Migrating files from the HPSS

Meeting on December 4th: http://www.fin.ucar.edu/it/mms/v2/fl-1001-chat.html

Overview

Read only Jan 20, 2020

End of life Oct 1, 2021

Deadlines won't be extended.   Oracle is not manufacturing the tape drives we need.

87PB on HPSS. Only time to migrate 50% of data before EOL.

Migration Plan

1. Form Migration plan with PI
   a. What data to save?
   b. Where will it go?

You can run gladequota to look at your CS quotas (cheyenne).

https://www2.cisl.ucar.edu/user-support/allocations

https://www2.cisl.ucar.edu/user-support/allocations/ncar-allocations

Understand your HPSS data footprint

On cheyenne, run 'id' to list your groups (projects). (Can also look at sam.ucar.edu).

If you know what paths your files exist:

```
hsi ls -lRU /path >& ~/hsi.ls.path.txt
```

If you don't know where your files are. Go to support.ucar.edu. Submit ticket and ask for an HPSS inventory.

Organize files on HPSS

You can move files from the HPSS to the HPSS for free (i.e. hsi mkdir & hsi mv cost basically nothing).

Recommendation - create directories for migration plan:

`/home/user/project/campaign`

`/home/user/project/delete`

`/home/user/project/local_storage`

Then you can use recursive commands to delete, get, etc.

To move to Campaign Storage via data-access posix mount

Go to data-access node.

ssh data-access.ucar.edu

cd /glade/campaign/ral/

cd path/to/cs/dir

```
hsi cget -RA /home/user/project/campaign
```

-R Recursive

-A Tape Ordering (makes a huge difference - gets everything from one tape before moving to the next tape)

(This is also why it is useful to move everything into a single directory, so HSI can figure out the most efficient way to get data from the tapes).

(RUN hsi cget in tmux or screen or something so that it doesn't die on you).

https://www.hamvocke.com/blog/a-quick-and-easy-guide-to-tmux/

https://thoughtbot.com/blog/a-tmux-crash-course

To move to Campaign Storage via Globus

Advantage - globus resumes, and provides tracking.

Disadvantage - need to stage data locally via hsi, then globus to push to CS.

To move to external storage with no Globus

If you are putting data somewhere that is not a globus endpoint:

scp, rsync, bbcp (requires install on both ends - uses multiple streams, so faster)

Verify data migration was ok.

confirm that file sizes are as expected. (ls)

Remove data from HPSS

Remove data you don't need and data that you have transfered to CampaignStore

- Help CISL track migration process.
- Make sure you don't transfer the same files multiple times.

Verify first! Data removed from HPSS is not recoverable!

Best Practices

Only get files once - getting them multiple times is a waste of valuable tape time.

Record commands into log files so you can figure out what you did.

Use verbose mode when available

Putting data onto HPSS will slow your migration (concurrent transfer limit of 5 files on HPSS).

Questions

Dave Hart told me that he thought projects at NCAR could purchase additional data storage from CISL at the rate of $45/TB/year, but that he was not sure. Can you confirm that this data storage is available and let me know if the rate is correct. Can you also tell me anything more about this storage? i.e. what will the interface and medium be for the storage?

Send ticket to support.ucar.edu.