Contribution ID: **159**   Type: **Talk**

Stitch It Up: Using Progressive Data Storage to Scale Science

Generally, scientific simulations load the entire simulation domain into memory because most, if not all, of the data changes with each timestep. This has driven application structures that have, in turn, affected the design of popular IO libraries, such as HDF-5, ADIOS, and NetCDF. All of these libraries assume that each output written will be a complete simulation domain. While a time dimension may be an "unlimited" dimension in NetCDF, the size of an array variable is fixed for the entire file. While for many cases, this assumption makes sense, there is a significant collection of simulations where this approach results in vast swaths of unchanged data written each timestep.

Some prior work has looked at compressing the output by only storing a difference. This reduces data size, but does not reduce the compute requirement. This paper explores a two-pronged approach to addressing the computation and the data sizes. First, an out-of-core-like computational framework is developed enabling simulation across a large domain by only computing over the part that is affected by the simulation. Second, and the primary focus of this paper, is a new IO approach that is capable of stitching together a coherent global view of the total simulation space at any given time. This benefit is achieved with no performance penalty compared to running with the full data set in memory, at a radically smaller process requirement, and a radical data reduction with no loss in fidelity. Additionally, the structures employed offer significant additional capabilities for simulation monitoring and easy data analytics.

Supported Student

Primary author: LOFSTEAD, Jay (Sandia National Laboratories)

Presenter: LOFSTEAD, Jay (Sandia National Laboratories)

Session Classification: HPC Technology

Track Classification: Storage and IO