

Innovative Data Acquisition Solutions with HSDS at MAX IV Synchrotron

Tuesday 27 May 2025 09:00 (25 minutes)

MAX IV is an accelerator-based light source located in southern Sweden. It continuously operates 16 experimental stations using X-rays for material and life sciences. User data analysis primarily occurs on a small edge HPC, while automatic scientific data processing, from around 100 data sources, runs in an edge-cloud environment. These pipelines provide rapid feedback on data acquisition and analysis to control workstations and users' laptops at the facility or accessed remotely. The vast majority of raw experimental data is stored in the HDF5 format.

Data analysis and visualization at synchrotron light sources is inherently a distributed computing problem (HPC, edge, users' terminals). Traditionally, we have used distributed file systems together with HDF5, which motivated the development of VDS and SWMR. The HDF group introduced HSDS over half a decade ago, and POSIX HSDS storage was implemented shortly after.

This allowed us at MAX IV to start HSDS microservices for our experimental stations, even without object storage. We routinely use HSDS to store intermediate data for our distributed data reduction pipelines, particularly for streaming tomography. This offers several advantages, including no dependencies on distributed file systems and straightforward data access from users' terminals. We can also utilize the existing HDF5 tools, such as the silx viewer and other standard software like h5pyd.

In many cases, the intermediate HDF5 data do not require persistent storage. Therefore, we developed an extension for dranspose –the distributed data analysis pipelines framework, which allows in-memory Python dictionary data to be exposed with the HSDS/h5pyd interface.

Keywords: data acquisition, distributed applications, hybrid environments, memory resident datasets, HSDS, synchrotron, tomography

May we record your session?

Yes

Primary authors: ENGELMANN, Felix (Institute for Cybersecurity & Digital Trust (ICDT), The Ohio State University); MATEJ, Zdenek (MAX IV Laboratory, Lund University)

Co-authors: BRUDVIK, Jason (European Spallation Source); CASCELLA, Michele (MAX IV Laboratory, Lund University)

Presenter: MATEJ, Zdenek (MAX IV Laboratory, Lund University)