

Seamless Integration of Blosc2 and HDF5 for High-Performance Data Compression

Monday 26 May 2025 10:35 (25 minutes)

HDF5 is the de facto standard for storing large volumes of binary data in files. Blosc2, an award-winning high-performance library, excels at compressing binary data in memory. Both are widely used, making their integration natural. This talk will cover using Blosc2 as an HDF5 filter and HDF5 as a Blosc2 backend.

We will outline the current state of the Blosc2 plugin for HDF5 (<https://github.com/Blosc/HDF5-Blosc2>) and provide instructions on its usage. Additionally, we will introduce the b2h5py package (<https://github.com/Blosc/b2h5py>), which bypasses the slow HDF5 filter pipeline to achieve high performance. Sparse datasets will be used to demonstrate Blosc2's performance on HDF5.

Finally, we will discuss various codecs available in Blosc2, with a focus on the Grok codec (<https://github.com/GrokImageCompression/grok>) which efficiently compresses data in the JPEG2000 format. We will also touch on the enhancements made to BTune (<https://ironarray.io/btune>), a Blosc2 plugin, to support lossy compression and automatically select the best codec/filter.

This work has been carried out as part of the LEAPS-INNOV program (<https://leaps-innov.eu/>), which strives to build a European ecosystem for photon sciences. The integration of Blosc2 and HDF5 ensures efficient storage and retrieval of large datasets, a critical factor for the program's success.

Compression #HighPerformance #HDF5 #Blosc2 #JPEG200 #LEAPS-INNOV

May we record your session?

Yes

Primary author: ALTED, Francesc (Blosc project)

Presenter: ALTED, Francesc (Blosc project)