**Data Archival**

Surprisingly, data generated by ASCOT5 at NERSC is not automatically archived. **Long-term data archivel is the responsibility of the user at NERSC**. **Data stored in the production directories, e.g. project/projectdirs/m3195/ascot/ascot\_run\_output is subject to being erased without warning.**

Files are archived to the HPSS long-term storage using the commands ‘his’ and ‘htar’. I have a shell script my\_htar.sh in my NERSC home directory that invokes the htar utility with appropriate switches.

sscott@cori04:~> cat my\_htar.sh

#

# $1 is the name of the tar file, e.g. july23.tar

#

# $2 is a filename containing a list of directories and/or files, one entry per line

#

echo " "

echo " I will htar the following directories and files "

cat $2

echo " "

echo " I will save this data into file: ", $1

echo " "

echo " I will pause 5 seconds before starting the htar ..."

echo " "

sleep 5

htar -cvf -Hnostage $1 -L $2

Invoking that shell script will cause the text generated by the script to be routed to the screen, which is helpful to know whether there are any error messages. But we could add a > my\_tar.log to the last line in the script, which would then provide a log of the htar for all posterity.

Some useful links for long-term data archival:

* <https://docs.nersc.gov/filesystems/archive/>
* <https://www.nersc.gov/assets/pubs_presos/archive-best-practices-JFUF-final.pdf>

NERSC recommends that you try to store between 100 GB and 2 TB each time you invoke htar.

I recently archived output files that were generated between July 28 and December 22. The total data transferred was 528 GB for ~3700 files; the wall clock time to do the archive was ~8 minutes. The typical size of my my hdf5 files is 1-3 GB, with a few extending to 5 GB. Depending on the frequency of running simulations, I would recommend that we do an archive every 1-2 months.