
Subject: Re: How to improve the efficiency of IDL on cluster
Posted by [Michael Galloy](#) on Mon, 03 Jun 2013 21:25:01 GMT
[View Forum Message](#) <> [Reply to Message](#)

On 6/3/13 9:09 AM, Chang Liao wrote:

> Hi! I am currently running IDL on a Linux cluster. Due to the massive
> mount of remote sensing dataset, it takes days to finish just one
> year's calculation. I can specific the computer node,
> memory(48G/96G), and cores(24 cores) for my QBS job. So I am thinking
> shall I use multi-threading or other ways to speedup the
> calculation. Does anyone have any suggestion or experience in this
> related topics, I couldn't find too much on Exelis or Google. I do
> remember IDL has some demo in earlier version, but it's gone in IDL
> 8.0 +.

Do you want a way to use multiple nodes of this cluster or speed up
processing on a single node? Do the nodes have GPUs?

The thread pool is built into IDL and automatically gives some
multi-threading help on certain operations. There is not much you can do
here except tweak some parameters. See the online help for the CPU routine

<http://www.exelisvis.com/docs/CPU.html>

and the Thread Pool:

http://www.exelisvis.com/docs/The__Thread_Pool.html

We at Tech-X have some third-party solutions to these types of problems:

1. TaskDL is a queue-based task farming solution. You create workers
corresponding to cores on a local or networked machine and then add
tasks to a queue. Workers take new tasks as they finish old ones. This
is for "embarrassingly parallel" problems.

2. mpiDL is a set of wrappers for MPI (Message Passing Interface).
This is more complicated and is for problems which require more
communication between workers.

3. GPULib is an easy way to speed up supported operations using
CUDA-enabled NVIDIA graphics cards.

Contact me for more information about any of these products!

Mike

--

Michael Galloy
www.michaelgalloy.com

Subject: Re: How to improve the efficiency of IDL on cluster
Posted by [Craig Markwardt](#) on Tue, 04 Jun 2013 04:14:21 GMT
[View Forum Message](#) <> [Reply to Message](#)

On Monday, June 3, 2013 11:09:00 AM UTC-4, Chang Liao wrote:

> Due to the massive mount of remote sensing dataset, it takes days to finish just one year's calculation.

>

> I can specific the computer node, memory(48G/96G), and cores(24 cores) for my QBS job. So I am thinking shall I use multi-threading or other ways to speedup the calculation.

I guess it's an obvious question but I'll ask it any way: have you already vectorized your IDL code?

I see many scientist types who prefer to write IDL code like this,

```
for i = 0, n_elements(v)-1 do begin
```

```
  v[i] = a[i] + b[i]
```

```
endfor
```

even though the obvious vectorized IDL version is,

```
v = a + b
```

which is both faster *and* easier to read.

The first step is to vectorize the inner loops of your calculations.

Use the profiler to find out where your computational bottlenecks are.

Craig
