

Framework for Distributed Near Real-Time Data Processing Pipelines

Processing large amount of data in near real-time during experiments at synchrotrons is enabling scientists to make the best use of limited beamtime [1]. However, building systems capable of handling data rates of several gigabytes per second over long periods of time requires specialized expertise in distributed computing [2], which limits the broader adoption of such systems at beamlines.

The presented framework, designed and developed as part of the ROCK-IT project [3], aims to simplify the creation and operation of distributed near real-time data processing pipelines. Users of this framework will create data processing pipelines by assembling together existing data processing units (workers) in a way similar to existing Flow-Based Programming [4] or Workflow frameworks [5] [6] mainly used for batch processing. When needed, developers will have the possibility to develop their own workers, for example using the AsapoWorker library [7]. In addition, the framework will provide tools to deploy and manage these pipelines on HPC clusters, visualize data from different workers, and save the relevant data into standard file formats such as NeXus [8].

- [1] “Real-time data processing for serial crystallography experiments”, Thomas White et al., 2025, <https://doi.org/10.1107/S20522525240118>
- [2] “Eight Fallacies of Distributed Computing”, Gareth Wilson, 2015, <https://web.archive.org/web/20171107014323/http://blog.fogcreek.com/fallacies-of-distributed-computing-tech-talk/>.
- [3] “Remote, Operando Controlled, Knowledge-driven, and IT-based (ROCK-IT)”, <https://www.rock-it-project.de/>
- [4] “Flow-Based Programming 2nd Edition: A New Approach to Application Development”, J. Paul Morrison, 2011, <https://www.jpaulmorrison.com/fbp/book.html>
- [5] “Airflow is an open-source platform for developing, scheduling, and monitoring batch-oriented workflows”, <https://airflow.apache.org/docs/apache-airflow/stable/index.html>.
- [6] “Extensible Workflow System (Ewoks)”, <https://ewoks.esrf.fr/en/latest/>.
- [7] “AsapoWorker” library, <https://gitlab.desy.de/fs-sc/asapoworker>.
- [8] “The NeXus data format”, J. Appl. Cryst. (2015). 48, 301-305, <https://doi.org/10.1107/S1600576714027575>.

I want to give a Lightning Talk

no

Author: ANDREZ, Marc-Olivier (Deutsches Elektronen-Synchrotron (DESY))

Co-authors: TOLSTIKOVA, Aleksandra (Deutsches Elektronen-Synchrotron DESY); BARTY, Anton (Deutsches Elektronen-Synchrotron DESY); RUEDA, Diana (Deutsches Elektronen-Synchrotron DESY); KARNEVSKIY, Mikhail (Deutsches Elektronen-Synchrotron DESY); Dr WHITE, Thomas (Deutsches Elektronen-Synchrotron (DESY)); SCHOOF, Tim (Deutsches Elektronen-Synchrotron DESY); KARTIK, Vijay (DESY)

Presenter: ANDREZ, Marc-Olivier (Deutsches Elektronen-Synchrotron (DESY))

Session Classification: Poster Session