GWDG
Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

HPC filesystems and a suggested workflow

Sebastian Krey

02.03.2022

Outline

HPC Filesystems

Most HPC systems provide several filespaces distributed over different filesystems. For the HPC systems operated by GWDG these are:

- Homedirectory (`$HOME`)
- Workstorage (`$SCRATCH` for SCC, `$WORK` for HLRN/NHR)
- Projectspace (directories below `/scratch/projects`)
- Node local SSDs
- Tapearchive (`$AHOME` for SCC, folders below `/perm` for HLRN/NHR)
- Shared SSDs

Intended usage and capabilities

Homedirectory:

Usage: Software, scripts, configuration files (only for SCC: important results, that are still in use and need a backup)

Backup: Daily backups (in the night) to tape (only HLRN/NHR: additional daily filesystemsnapshot for fast restores)

Quota: No inode quota (number of files and directories), but strict volume quota, extension requires support ticket (for SCC: extension usually possible, for HLRN/NHR extension only in limited cases permitted)

Speed: Medium/Low, not suitable for running I/O intensive computation

Intended usage and capabilities

Workstorage:

Usage: All data concerning compute jobs (input files, intermediate results, checkpoints, final results)

Backup: None

Quota:

- **SCC:** No quota limit, but regular request to clean up unused files, when the filesystem is getting too full
- **HLRN/NHR:** Inode and volume quota, hardlimit 10-times higher than softlimit, so a temporary increase (grace period 2 weeks) possible, extension needs support ticket and reasoning (preferred usage of projectspace)

Speed: High/Very high, perfect for high sequential performance, but random IO requires SSDs

Intended usage and capabilities

Projectspace:

Usage: Same usage as the workstorage, but shared between different members of a working group or project.
Location `/scratch/projects/PROJECTID`
Space created upon request (SCC: support ticket) or automatically as part of a compute project application (HLRN/NHR)

Backup: None

Quota: Like workstorage but easy quota extension for HLRN/NHR

Speed: High/Very high, perfect for high sequential performance, but random IO requires SSDs

Intended usage and capabilities

Node local SSDs:

Usage: Temporary files with high (random) I/O activity, which are needed only locally on the compute, access via `$TMP_LOCAL` (SCC) or `$LOCAL_TMPDIR` (HLRN/NHR)

Backup: None, automatic deletion at the end of the compute job, copying of important results has to be handled by the user in the jobscript.

Quota: Limited by capacity of node SSD (node dependent 250 GB up to 2 TB)

Speed: Very high, esp. for random I/O

Intended usage and capabilities

Shared SSDs:

Usage: Temporary files with high (random) I/O activity, which have to be shared between all nodes of a compute job
Access via `$TMP_SCRATCH` (SCC), similar access mode will be available on the HLRN/NHR systems in the near future.
The HLRN/NHR system Emmy provides additionally a NVMe based burst buffer (DDN IME), for documentation see:
https://www.hlrn.de/doc/display/PUB/IME+Burst+Buffer%2C+File+System+Cache

Backup: None, automatic deletion at the end of the compute job, copying of important results has to be handled by the user in the jobscript.

Speed: Very high, for sequential and random I/O

Intended usage and capabilities

Tape archive:

Usage: Inactive files, which have to be kept for reference or later usage, only container files possible (tar or zip archives), preferred size between 1 GB and 4 TB.

Quota: None for SCC
Inode and volume quota for HLRN/NHR (extension upon request via support ticket)

Speed: Very slow

Workflow

- Setup your software, configuration, scripts in the homedirectory
- Create a folder for the compute project in your workstorage, e.g. `/scratch/users/$USER/2022a-PaperXY`
- Copy all input files for your compute jobs to this folder
- Run your compute jobs
- Analyze your results
- Copy the important final results to your homedirectory (SCC) or local storage (HLRN/NHR)
- Cleanup the workstorage from temporary files, unneeded intermediate results, etc.

Workflow

- Create a tar/zip archive of unneeded files (please use threaded compression tools), which have to be kept for reference or future use and move the file to the tape archive, cleanup the folder in the workstorage

```
PIGZ="-1 -p 8 -R" tar -I pigz -cf $Project-archive.tar.gz $Project
mv $Project-archive.tar.gz $AHOME/
rm -rf $Project
```

or for xz compression

```
XZ_OPT='-0 -T8' tar -cJf $Project-archive.tar.xz $Project
mv $Project-archive.tar.xz $AHOME/
rm -rf $Project
```