Contribution ID: **4**   Type: **20-minute presentation + 5-minute Q&A**

Accelerating Data Compression in HDF5 through Parallel Filter Processing

Wednesday 28 May 2025 09:00 (25 minutes)

Modern science and engineering creates and accumulates huge amounts of data which is persisted through tools like HDF5 in order to be available for further analysis, display and many other operations. Increasing efficiency in this data processing is critical for nowadays growing data quantities, not only for saving time, but also to efficiently use available resources.

This thesis aimed to provide a working prototype and analysis on parallel application of data filters within the HDF5 environment with special emphasis on HDF5 registered filters, such as LZ4 compression.

This prototype is embedded into the HDF5 framework, can be freely accessed as any other provided function and is available ready to use after building the project, while maintaining standard library behavior for all other use cases. Generally implementation is based on an all purpose thread pool with variable functionality based on registered callback function and can therefore be used in later versions of development. Analysis is based on comparison of standard library performance and prototype, based on identical example datasets, as well as statical analysis of provided program code. Furthermore CPU utilization and I/O performance are evaluated.

Results suggest a great potential for a fully implemented design including most capabilities of the stock HDF5 library. Near full CPU utilisation is shown with little to no wait for I/O completion and therefore cutting down runtime by quite an extensive amount.

This prototype, related work and analysis show the great improvement possible by adjusting the existing framework to a multithreaded solution while still maintaining full standard behaviour.

May we record your session?

Yes

Primary author: NEU, Frederick

Presenter: NEU, Frederick