Use of Btune for finding best codecs/filters for Blosc2

Francesc Alted / @FrancescAlted

Marta Iborra / @Marta_Iborra4

The Blosc Development Team / @Blosc2

2023 European HDF User Group (HUG) plugins and data compression summit

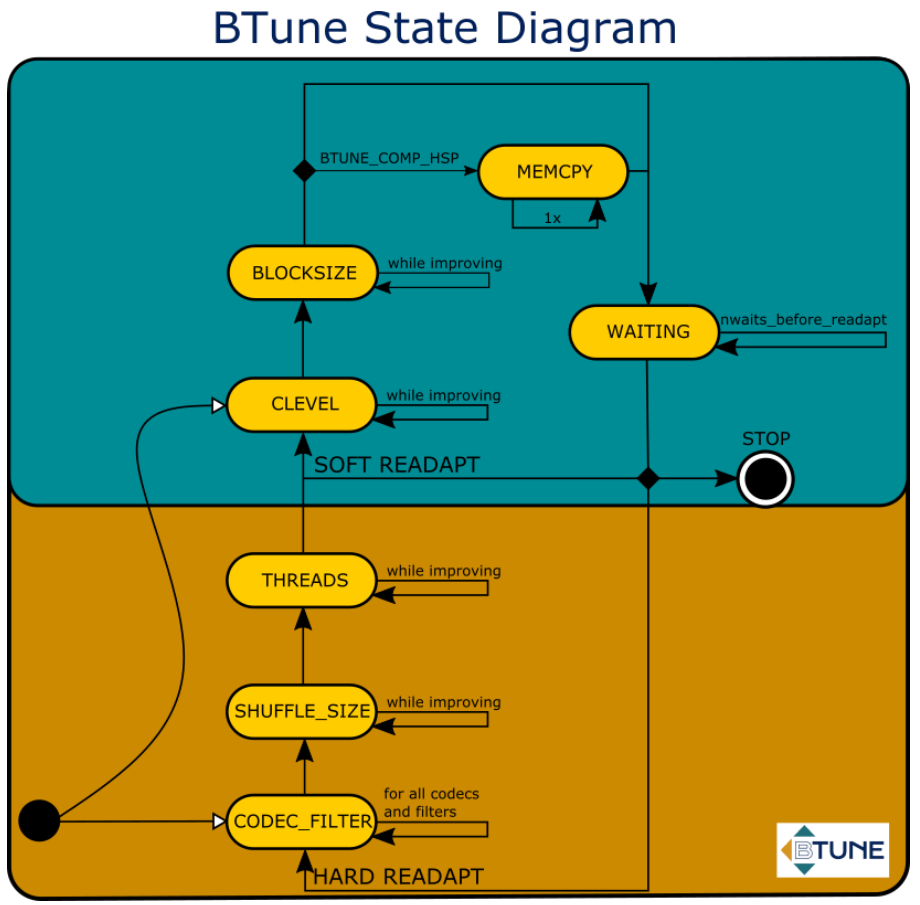
September 19th 2023

Fine Tuning Performance with BTune

https://btune.blosc.org

- BTune can fine tune the different parameters of the underlying Blosc2 storage to perform as best as possible.

- Can be trained to find the best codec & filter with **deep learning**.



Requeriments for the tutorial

- Laptop (or a remote machine, as you wish)
- Operating systems supported
  - Linux
  - MacOS
  - Windows: only via WSL. Please install it prior the tutorial; instructions here: https://learn.microsoft.com/en-us/windows/wsl/install
- Pyenv/Conda/mamba environment with a Python 3.10 or 3.11 installed
- Advise to do groups of 3

Different tiers of support for Btune

- **Genetic (Btune Free):** test different combinations by brute force

- **Trained (Btune Models):** Blosc team train datasets for you:
  - More accurate predictions for all chunks (specially first ones!)

- **Fully managed (Btune Studio):** you can train on your own

For the tutorial, follow the instructions:

https://github.com/Blosc/Btune-tutorial#btune-tutorial

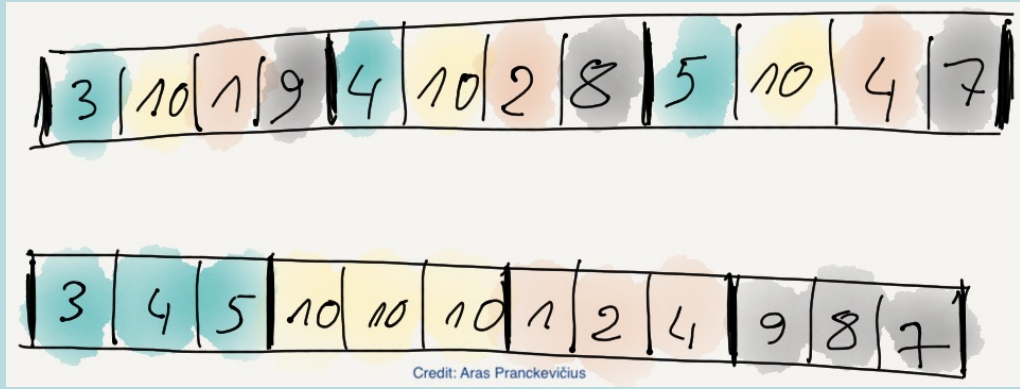
Bytedelta

A new filter for Blosc2

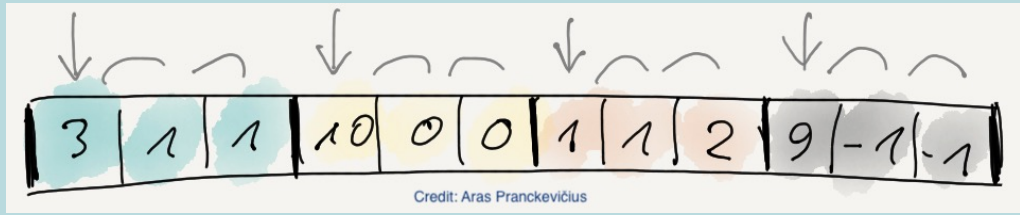
Bytedelta: How it works

Blosc2 Filter Pipeline

shuffle ->



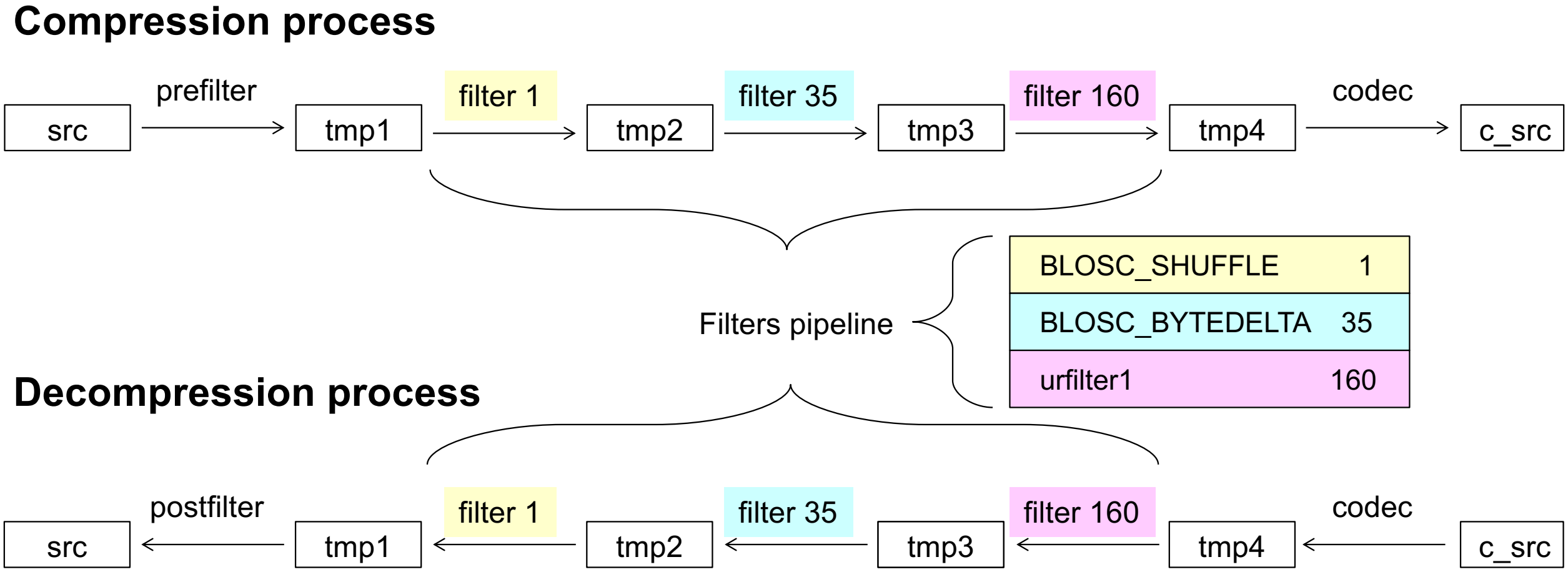
bytedelta ->



Based on initial work by Aras Pranckevičius

https://www.blosc.org/posts/bytedelta-enhance-compression-toolset/

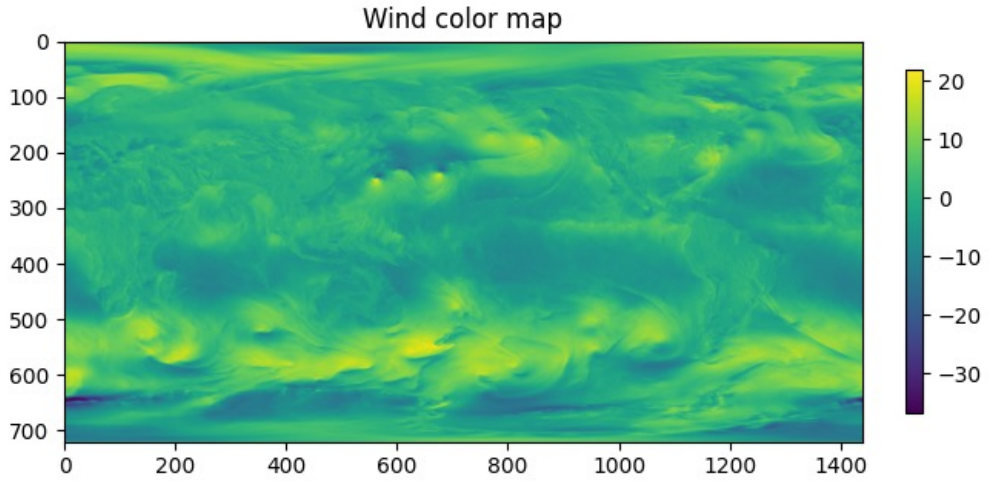
Filter pipeline: composing filters + codecs

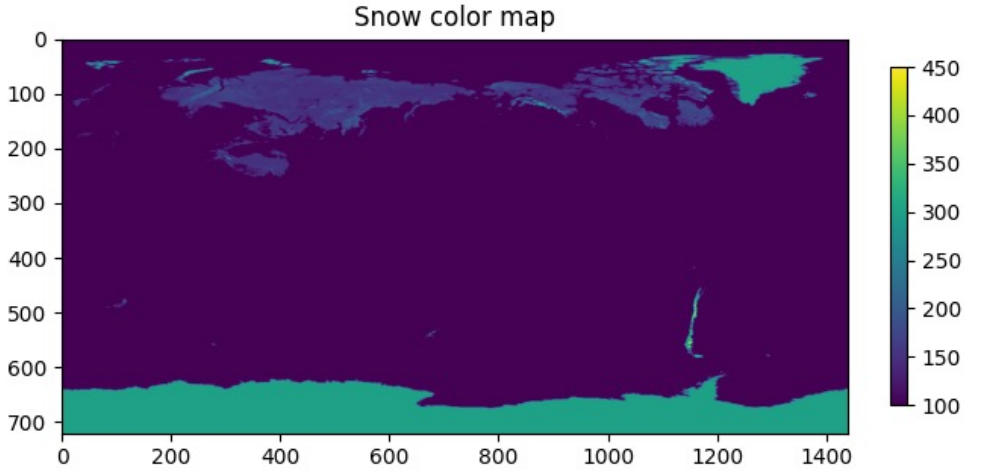


Bytedelta: How does it perform?

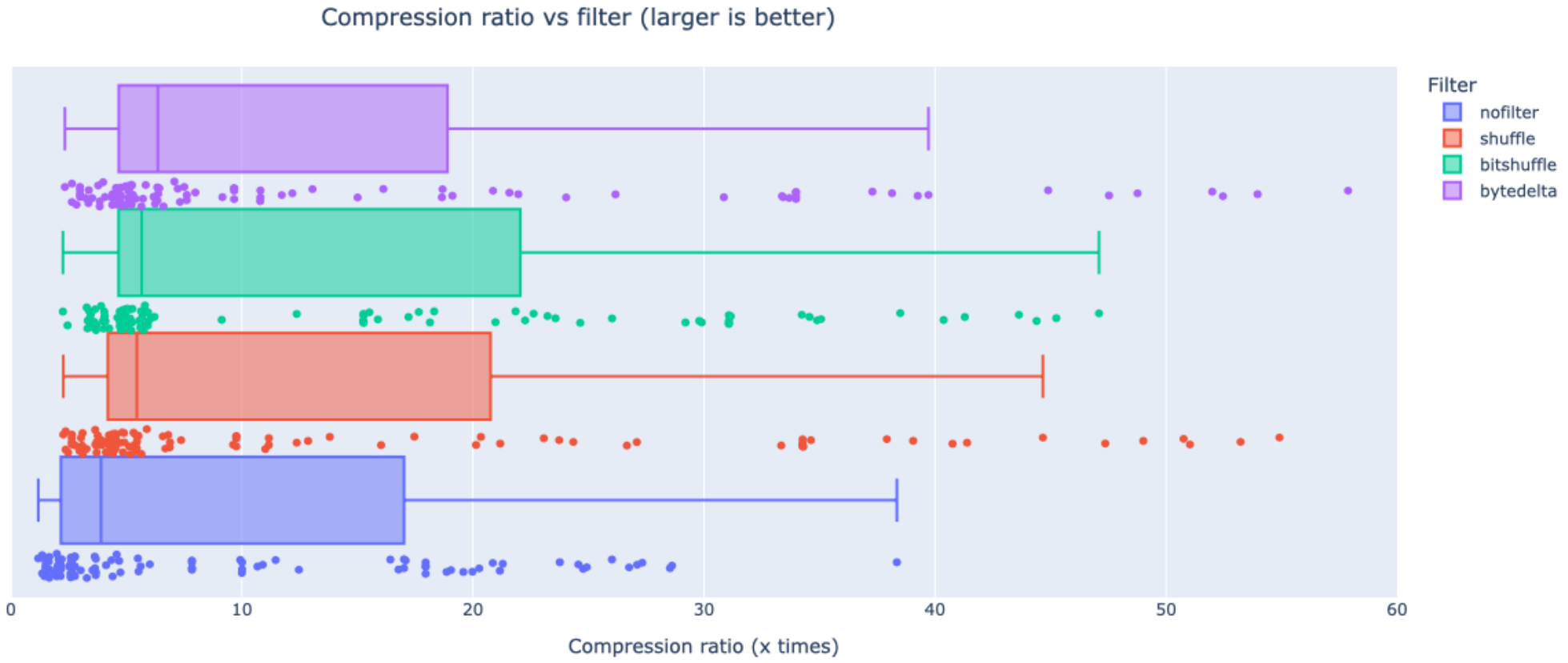
Tested on 5 different ERA5 datasets
(atmospheric reanalysis of the global climate):
wind, snow, flux, precip and pressure

- Some show some complex structure (wind)
- Others are simpler (snow)





Bytedelta: How does it perform?



Bytedelta compress better than shuffle or bitshuffle on average
Median for bytedelta (best): 6.36 x
Median for bitshuffle (second best): 5.66 x