

File Transfer from appnodes to NIRD/SAGA

- Home directories: 20GB – 100k files (`/cluster/home/<username>` or `/nird/home/<username>`)
- `/nird/home/<username>/backup/<hpc>` contains backup of hpc homedirs.

Currently SAGA homedirs are not backed-up

- hpc projects (nnxxxxk): 1TB up to 10TB(`/cluster/projects/nnxxxxk`).
- shared projects in `/cluster/shared` on request to sigma2.
- `/cluster/work/users/<username>`: no quota limit but no backup and deletion after 42d max (eventually 21d)

- NIRD projects (NsxxxxK): depends of application and RFK (`/projects/NSxxxxK`)
- scratch 15TB in total (`/scratch/<username>`)
- no shared project (but project owners can eventually open their projects or use acls).

Network speed between OSL and SAGA (Forskningnett)
100Gb/s

Inside OSL
20Gb/s (?)

App-nodes – Abel
1Gb/s → ~8TB per day.

Network speed between SAGA and NIRD
2x 25Gb/s → ~20TB/h

On your appnode to transfer your data which are in folder
/<some_path>/MyDataset

- case “**not too many files and few TBs of data**”

- for SAGA

```
rsync -avxh /<some_path>/MyDataset username@saga.sigma2.no:/cluster/projects/nn1234k/
```

- for NIRD

```
rsync -avxh /<some_path>/MyDataset username@login.nird.sigma2.no:/projects/NS1234K/
```

Typical speed: ~50MB/s → 4TB/day

Tips: check rsync options – for example “-z” to compress before sending or –partial-dir=.partial-dir to resume transfer in case of connection drop while transferring a large file. --progress to print out status...

- case “**not too many files but many TBs of data**”

Use if possible Abel module parsyncfp (parallel rsync program)

- for SAGA:

```
parsyncfp --rsyncopts="-ax --ignore-existing --relative" --NP=4 --startdir="/<some_path>"
```

```
MyDataset username@saga.sigma2.no:/cluster/projects/nn1234k/
```

- for NIRD:

```
parsyncfp --rsyncopts="-ax --ignore-existing --relative" --NP=4 --startdir="/<some_path>"
```

```
MyDataset username@login.nird.sigma2.no:/projects/NS1234k/
```

NP:number of parallel threads (do not use too many threads as it probably won't make the transfer faster)

--maxbw=xxxxx (for example 100000 to limit bandwidth usage to 100MB/s)

--maxload=xx (for example 10 to suspend threads if load on the appnode is higher than 10 – use top command to check the load)

- case “**many files and few TBs of data**”

First archive your dataset with tar:

```
tar zcvf MyDataset.tar.gz /<some_path>/MyDataset
```

then transfer with rsync and untar:

- for SAGA:

```
rsync -avxh MyDataset.tar.gz username@saga.sigma2.no:/cluster/projects/nn1234k/
```

Then on SAGA

```
tar zxvf MyDataset.tar.gz
```

- for NIRD:

```
rsync -avxh MyDataset.tar.gz username@login.nird.sigma2.no:/projects/NS1234k/
```

Then on NIRD

```
tar zxvf MyDataset.tar.gz
```

- case “**many files and many TBs of data**”

Combine previous solutions!

- File transfer between SAGA and NIRD

Currently the NIRD storage is not mounted on SAGA. But it will be NFS mounted on the login nodes at some point. In the meantime users can apply same methods as describe above to migrate data between the two systems.