

Data Analysis Integrated Software System in IHEP, design and implementation

Tuesday, March 22, 2022 3:30 PM (30 minutes)

Large scale research facilities are becoming prevalent in the modern scientific landscape. One of these facilities' primary responsibilities is to make sure that users can process and analyse measurement data for publication. In order to allow for barrier-less access to those highly complex experiments, almost all beamlines require fast feedback capable of manipulating and visualising data online to offer convenience for the decision process of the experimental strategy. And recently, the advent of beamlines at fourth-generation synchrotron sources and high resolution with high sample rate detector has made significant progress that pushes the demand for computing resources to the edge of current workstation capabilities. On top of this, most synchrotron light sources have shifted to prolonged remote operation because of the outbreak of a global pandemic, with the need for remote access to the online instrumental system during the operation. Another issue is the vast data volume produced by specific experiments makes it difficult for users to create local data copies. In this case, on-site data analysis services are necessary both during and after experiments.

Some state-of-the-art experimental techniques, such as phase-contrast tomography and ptychography approaches, will be deployed. However, it poses a critical problem of integrating this algorithmic development into a novel computing environment used in the experimental workflow. The solution requires collaboration with the user research groups, instrument scientists and computational scientists. A unified software platform that provides an integrated working environment with generic functional modules and services is necessary to meet these requirements. Scientists can work on their ideas, implement the prototype and check the results following some conventions without dealing with the technical details and the migration between different HPC environments. Thus, one of the vital considerations is integrating extensions into the software in a flexible and configurable way. Another challenge resides in the interactions between instrumental sub-systems, such as control system, data acquisition system, computing infrastructures, data management system, data storage system and so on, which can be quite complicated.

In this paper, we propose a platform for integration and automation across services and tools, which ties together existing computing infrastructure and state-of-the-art algorithms. With modular architecture, it comprises loosely coupled algorithm components that communicate over the heterogeneous in-memory data store, and scales horizontally to deliver automation at scale based on kubernetes. To produce and integrate into applications high-performance products embodying data analysis and visualization methods, the platform also has native PyQt GUIs, Web UIs based on JupyterLab, ipython CLI clients, and APIs over ZeroMQ.

Primary authors: HU, Yu (IHEP); Dr TIAN, Haolai (Institute of High Energy Physics, CAS); LI, Ling (Institute of High Energy Physics, CAS); QIU, XiaoMeng (Zhengzhou university)

Co-authors: Mr LIU, Zhibing (Institute of High Energy Physics, CAS); HU, Qingbao (IHEP); QI, Fazhi (Institute of High Energy Physics, CAS)

Presenter: Dr TIAN, Haolai (Institute of High Energy Physics, CAS)

Session Classification: VRE

Track Classification: Track 5: Virtual Research Environment (including tools, services, workflows, portals, ... etc.)