Technical guidelines for group data migration from Hutch IT to CSCS services

Data migration workflow



Group drive permissions

Due to cross-domain data migration, it is not possible to retain permissions when disabled inheritance has been applied. As such the following will apply:

There will be only one **security group** per **groupdrive** per **PI group** to be migrated.

In limited cases of **shared** groupdrives where users of different groups will need to collaborate with each other these will be treated as different groupdrives (**nested shares**) where both security groups from each group will be added

Users will need to **remove any confidential data** and/or **PID** as data migrating across will be accessible by the entire group (Please refer to the Data Storage and Management document for suggested locations to store this data to comply with University [guidelines](https://help.uis.cam.ac.uk/service/security/data-sec-classes)).

IFS drive sizing and data cleansing

The maximum size for any IFS share is 100TB (this includes snapshot reserve). As a result of this groups with large datasets will therefore need to categorise their data **into chunks of max 80TB** (This is to provide room for data growth and mitigate snapshot overflow).

CSCS request groups complete data cleansing and organisation to assist with the migration and complete the following:

* Groups to provide source path of their data (e.g [\\minerva\centre\GroupName](file:///%5C%5Cminerva%5Ccentre%5CGroupName) etc)
* Groups with large data sets (larger than 80tb total) need to follow below structure for data migrations.
	+ Group Name
		- Folder\_x (max 80tb)
		- Folder\_y (max 80tb)
		- Folder\_z (max 80tb)

SMB / NFS Shares and drive structuring

IFS does not support mixed mode protocol. All shares must be either CIFS (SMB) or NFS shares. CSCS recommends using **SMB(CIFS) shares** as these are compatible with both Windows, Linux and MAC and can be mounted via fstab (on Mac and Linux) in the same way as NFS.

SMB shares are compatible with **DFS namespace** ([\\medlan.cam.ac.uk\](file:///%5C%5Cmedlan.cam.ac.uk%5C)) we will use to centralise Groups shares so users can access everything from one path. Any NFS shares created would need individually mounting.

Example structure:

DFS structure of a nested configuration will be as below.

* DFS Namespace [\\medlan.cam.ac.uk\ed](file:///%5C%5Cmedlan.cam.ac.uk%5Ced)
	+ GroupName\_x
		- Folder\_a (IFS Share)
		- Folder\_b (IFS Share)
		- Folder\_g (IFS nested share)
	+ GroupName\_y
		- Folder\_c (IFS Share)
		- Folder\_d (IFS share
		- Folder\_g (IFS nested share)
	+ GroupName\_z