

Lilac Storage Primer

This page documents the various questions related to lilac storage. Lilac storage is primarily divided into 4 categories.

Lilac home storage :

- **Description** : GPFS shared parallel filesystem, not replicated, and not backed up.
- **Purpose**: To store software related code and scripts, default quota size is small and fixed.
- **Mount**: /home/<user>
- **Access**: All Lilac nodes, including compute storage and login nodes.
- **Default quota**: 100GB
- **Snapshots**: 7 days of snapshots. (not backed up). Can be accessed in /home/.snapshots/<user>
- **Replicated**: no

Lilac compute storage :

- **Description** : GPFS shared parallel filesystem, not replicated, and not backed up.
- **Purpose**: For jobs to read and write compute data from login and compute nodes, default quota size is larger with flexibility to request larger quota.
- **Mount**: /data/<lab group>
- **Access**: All Lilac nodes, including compute storage and login nodes.
- **Default quota**: 5TB (Increased/Decreased on request)
- **Snapshots**: 7 days of snapshots. (not backed up). Can be accessed in /data/.snapshots/<date>/<lab group>
- **Replicated**: no

Lilac warm storage :

- **Description** : GPFS shared parallel filesystem, not replicated but will be replicated in near future, and not backed up. Comparatively slower than lilac compute storage.
- **Purpose**: To store long term data. Only accessible from login nodes and cannot be accessed from compute nodes.
- **Mount**: /warm/<lab group>
- **Access**: Only lilac and luna login nodes.
- **Default quota**: 5TB (Increased/Decreased on request)
- **Snapshots**: 7 days of snapshots. (not backed up). Can be accessed in /warm/.snapshots/<date>/<lab group>
- **Replicated**: no (will be replicated in near future)

Lilac local scratch storage :

- **Description** : XFS filesystem, not replicated, and not backed up. Local and not a shared filesystem, slower than GPFS.
- **Purpose**: To store local temporary data related to compute jobs. Since this is not a shared filesystem, the temporary data needs to be cleaned up and copied back to shared filesystem after job completion.
- **Mount**: /scratch/
- **Access**: Only lilac compute nodes.
- **Default quota**: No quota and limited to free disk space in /scratch.
- **Snapshots**: No snapshots.
- **Replicated**: no

How to :

Check Quota for GPFS filesystem:

- Lilac home storage :

Command line

```
mmlsquota lilac:home --block-size auto
```

- Lilac compute storage :

Command line

```
mmlsquota -j data_<lab group name> lilac --block-size auto
```

- Lilac warm storage (oscar) :

Command line

```
mmlsquota -j warm_<lab group name> oscar --block-size auto
```

mmlsquota gives information about quota on number of files too, along with information about block quota.

Filesystem	Fileset	type	blocks	quota	limit	in_doubt	grace
Filesystem name	Fileset name	fileset/usr/grp	Blocks currently occupied	Your block quota	Your limit for "7 days" beyond quota.	Blocks in doubt that will be counted towards your quota. Happens when many files are added/deleted recently.	Countdown of once you occur than mentioned

Once the number of blocks or number of files reach the value mentioned in "quota" - Storage system will give 7 days as a grace period, to fill up until the max value mentioned in "limit" is reached. Storage system will not allow any more data to be written when:

1. The block limit/file limit is reached.
2. 7 days have passed since the blocks/files have occupied more than "quota". The grace field will show you the number of days left, before which the number of blocks/files need to go less than the value mentioned in "quota".

2. Copy files from other clusters:

HAL cluster is outside the firewall, so lilac cannot be accessed directly from HAL cluster

- SABA/LUNA/LUX:
To copy files from other clusters, first `ssh -A` into the other cluster to forward your keys.

Command line

```
ssh -A $USERNAME@$CLUSTER
```

We recommend `rsync -va` to copy files and directories.

Make note of the source directory/source files and destination directory/files on Lilac and copy them as below:

Command line

```
rsync -av --progress $SOURCEPATH lilac:$DESTPATH
```

- HAL:
Remember that the `hal` cluster is outside the MSKCC network, and does not have access to `lilac`.
First - Make note of the source directory/source files on HAL and destination directory/files on Lilac:
To transfer data, `ssh` into `lilac` as below :

Command line

```
ssh -A $USERNAME@lilac.mskcc.org
```

Then pull files from HAL:

Command line

```
rsync -av --progress hal:$SOURCEPATH $DESTPATH
```

- Depending on the size and number of files to copy, you may run multiple `rsync` commands simultaneously to copy different directories.
- The HPC private network is faster than the MSKCC campus network, so using short names (`lilac`, `saba`, `luna`, `selene`, etc.) will often make transfers faster than using fully qualified domain names such as `luna.mskcc.org`. This does not apply to `hal`, though