GROMACS - Bug #3440
Multi-nodes run exits with error with openmpi/4.0.0
03/11/2020 04:40 PM - Yuxuan Zhuang

<table>
<thead>
<tr>
<th>Status:</th>
<th>New</th>
</tr>
</thead>
<tbody>
<tr>
<td>Priority:</td>
<td>Normal</td>
</tr>
<tr>
<td>Assignee:</td>
<td></td>
</tr>
<tr>
<td>Category:</td>
<td></td>
</tr>
<tr>
<td>Target version:</td>
<td></td>
</tr>
<tr>
<td>Affected version - extra info:</td>
<td></td>
</tr>
<tr>
<td>Affected version:</td>
<td>2020.1</td>
</tr>
<tr>
<td>Difficulty:</td>
<td>uncategorized</td>
</tr>
</tbody>
</table>

**Description**

Not sure it is related to [#3296](#3296), but I was trying to run multi jobs on three nodes, each with 32 logical cores and 4 gpus. Error output and tpr files are attached.

- Codes to reproduce the error

```
salloc -p lindahl --time=2:00:00 --nodes=3 -n 96
module load gromacs/2020.1
module load openmpi/4.0.0
mpirun -np 96 gmx_mpi mdrun -deffnm md -multidir walker(1..16) -ntomp 1 -nsteps 10000
```

- Codes that run successfully

```
salloc -p lindahl --time=2:00:00 --nodes=3 -n 96
module load gromacs/2020.1
module load openmpi/4.0.2
mpirun -np 96 gmx_mpi mdrun -deffnm md -multidir walker(1..16) -ntomp 1 -nsteps 10000
```

- Codes that works but only utilize one node.

(raise WARNING: On rank 0: oversubscribing the available 32 logical CPU cores per node with 96 threads. This will cause considerable performance loss.)

```
salloc -p lindahl --time=2:00:00 --nodes=3 -n 96
module load gromacs/2020.1
module load openmpi/4.0.0
mpirun -np 16 gmx_mpi mdrun -deffnm md -multidir walker(1..16) -ntomp 6 -nsteps 10000
```

**History**

**#1 - 03/20/2020 04:20 PM - Berk Hess**

Which MPI version was GROMACS built with (I think this is written in the log file)? You can't freely mix MPI versions, but normally you would expect mixing minor versions would work.

**#2 - 03/23/2020 02:10 PM - Szilárd Páll**

The linked output seems to indicate an hwloc crash, so probably not GROMACS-related. Can you try to build against a different OpenMP version?
Secondly, regarding launching multi-node, have you checked/do you know that the OpenMPI installation is configured with SLURM such that node allocations get passed to the MPI launcher?

On our cluster I've use the manual way where I allocate nodes first, then launched using mpirun --hostfile --host.

Files

<table>
<thead>
<tr>
<th>File</th>
<th>Size</th>
<th>Date</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>md.tpr</td>
<td>15.3 MB</td>
<td>03/11/2020</td>
<td>Yuxuan Zhuang</td>
</tr>
<tr>
<td>error_openmpi_4_0_0.txt</td>
<td>151 KB</td>
<td>03/11/2020</td>
<td>Yuxuan Zhuang</td>
</tr>
</tbody>
</table>