eventually over
> 33,000,000] with the structure
> dataShort = {dataShort, returnNum:0L, x:0D,y:0D,z:0D}
>
> At first, either when reading in or processing, I ran out of memory with
> a message about IDL not being able to allocate memory to make the
> array. So I read this into an Assoc variable which works fine. Now ,
> however, when I try to access parts of this data [dataS is the
> Associated structure array]:
> datSvX = (dataS.y-yll)/cellsize
> I get the error:
> % File expression not allowed in this context: DATAS.
>
> Any suggestions would be appreciated. It appears to me that Eric
> Korpela's VARRAY refered to in a thread earlier this week will not work
> on Win2000 Intel machines. I run Win2000 on a PC with 1Gb RAM and lots
> of hard disk space.

It looks like ultimately your 33,000,000 will just barely fit into
core. Thus, any strategy that involves reading the entire set of data
into memory at once seems doomed to failure. Also, recall that IDL is
fond of creating temporary arrays for expressions, so you will always
need a factor of 2-4 more memory in practice.

I have no direct experience with ASSOC or VARRAY.

However, I routinely process over 500 million events of a similar
type. The secret is to break things up into chunks, and process them
that way. Somehow I doubt that you really need all the data in memory
all the time (I hope). But if you do, then you probably need a huge
memory upgrade right now! :-)

Craig

--
---------- ----------
Craig B. Markwardt, Ph.D. EMAIL: craigmnet@cow.physics.wisc.edu
Astrophysics, IDL, Finance, Derivatives | Remove "net" for better response
---------- ----------

Subject: Re: Huge datasets under Win2000
Posted by david[2] on Wed, 04 Jul 2001 18:51:38 GMT
View Forum Message <> Reply to Message

Joe Means writes:

> have to read in over 13,000,000 observations [eventually over
> 33,000,000] with the structure
> dataShort = {dataShort, returnNum:0L, x:0D,y:0D,z:0D}
>
> At first, either when reading in or processing, I ran out of memory with
> a message about IDL not being able to allocate memory to make the
> array. So I read this into an Assoc variable which works fine. Now ,
> however, when I try to access parts of this data [dataS is the
> Associated structure array]:
> datSvX = (dataS.y-yll)/cellsize
> I get the error:
> % File expression not allowed in this context: DATAS.

Joe, sorry. I was distracted this morning what with
playing tennis and watching that Henman/Federer match
at Wimbledon. In fact, I don't know why in the world
I was even reading the IDL newsgroup! :-(

But I don't think this error has anything at all to
do with reading large data sizes. I think it has to
do with the way you are accessing the associated
variable. (Although I admit I've never seen the
error message before.)

If I understand your message correctly, DATAS is the
associated variable. Is this right?

```
dummy = {dataShort, returnNum:0L, x:0D,y:0D,z:0D}
OpenU, lun, 'mydatafile.dat', /Get_Lun
datas = Assoc(lun, {dataShort})
```

It is really incorrect to think of this as an
array of structures, which I think is what you may
be doing. In other words, this syntax doesn't work
with associated variables:

b = datas(4:6)

With associated variables, you are going to have
to pull each structure out to work with it:

```
struct1 = datas(5)
struct1.x = 5.5
datas(5) = struct1
```

I can't really tell what you are doing with this
syntax:

```
datSvX = (dataS.y-yll)/cellsize
```

But, I'm pretty sure this is the source of your
error.

Cheers,

David

--
David Fanning, Ph.D.
Fanning Software Consulting
Phone: 970-221-0438 E-Mail: davidf@dfanning.com
Coyote's Guide to IDL Programming: http://www.dfanning.com/
Toll-Free IDL Book Orders: 1-888-461-0155

Subject: Re: Huge datasets under Win2000
Posted by Joe Means on Wed, 04 Jul 2001 21:33:57 GMT
View Forum Message <> Reply to Message

David and Craig,
Thanks for your responses. You're right, David, dataS is the associated
variable with code like you wrote. The code below includes scalars yll
and cellsize, so it is designed to operate on the vector dataS.y to
produce the vector datSvY:

```
datSvX = (dataS.y-yll)/cellsize
```

I see now this will not work. The Assoc function is much less useful
than I had thought. Seems the additional fuctionality could be built in.

I'll look into Associating each vector of this structure individually,
before rewriting the whole program to process data one chunk at a time.
Or.... I could just go ahead and do that now, preparing the program for
yet larger datasets that will come eventually. Hmmmm....

Cheers,
Joe

Joe Means wrote:

> I have to read in over 13,000,000 observations [eventually over
> 33,000,000] with the structure
> dataShort = {dataShort, returnNum:0L, x:0D,y:0D,z:0D}
>
> At first, either when reading in or processing, I ran out of memory
> with a message about IDL not being able to allocate memory to make the
> array. So I read this into an Assoc variable which works fine. Now ,
> however, when I try to access parts of this data [dataS is the
> Associated structure array]:
> datSvX = (dataS.y-yll)/cellsize
> I get the error: % File expression not allowed in this context: DATAS.
>
> Any suggestions would be appreciated. It appears to me that Eric
> Korpela's VARRAY refered to in a thread earlier this week will not
> work on Win2000 Intel machines. I run Win2000 on a PC with 1Gb RAM
> and lots of hard disk space.
>

--
Joseph E. Means
Assistant Professor, joe.means@orst.edu
Department of Forest Science
Oregon State University
Corvallis, OR 97331-5752
541-750-7351