Contribution ID: **122**   Type: **Poster**

Easy columnar file conversions with "odapt"

Thursday 14 March 2024 16:10 (30 minutes)

When working with columnar data file formats, it is easy for users to devote too much time to file manipulation. With Python, each file conversion requires multiple lines of code and the use of multiple I/O packages. Some conversions are a bit tricky if the user isn't very familiar with certain formats, or if they need to work with data in smaller batches for memory management. To try and address this issue, we are developing Python package 'odapt.' This package allows users to convert files with just one function call, with automatic memory management, compression settings, and other features added based on user feedback. Some such features include merging ROOT files (hadd-like), adding and dropping branches or TTrees from ROOT files. Odapt uses reliable columnar I/O packages h5py, Uproot, Awkward, and dask-awkward.

Significance

Though the project is still in development, we have gotten a lot of interest and feature-requests from users who frequently need to do columnar file conversions.

References

Experiment context, if any

Converting large files between different columnar formats.

Primary author: BILODEAU, Zoë (Princeton University (US))

Presenter: BILODEAU, Zoë (Princeton University (US))

Session Classification: Poster session with coffee break

Track Classification: Track 2: Data Analysis - Algorithms and Tools