Blosc is a high performance compressor optimized for binary data. It has been designed to transmit data to the processor cache faster than the traditional, non-compressed, direct memory fetch approach via a memcpy() OS call.

Blosc is the first compressor that is meant not only to reduce the size of large datasets on-disk or in-memory, but also to accelerate memory-bound computations (which is typical in vector-vector operations).

USE CASE

PyTables is a package for managing hierarchical datasets and designed to efficiently and easily cope with extremely large amounts of data. PyTables is built on top of the HDF5 library, using the Python language and it uses Blosc for enhanced I/O performance.

This package provides an interface to the HDF5 library for the Julia language. It uses Blosc as the default compressor for data I/O.

Zarr is a format for the storage of chunked, compressed, N-dimensional arrays. It uses Blosc as its default compressor for accelerated I/O to memory/disk or network.

PLANNED FEATURES

+ Plugin capabilities for allowing users to add more filters and codecs. There should also be a plugin register capability so that the info about the new filters and codecs can be persistent and propagated to different machines.

+ Lock support for super-chunks: when different processes are accessing concurrently to super-chunks, make them to sync properly by using locks, either on-disk (frame-backed super-chunks), or in-memory. Such a lock support would be configured in build time, so it could be disabled with a cmake flag.

+ Improve the Blosc website: create a nice, modern-looking and easy to navigate website so that new potential users can see at first glimpse what's Blosc all about and power-users can access the documentation part easily. Ideally, a site-only search box would be great.

PROJECT NEEDS

Support for network storage	150 Hours
Add a multidimensional compression codec to Caterva/cat4py	200 Hours
Checksums for chunk/index/metalayer.	150 Hours

Blosc is a Sponsored Project of NumFOCUS, a US 501(c)(3) public charity.

NumFOCUS Sponsored Projects rely on the generous support of corporate sponsors, institutional partners, and individual donors.

For more information on Blosc!, including our governance structure and project roadmap, please visit

http://www.blosc.org/

For more information:
info@numfocus.org | +1 (512) 831-2870.