



Contribution ID: 71

Type: Poster

Declarative paradigms for analysis description and implementation

Wednesday, 13 March 2024 16:15 (30 minutes)

The software toolbox used for “big data” analysis in the last few years is changing fast. The adoption of approaches able to exploit the new hardware architectures plays a pivotal role in boosting data processing speed, resources optimisation, analysis portability and analysis preservation.

The scientific collaborations in the field of High Energy Physics (e.g. the LHC experiments, the next-generation neutrino experiments, and many more) devote increasing resources to the development and implementation of bleeding-edge software technologies, pushing the reach of the single experiment and the whole HEP community.

The introduction of declarative paradigms in the analysis description and implementation is gaining interest and support in the main collaborations. This approach can simplify and speed-up the analysis description phase, support the portability of an analysis among different datasets/experiments, and strengthen the preservation of the results.

Furthermore, this approach - providing a deep decoupling between the analysis algorithm and back-end implementation - is a key element for present and future processing speed.

In the panorama of the approaches currently under study, an activity is ongoing in the ICSC (Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing, Italy) which focuses on the development of a framework characterized by the use of a declarative paradigm for the analysis description and the ability to operate on datasets from different experiments.

Using as a building base for a demonstrator the NAIL (Natural Analysis Implementation Language) Python package (developed in the context of the CMS data analysis for the event processing), the activity focuses both on the development of a general and effective interface able to support the data format of different experiments, and on the extension of the declarative approach to the full analysis chain.

Significance

The application of declarative paradigms to data analysis has been an active field of development in the last decade. The presented developments focus on aspects not yet fully explored by previous studies/demonstrators: configurable input interface (i.e. access to multiple experiments) and the definition of the full analysis chain (not only the “event-loop”).

References

https://indico.cern.ch/event/769263/contributions/3413006/attachments/1840145/3016759/NAIL_Project_Natural_Analysis_Implementat

Experiment context, if any

Authors: CMS, ATLAS - target application: HEP main collaborations (e.g. LHC experiments)

Primary authors: ANNOVI, Alberto (INFN Sezione di Pisa); RIZZI, Andrea (Universita & INFN Pisa (IT)); VASELLI, Francesco (Scuola Normale Superiore & INFN Pisa (IT)); MASTRANDREA, Paolo (Universita & INFN Pisa (IT)); Dr BOCCALI, Tommaso (INFN Sezione di Pisa)

Presenter: VASELLI, Francesco (Scuola Normale Superiore & INFN Pisa (IT))

Session Classification: Poster session with coffee break

Track Classification: Track 1: Computing Technology for Physics Research