

E3SM simulations generate large amounts of data that need be archived on HPSS (High Performance Storage System). For optimal performance, storage on HPSS should consist of a relatively small number of large files. Therefore, it is not possible to directly archive individual E3SM model output files on HPSS.

zstash is a python command line utility developed to serve E3SM long-term archiving needs. With zstash, files are archived into standard tar files with a user specified maximum size. Tar files are created locally, then transferred to HPSS. For improved performance, md5 checksums of input files are computed on-the-fly during archiving. Checksums and additional metadata are stored in a database. File integrity is verified by computing checksums on-the-fly during extraction.

zstash v0.4.2 has the ability to be used on machines without HPSS and the ability for the user to specify a path for the local archive (cache).

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.