

Developing implicitly-parallel Python analysis tools for NOvA

Thursday 15 September 2022 18:00 (10 minutes)

The NOvA collaboration together with a Dept. of Energy ASCR supported SciDAC-4 project, have been exploring Python-based analysis workflows for HPC platforms. This research has been focused on adapting machine-learning application workflows using highly-parallel computing environments for neutrino-nucleon cross section measurements. This work accelerates scientific analysis and lowers the learning barriers required to leverage leadership computing platforms.

Users of these HPC workflows have often been required to have significant experience with parallel computing libraries and principles in addition to dedicated access to accounts and resource allocations at off-site HPC/Supercomputing centers such as NERSC and the Argonne Leadership Computing Facility. With the commissioning of the new Fermilab analysis cluster, which will provide dynamically provisioned pools of HPC resources, we are now exploring ways to improve the efficiency and approachability of Python-based analysis tools. These include data organizations using HDF5 and the PandAna analysis framework, both of which natively support highly data-parallel operations.

Our research enables fully data-parallel exploration, selection, and aggregation of neutrino data, which are the fundamental operations required for neutrino cross section analysis work in NOvA. These operations are executed with the analysis cluster through Jupyter notebook interfaces and have been demonstrated to achieve low execution latencies, which are highly compatible with interactive analysis time-scale(s). We have developed a method for constructing large monolithic HDF5 based files, which each represent an entire NOvA dataset, and we have demonstrated a factor of more than $10\times$ speedup of basic event selection using this data representation, relative to equivalent multi-file composite representations of the datasets. We have developed a complete, implicitly-parallel analysis workflow with basic histogram operations and demonstrated its scalability using a realistic neutrino cross section measurement on the Perlmutter system at NERSC. These tools will enable real-time turnaround of more physics results for the NOvA collaboration and wider HEP community.

Author: DOYLE, Derek (Colorado State University)

Co-authors: NORMAN, Andrew; KOWALKOWSKI, Jim (Fermilab); PATERNO, Marc; GROH, Micah (Fermi National Accelerator Laboratory); SEHRISH, Saba (Fermilab)

Presenter: DOYLE, Derek (Colorado State University)

Session Classification: Plenary Session Thursday