

# How to Get More Memory for your Job

## Category: Memory Usage on Pleiades

If your job was terminated because it needed more memory than what was available in the nodes that it ran on, consider the following examples.

### Harpertown Nodes

Among the Harpertown nodes, the 64 nodes in rack 32 have 16 GB per node ( GB/node) instead of 8 GB/node. You can request running your job on rack 32 with the keyword **bigmem=true**. For example, change:

```
#PBS -lselect=1:ncpus=8  
  
to  
  
#PBS -lselect=1:ncpus=8:bigmem=true
```

Instead of running your jobs on Harpertown nodes, run them on Nehalem-EP, Westmere, or Sandy Bridge nodes. For example, change:

```
#PBS -lselect=1:ncpus=8:model=har  
  
to  
  
#PBS -lselect=1:ncpus=8:model=neh  
  
or  
  
#PBS -lselect=1:ncpus=8:model=wes  
  
or  
  
#PBS -lselect=1:ncpus=8:model=san
```

### Westmere Nodes

Among the Westmere nodes, 17 nodes have 48 GB/node and 4 nodes have 94 GB/node instead of 24 GB/node. You can request using some of these nodes with the keyword **bigmem=true** and **model=wes**. For example, change:

```
#PBS -lselect=1:ncpus=12:model=wes  
  
to
```

```
#PBS -lselect=1:ncpus=12:bigmem=true:model=wes
```

If you submit your resource request as shown above, your job will be assigned either a 48 GB or a 94 GB bigmem node, depending on availability.

To explicitly request a bigmem node with 94 GB of memory, add the `:mem` attribute with a memory size between 48 and 94 GB. For example:

```
#PBS -l select=1:ncpus=12:bigmem=true:mem=90GB:model=wes
```

Please note that these Westmere bigmem nodes can be used for jobs requesting the **normal**, **long**, **debug**, and **low** queues. They are not available for the **devel** or **gpu** queues.

## All Nodes

If all processes use about the same amount of memory and you cannot fit 8 processes per node (for Harpertown or Nehalem-EP), 12 processes per node (for Westmere), or 16 processors per node (for Sandy Bridge), then reduce the number of processes per node and request more nodes for your job. For example, change:

```
#PBS -lselect=3:ncpus=8:mpiprocs=8:model=neh
```

to

```
#PBS -lselect=6:ncpus=4:mpiprocs=4:model=neh
```

For a typical MPI job where rank 0 does the I/O and uses a lot of buffer cache, assign rank 0 to one node by itself. For example, if rank 0 needs up to 22.5 GB of memory by itself, change:

```
#PBS -lselect=1:ncpus=12:mpiprocs=12:model=wes
```

to

```
#PBS -lselect=1:ncpus=1:mpiprocs=1:model=wes+1:ncpus=11:mpiprocs=11:model=wes
```

If rank 0 needs 22.5 - 48 GB of memory by itself, use:

```
#PBS -lselect=1:ncpus=1:mpiprocs=1:bigmem=true:model=wes+1:ncpus=11:mpiprocs=11:model=wes
```

Note that due to formatting issues, the above may appear as two lines; it should be entered as a single line.

If you suspect that certain nodes which your job ran on had less total physical memory than normal, report it to the NAS Control Room. Those nodes can be "off-lined" and taken care of by NAS staff. This prevents you and other users from using those nodes before they are

fixed.

For certain pre- or post-processing work that needs more memory, you can use one of the Westmere bigmem nodes in a PBS batch job or run the job interactively on the bridge nodes (bridge[1-4]). Note that an interactive job cannot use more than 56 GB on bridge[1,2] or 192 GB on bridge[3,4]. Also, MPI applications that use SGI's MPT library cannot run on the bridge nodes.

For a multi-process or multi-thread job, if any of your processes/threads need more than 94 GB, the job won't run on Pleiades. Instead, run it on a shared-memory system such as Columbia.

---

Article ID: 222

Last updated: 08 Jan, 2013

Computing at NAS -> Best Practices -> Memory Usage on Pleiades -> How to Get More Memory for your Job

<http://www.nas.nasa.gov/hecc/support/kb/entry/222/?ajax=1>