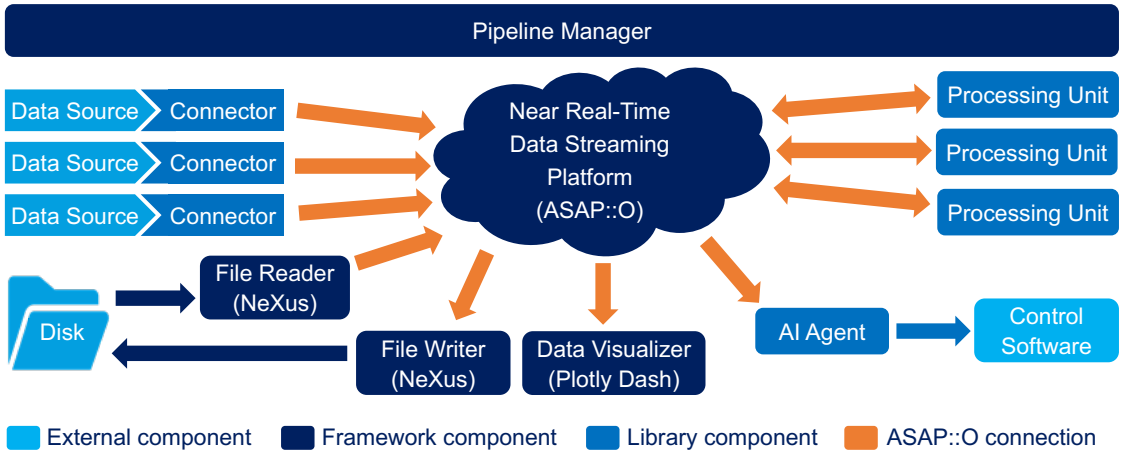
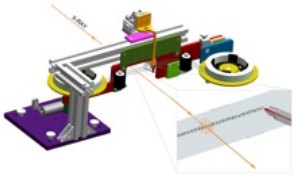
Framework for Distributed Near Real-Time Data Processing Pipelines

Marc-Olivier Andrez, Diana Rueda, Tim Schoof, Mikhail Karnevskiy, Aleksandra Tolstikova, Vijay Kartik, Thomas White, Anton Barty

Deutsches Elektronen-Synchrotron (DESY), Hamburg, Germany



Enable scientists to make the best use of limited beamtime at synchrotron facilities



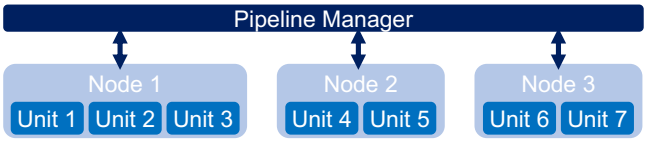
- Due to the high number of applications to run experiments on beamlines, scientists have limited beamtime allocated for their experiments.
- High performance data processing pipelines provide feedback on experiments in near real-time so that scientists can make informed decisions on follow-up experiments.
- When scientists are away from the beamlime, for example at night, AI agents can modify experiment plans when relevant.

Create high performance data processing pipelines from existing components

- No advanced skills in distributed computing required!
- Write a pipeline configuration file that describes connections between existing components (connectors, processing units).
- At runtime, the connectors and data processing units will use ASAP::O, a high performance distributed streaming platform, to exchange data up to several GB/s sustained over hours.

Deploy and run data processing pipelines on HPC clusters

- Use the pipeline manager that we are developing at DESY to deploy, start and stop data processing pipelines on HPC clusters from pipeline configuration files.



- The first version of the pipeline manager will use HashiCorp Nomad to install programs on the cluster nodes when required, to execute these programs and collect logs.

Implement new data processing units

- Develop new connectors, data processing units and AI agents using ASAP::O client libraries (available in C, C++ or Python), or the AsapoWorker library (recommended).
- The AsapoWorker library makes it easy to wrap existing Python data processing code (error handling, processing of late events, standardized input and output ports, compatible serialization formats).

Benefit from additional tools

- Use tools compatible with the data streamed into ASAP::O such as a web based data visualizer or a NeXus file writer under development at DESY.

Example of data processing pipeline at the DESY beamlime P65 for X-ray Absorption Spectroscopy (XAS)

Question / Comments: marc-olivier.andrez@desy.de, tim.schoof@desy.de, diana.rueda@desy.de

We gratefully acknowledge funding from the Helmholtz Association HGF for the ROCK-IT project.

rock-it-project.de