

Managing data on Colossus

Available file systems on Colossus

File system	Path	Recommended use
Colossus project directory	/cluster/projects/pXX	Software, job configurations, input files, processing data in a job. DO NOT use it for long-term data storage unless you have backup enabled. 1 TiB * disk quota
Home directory	/tsd/pXX/home/user or \$HOME	Software, job configurations. DO NOT use it for processing data in a job. 100 GiB disk quota
Scratch	/cluster/work/jobs/ <i>jobid</i> or \$SCRATCH	Processing data in a job, <u>chkfile</u> to retain output data.
Local disk space	\$LOCALTMP	Processing data in a very high I/O job. 100-200 GiB disk quota

(*) Additional disk space can be requested.

Colossus project directory

Each project with access to Colossus has a project directory on the high-performance parallel filesystem: `/cluster/projects/pXX`, where `pXX` is your project number. By default, the disk quota of the project directory is 1 TiB.

By default, there is no backup of the data stored in `/cluster/projects/pXX`, but snapshots are available for the last 7 days in a `.snapshot` subdirectory. If you need to work on several TiB of data and moving the data between the `/tsd/pXX/data/durable` (HNAS) and the Colossus project directory is not convenient, then we strongly recommend to have backup of `/cluster/projects/pXX`. Backup is a service on demand. Project administrators can order it by contacting TSD. [Pricing for the backup can be found here](#). A list of projects with backup enabled can be found [here](#).

Serving and keeping track of a parallel cluster file system is a complicated task. While the hardware is high-end, once in a while the GPFS software will get in trouble. When this happens users might experience delay when doing simple commands as "ls" or even hangs. Usually these problems lasts for a very short time, but if there is a serious problem there will be an announcement on the [TSD Operational Log](#) and on the Colossus Users email list.

On the submit host

On the submit hosts this file system is available using NFSv4 with Kerberos authentication (see below) and mounted under `/cluster/p/pXX/cluster` with symlinks pointing to `/cluster/projects/pXX` for historical reasons. We advice to use references to `/cluster/projects/pXX` in your scripts.

Kerberos authentication

Access to the `/cluster/projects/pXX` file system over NFSv4 requires a valid [Kerberos](#) ticket. A valid ticket will grant you access to the `/cluster/projects/pXX`, whereas an expired, invalid ticket will deny access.

If you connect to the submit host (via ssh or PUTTY) you'll automatically be granted a ticket for a 10 hour session which is automatically renewed up to a week. If the ticket expires after a week, you'll have to log out and back in to restore access to `/cluster/projects/pXX`. This is the preferred method of obtaining a ticket.

You can also manually obtain a ticket using the `kinit` command. However, this ticket will **not** renew and expire after 10 hours. We advice you **not** to use this command. Obtaining automatic and manual tickets at the same time may result in `permission denied` errors if one of the tickets expires while the other is still valid.

Kerberos authentication requires password authentication, hence you'll **not** be given a ticket if you connect using ssh keys. Please disable ssh keys on the submit host and use password authentication instead.

You can list your current ticket status using:

```
klist
```

In the initial stage it may only list the entry for the Ticket Granting Ticket (TGT) indicating a successful password verification:

```
-bash-4.2$ klist
Ticket cache: FILE:/tmp/krb5cc_7927__Vx2FH
Default principal: p11-bartt@TSD.USIT.NO

Valid starting    Expires          Service principal
07/06/2020 14:50:43 07/07/2020 00:50:43 krbtgt/TSD.USIT.NO@TSD.USIT.NO
        renew until 07/13/2020 14:50:21
```

Once you access `/cluster/projects/pXX`, entries for access to `nfs/ess01.tsd.suit.no` will be added, indicating successful authorization to the nfs mount:

```
-bash-4.2$ klist
Ticket cache: FILE:/tmp/krb5cc_7927__Vx2FH
Default principal: p11-bartt@TSD.USIT.NO

Valid starting    Expires          Service principal
07/06/2020 14:50:43 07/07/2020 00:50:43 krbtgt/TSD.USIT.NO@TSD.USIT.NO
        renew until 07/13/2020 14:50:21
07/06/2020 14:52:44 07/07/2020 00:52:44 nfs/ess01.tsd.usit.no@TSD.USIT.NO
        renew until 07/13/2020 14:50:21
```

If your ticket expires, you'll need to re-authenticate to obtain a new ticket. Log out and back in. If you get permission denied or cannot list the contents when you access a directory for the first time, but you do have a valid ticket, Kerberos authentication may have been delayed and will succeed if you try again.

On Colossus

On Colossus the file system is concurrently available on all compute nodes using the global parallel file system (GPFS). On the compute nodes its mounted under `/gpfs`, with symlinks pointing to `/cluster/projects/pXX` to resemble the NFS mount points on the submit hosts. On the compute nodes, the communication with this storage-system is over ultra-fast 56 Gbps Infiniband.

Home directory

Each user has a home directory (`$HOME`) on a low-performance file system. By default, the disk quota for the home directory is 100 GiB (see below).

The home directory is backed up regularly (see below), but anything inside directories named *nobackup* is skipped. Backup is slow and expensive, so please put temporary files, files that can be downloaded again, installed software and other files that can easily be recreated or do not need to be backed up inside a *nobackup* directory.

Also note that one is not supposed to use the home directory as read/write area for jobs, especially not I/O intensive jobs. Use the scratch area for that (see below).

Scratch disk space

While a job runs, it has access to a temporary scratch directory on `/cluster/work/jobs/jobid` which resides on the high-performance GPFS filesystem. The directory is individual for each job, is automatically created when the job starts, and is deleted when the job finishes (or gets requeued). *There is no backup of this directory.* The name of the directory is stored in the environment variable `$SCRATCH`, which is set within the job script.

In general, jobs should copy their work files to `$SCRATCH` or `/cluster/projects/pXX` and run there. This is especially important for I/O intensive jobs. The scratch disk is faster than the home directory disk, and running I/O intensive jobs in `$HOME` slows down not only the job, but also interactive work for other users.

If you need to access the scratch directory from *outside* the job (for instance for monitoring a running job), the directory is `/cluster/work/jobs/jobid`, where *jobid* is the job id of the job in question.

Local disk space

For very intensive IO, it can be useful to use the local drives on the compute nodes. The path to the directory is stored in the environment variable `$LOCALTMP`. The compute and GPU nodes have 100 GB and 200 GB of local storage, respectively. Add the following to your batch script to request (e.g. 20 GB) of local temporary storage on the node:

```
#SBATCH --gres=localtmp:20
cleanup cp $LOCALTMP/outputfile $SLURM_SUBMIT_DIR
```

Disk quotas

All projects receive a 1 TiB Colossus project directory and a 100 GiB home directory disk space quota. Projects may apply for extra disk space from [Sigma2](#).

On Colossus and the submit hosts, the UNIX `df` utility can be used to query disk usage on the GPFS file system. A disk will be full if either the space or the number of inodes (files) runs out. To query disk space, use:

```
$ df -h /cluster/projects/p11
Filesystem                Size      Used Avail Use% Mounted on
ess01.tsd.usit.no:/p11    1.0T    966G   59G   95% /cluster/p/p11
```

To query the number of inodes, use:

```
$ df -ih /cluster/projects/p11
Filesystem                Inodes     IUsed IFree IUse% Mounted on
ess01.tsd.usit.no:/p11    1.1G     22M   1.1G    3% /cluster/p/p11
```

Data transfer to/from Colossus

To copy data from the the `/tsd/pXX/data/durable` project directory to the Colossus `/cluster/projects/pXX` project directory you'll have to connect to the submit host which has access to both.

Secure copy with `scp` is a common way of transferring files. However, if the communication fails half way through `scp` will not pick up where it stopped. In addition the transfer is not compressed, which is an advantage for bandwidth limited transfers. The `rsync` utility is the preferred way of moving large files.

```
rsync [OPTION]... SRC [SRC]... [USER@]HOST:DEST
```

Please see the `rsync` manual page (`man rsync`) for more information.

Data compression

Millions of small files pose a challenge for GPFS and is to be avoided. If possible pack the small files in archives making operations on them easy. One way is to copy the archive to `$SCRATCH` or `/cluster/projects/pXX`, unpacking them there and working on the local file tree.

A standard unix/linux utility is `gzip`. Read the man pages for more information.

```
gzip file.dta
```

This will produce a file `file.dta.gz`, hopefully a much smaller file. Not all types of data compress equally well. Text compresses well, jpg pictures not well at all. For files that are to be unpacked at Windows machines the `zip` utility can also be used. One limitation of `zip` is that neither the input files nor the resulting archive can be larger than 4 GB. For files larger than 4 GB use `gzip`. Giving it a numeric argument like `-9` forces higher compression at the expense of longer compression time. A more efficient alternative is `bzip2`.

To unpack the file :

```
gunzip file.dta.gz
```

This will result in the original file in its uncompressed form.

Backup and restore

See [here](#) for more information on backup and restore options.



Contact support



Chat



Register
case



Video call



Phone

Did you find what you were looking for?

[Give us feedback](#)

Published June 21, 2021 10:35 AM - Last modified Feb. 25, 2022 2:16 PM



Share by
E-mail



Share on
Facebook



Share on
Twitter



UNIVERSITY OF OSLO

CONTACT INFORMATION

[Contact us](#)

[Find us](#)

ABOUT THE WEBSITE

[Cookies](#)

[Accessibility
statement](#)

[\(in
Norwegian
only\)](#)

RESPONSIBLE FOR THIS PAGE

[Web editor USIT](#)

[Log in](#)

