

# Lisa User Day 2011

## Efficient use of file systems

John Donners  
John.Donners@sara.nl

# What filesystems are available?

- ▶ home directory
- ▶ scratch directory
- ▶ archive
  
- ▶ on request: global scratch directory
- ▶
- ▶ **DO NOT use /tmp ! This is needed for system processes**

# home directory

- + Global filesystem, available everywhere
- + Backup
- - Slow
- - Limited storage (75GB)
  
- home-directory is /home/donners, or \$HOME
- Please note that changes on one node are not immediately visible on other nodes.
- Use the home directory in your job only to:
  - copy input files from home to scratch
  - copy output files from scratch to home
- Too much I/O in your home directory will dramatically slow down your job, but also jobs of everyone else!

# scratch directory

- ▶ + Fast
- ▶ - Only available during a job
- ▶ - Local filesystem
- ▶ - Limited storage: 220GB on 12-core nodes, 85GB on other nodes
  
- ▶ \$TMPDIR points to the scratch directory on each node.
- ▶ It is recommended to run jobs from the scratch directory.

# Example job using scratch

```
#PBS -lnodes=1 -lwalltime=5:00:00  
cp -r $HOME/datadir/problem1 $TMPDIR  
cp -r $HOME/datadir/problem2 $TMPDIR  
(  
  cd $TMPDIR/problem1  
  myprogram >out 2>err  
  cp out err results $HOME/datadir/problem1  
) &  
(  
  cd $TMPDIR/problem2  
  myprogram >out 2>err &  
  cp out err results $HOME/datadir/problem2  
) &  
wait
```

# scratch directory: mpicopy

- ▶ Sometimes, an application needs large input files, which have to be available on all nodes.
- ▶ mpicopy can be used to efficiently copy files from the home directory to the scratch directories on all nodes.
- ▶ all parameters of mpicopy are files that are copied.
- ▶ the files are always copied to \$TMPDIR.

```
module load mpicopy openmpi/gnu
```

```
mpicopy my_input.dat
```

```
module unload mpicopy openmpi/gnu
```

- ▶ there is no special mechanism to copy output files back to your home directory, just use cp.

# Global scratch directory

- For multi-node jobs, there's also the possibility for a global scratch space, using GlusterFS.
- The scratch disks of the first two nodes are combined into one filesystem of 170-440GB.
- That filesystem is globally visible on all nodes of your job as the directory **/global**.
  
- Note that the variable `$TMPDIR` still points to the local scratch directory (which is still usable)
- Explicitly use **/global** in your job to access the global scratch space.
  
- Interested? Send an email to [hic@sara.nl](mailto:hic@sara.nl) with a request to use GlusterFS.

# scratch directory interactively

- ▶ **The interactive nodes also have a scratch directory.**
- ▶ **It is much faster to use this directory, e.g. for:**
  - ▶ **compiling your program**
  - ▶ **analyzing output files**
  - ▶ **unpacking archives**



# archive directory

- + Backup
  - + Unlimited storage
  - + Also visible from huygens
  - - Slow
  - - Only directly accessible from login nodes
- 
- archive directory is /archive/donners
  - Don't use the archive in your batch job
  - Only copy files interactively to or from your archive.
  - More details about using the archive are shown in a later presentation.

# Too much output?

- ▶ If there is not enough space in your home-directory for all output, try to compress the files in the scratch-directory in your batch job. Especially text files can be compressed very well (approx. 10x).
- ▶ For many files or directories: use zip or tar
- ▶ For large files, please use a parallel compression program:
  - ▶ pigz to create a .gz-file, fast
  - ▶ pbzip2 to create a .bz2-file, slower but better compression
  - ▶ by default, these programs use all available cores.
- ▶ Then move the compressed files to your home directory:

```
pigz output_file.dat
```

```
cp output_file.dat.gz $HOME/output_dir
```

40  
JAAR  
1971  
2011



# Thank you for your attention!

Any questions?