PUNCHLunch –

the online lunch seminar of the PUNCH4NFDI Consortium
Every third Thursday, 12:30-13:30 hrs: keynote and discussion

_____

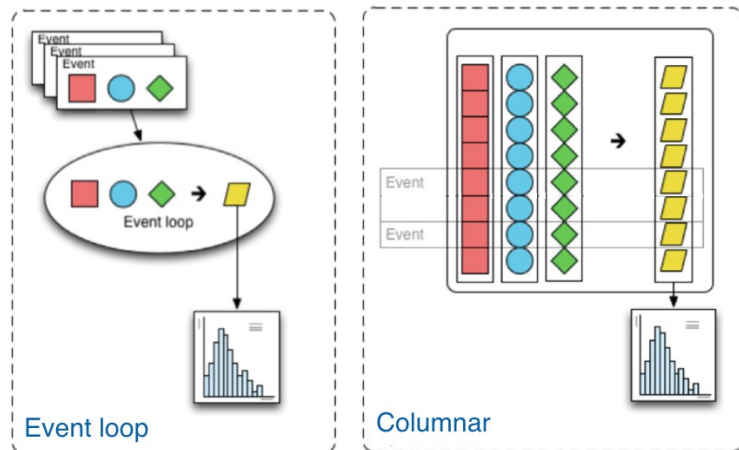Thursday, February 15 2024, 12:30 hrs:
Dr. Nikolai Hartmann (LMU Munich)
Columnar data analysis in HEP
challenges and prospects

Abstract:

The rise of machine learning techniques in recent years has also brought the concept of Array Programming back into mainstream data analysis. Thinking in terms of operations on whole arrays instead of individual quantities that are looped over allows a separation of high level steering code from the low level number crunching. This enabled high level languages like python to become the main platform for analyzing large amounts of data. In HEP, the array-at-a-time paradigm, often termed "columnar data analysis" amounts to move away from writing code for event-by-event processing to a per-column/per-object-list description. Packages and code developed in the HEP community like Uproot, Awkward Array and ROOT's RDataFrame provide the necessary additions for HEP analysis to be viable. To get the maximum performance out of this analysis technique, data reading has to be organized with columnar analysis in mind. While the ROOT TTree format is a columnar storage, it has limitations that could lead to suboptimal performance in certain use cases. The ROOT RNtuple format provides a promising future improvement on this as well as data formats like Apache Parquet that are also used outside the HEP community.

This presentation will give an overview of the current status of tools, as well as some experiences with ATLAS DAOD_PHYSLITE format which is intended to be a format usable for columnar data analysis in the HL-LHC era.

Connection details:

ZOOM Meeting "PUNCHLunch seminar": *https://indico.desy.de/event/XXXXX/*
Webinar ID: 919 1665 4877, passcode: 481572
Next event: to be announced

_____

Connect to PUNCH4NFDI:
Mastodon: https://nfdi.social@punch4nfdi   Mail: punch4nfdi@desy.de   Web: www.punch4nfdi.de