



Contribution ID: 393

Type: Talk

## Declarative paradigms for data analysis description and implementation

Monday 21 October 2024 14:24 (18 minutes)

The software toolbox used for “big data” analysis in the last few years is rapidly changing. The adoption of software design approaches able to exploit the new hardware architectures and improve code expressiveness 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) are devoting increasing resources to the development and implementation of bleeding-edge software technologies in order to cope effectively with always growing data samples, pushing the reach of the single experiment and of the whole HEP community.

The introduction of declarative paradigms in the analysis description and implementation is growing interest and support in the main collaborations. This approach can simplify and speed-up the analysis description phase, support the portability of the analyses among different datasets/experiments and strengthen the preservation and reproducibility 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, potentially even with back-ends not existing today.

In the landscape 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 characterised by a declarative paradigm for the analysis description and able to operate on datasets from different experiments.

The existing NAIL (Natural Analysis Implementation Language [1]) Python package, developed in the context of the CMS data analysis for the event processing, is used as a building base for the development of a demonstrator able to provide a general and effective interface characterised by a declarative paradigm and targeted to the description and implementation of a full analysis chain for HEP data, with support for different data formats.

Status and development plan of the demonstrator will be discussed.

[1][https://indico.cern.ch/event/769263/contributions/3413006/attachments/1840145/3016759/NAIL\\_Project\\_Natural\\_Analysis\\_Implement](https://indico.cern.ch/event/769263/contributions/3413006/attachments/1840145/3016759/NAIL_Project_Natural_Analysis_Implement)

**Primary authors:** ANNOVI, Alberto (INFN Sezione di Pisa); RIZZI, Andrea (Universita & INFN Pisa (IT)); MASTRANDREA, Paolo (Universita & INFN Pisa (IT)); Dr BOCCALI, Tommaso (INFN Sezione di Pisa)

**Presenter:** MASTRANDREA, Paolo (Universita & INFN Pisa (IT))

**Session Classification:** Parallel (Track 5)

**Track Classification:** Track 5 - Simulation and analysis tools