

If you run an application twice on the same machine will it produce identical HDF5 files?

Will netCDF4 make bit-for-bit (BFB) reproducible files? In other words will running a deterministic model twice on the same machine (without changing compiler, netCDF library, etc.) produce identical output files?

I thought that netCDF4 would, like netCDF3, use deterministic algorithms without any date-stamps within the file, and thus produce BFB files. Apparently I am wrong.

To determine whether files were BFB I checked their SHA1 sums. netCDF3 produces BFB files and netCDF4 does not. BFB output files make models easier to debug, so it would be helpful if netCDF4 continued the netCDF3 BFB "tradition". Is this possible? Am I missing something? Is HDF5 the culprit?

Answer:

If you turn off the create/modify/access time tracking for objects created (with the `H5Pset_obj_track_times()` routine), everything should be bit-for-bit reproducible. Coincidentally, it makes accessing those objects faster and the size of their metadata smaller also. You do lose the ability to know when the object was created/modified/accessed.