ExaHDF5: An I/O Platform for Exascale Data Models, Analysis and Performance

Prabhat (PI), John Wu, Jerry Chuo, Mark Howison,
E. Wes Bethel, *LBNL*
Quincey Koziol, *The HDF Group*
Bruce Palmer, Karen Schuchardt, *PNNL*

Overview

It is reasonably well accepted that one of the primary bottlenecks in modern computational and experimental sciences is coping with the sheer volume and complexity of data. Storing, reading, finding, analyzing, and sharing data are tasks common across virtually all areas of science, yet advances in data management infrastructure, particularly I/O, have not kept pace with our ability to collect and produce scientific data. This "impedance mismatch" between our ability to produce and store/analyze data continues to grow and could, if not addressed, lead to situations where science experiments are simply not conducted or scientific data not analyzed for want of the ability to overcome data-related challenges.

Our project consists of three thrust areas that address the challenges of data size and complexity on current and future computational platforms:

- We are extending the scalability of I/O middleware to make effective use of current and future computational platforms.
- We are incorporating advanced index/query technology to accelerate operations common to scientific data analysis.
- We are building upon our existing work on data model APIs that simplify simulation and analysis code development by encapsulating the complexity of parallel I/O.

We are conducting these activities in close collaboration with specific DOE science code teams to ensure the new capabilities are responsive to scientists' needs and are usable in production environments. Our approach includes a clear path for maintainability and production release.

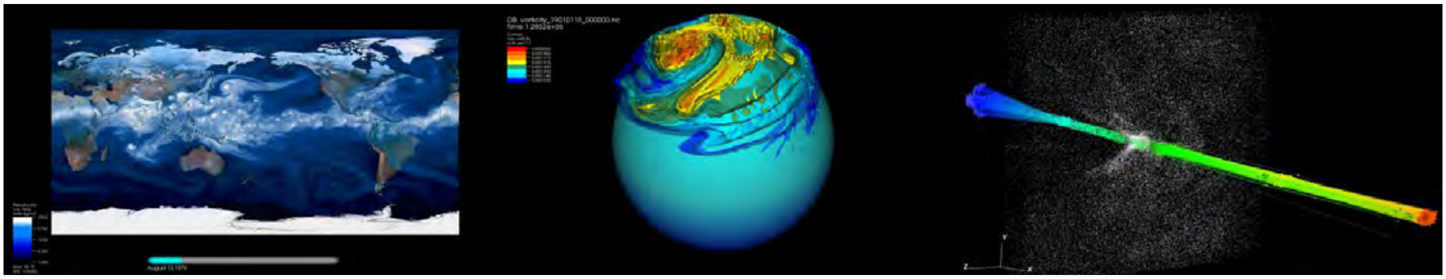
Technical Objectives

HDF5 Scalability: We are implementing a number of significant optimizations and improvements to the HDF5 library to enable Exascale I/O performance.

- Streamlining HDF5 metadata modifications by removing collective I/O operations.
- Improved speed and scalability of append-only modifications to HDF5 files through file format and algorithmic changes.
- Supporting asynchronous I/O operations.
- Auto-tuning HDF5's behavior for the underlying file system.
- Incorporating state-of-the-art indexing solutions for data stored in HDF5 files.
- Adding fault tolerance to HDF5 by supporting new MPI extensions, as well as incorporating a new file update mechanism that makes the library robust to application failure.

Extreme scale analysis: As dataset sizes increase, it becomes infeasible for scientists to load and examine entire datasets. Typically, they expect an interactive system that allows them to specify, and refine, complex criteria for regions or elements of interest in the dataset. Conventional approaches based on loading entire datasets at each step of the refinement/display loop perform poorly in such a context. Building on our success with FastBit and HDF5-FastQuery, we are using state-of-the-art index/query techniques within VisIt to accelerate analysis of extremely large datasets.

Exascale Data Models: We are developing data models in the context of three applications areas that are of strategic importance to DOE's mission: Climate Modeling, Groundwater Modeling and Accelerator Modeling. For Groundwater and Accelerator Modeling, we are leveraging our expertise with the H5Part and H5Block data models for handling particle and block structured data. For Climate Modeling, we are building upon our existing work for handling geodesic grids for Global Cloud Resolving Models. We expect this work to play a critical role in terms of presenting real scientific applications with a high-performance as well as high-productivity interface for doing parallel I/O.



Science Collaborators

Global Cloud Resolving Models (*Dave Randall, Colorado State University*). The largest source of uncertainty in climate models can be attributed to modeling of clouds. To simulate climate at high resolutions (<10 km), and scale to hundreds of thousands of processors, climate models are moving to new grid representations such as the geodesic and cubed-sphere grids. We are addressing a number of challenges resulting from the high-resolution application of the geodesic grid: developing a parallel I/O infrastructure to efficiently write large volumes of data (0.5 TB per snapshot), development of APIs to simplify integration with GCRM code, and efficiently reading and analyzing the data.

Community Climate System Model (*Bill Collins & Michael Wehner, LBNL, Mariana Vertenstein, UCAR*). CCSM is comprised of models for the atmosphere, ocean, land, and sea ice. The next generation of CCSM simulations will run at 25 km spatial resolution and a 10 minute temporal resolution. The total integration period can span anywhere from centuries to millennia. We are working with the CCSM team to tackle I/O challenges resulting from these massive simulation runs. We are optimizing the NetCDF-4/HDF5 PIO layer in the CCSM code enable large scale runs on systems such as hopper and jaguar.

Groundwater Modeling (*Tim Scheibe, PNNL*). Simulations of sub- surface flows are moving to higher resolutions and incorporating more sophisticated physics and chemistry. Because of the large uncertainties associated with parameterizing the subsurface, there is also increasing interest in running ensembles of simulations that can be used either to estimate the distribution of outcomes or in inverse modeling that can provide estimates of parameters that best match measured behavior. All of these trends are increasing the volume of data produced by these calculations. Particle based simulations using Smoothed Particle Hydrodynamics (SPH) will soon use billions of particles. In this project, we are working on incorporating HDF5 based data models into the SPH code. We are also profiling and optimizing parallel I/O for these codes to enable them to run at massive concurrency.

Accelerator Physics (*Robert Ryne, LBNL*). Particle accelerators are extremely important instruments for scientific research and discovery. Previously accelerators were seen mainly as tools for research in medium- and high-energy physics. Now, in addition, the Nation's light sources and neutron sources are seen as essential tools for research in materials science, chemistry, and the biosciences. In this project, we are working with the IMPACT/MaryLie suite of codes. We are working on integrating the H5hut data model, optimizing parallel I/O on large scale runs, and streamlining visualization and analysis capabilities for particle physics researchers.

Contact

For more information on the project, please contact Prabhat <prabhat@lbl.gov>