Contribution ID: **306**   Type: **Poster**

Custom event sample augmentations for ATLAS analysis data

Monday 24 October 2022 16:10 (30 minutes)

High Energy Physics (HEP) has been using column-wise data stored in synchronized containers, such as most prominently ROOT's TTree, for decades. These containers have proven to be very powerful as they combine row-wise association capabilities needed by most HEP event processing frameworks (e.g. Athena) with column-wise storage, which typically results in better compression and more efficient support for many analysis use-cases. The downside, however, is that all events (rows) need to contain the same attributes and therefore extending the list of items to be stored, even if needed only for a subsample of events, can be costly in storage and lead to data duplication.

The ATLAS experiment has developed navigational infrastructure to allow storing custom data extensions for subsample of events in separate, but synchronized containers. These extensions can easily be added to ATLAS standard data products (such as DAOD-PHYS or PHYSLITE) avoiding duplication of those core data products, while limiting their size increase. As a proof of principle, a prototype based on the Long Lived Particle search is implemented. Preliminary results concerning the event-size as well as reading/writing performance implications associated with this prototype will be presented.

Augmented data as described above are stored within the same file as the core data. Storing them in dedicated files will be investigated in future, as this could provide more flexibility to store augmentations separate from core data, e.g. certain sites may only want a subset of several augmentations or augmentations can be archived to disk once their analysis is complete.

Experiment context, if any

ATLAS

References

Significance

Derived data is a main consumer of storage resources (for ATLAS in Run 2, derived AOD occupied >30% of disk). The capability of custom augmentation will reduce duplication and reduce storage costs.

Author: VAN GEMMEREN, Peter (Argonne National Laboratory (US))

Co-authors: METE, Alaettin Serhan (Argonne National Laboratory (US)); BURZYNSKI, Jackson Carl (Simon Fraser University (CA)); CATMORE, James (University of Oslo (NO)); HEINRICH, Lukas Alexander (CERN); NOWAK, Marcin Jerzy (Brookhaven National Laboratory (US)); KRUMNACK, Nils Erik (Iowa State University (US))

Presenter: HEINRICH, Lukas Alexander (CERN)