

P11 processing pipeline

History

[Jan Meyer](#) sent email on April 13, 2021 presenting a need for updating the auto-processing; mainly speed-up, because of the faster and bigger Eiger detector. A possible template from EDNA/MaxIV exists.

[Johanna Hakanpaeae](#) emailed on April 21, 2021; expresses urgency to the matter of severe processing lag between data taking and result; users start complain that it negatively impacts the beamline experiments.

Meeting on April 27, 2021 (virtual)

Participants:

[Johanna Hakanpaeae](#)

[Jan Meyer](#)

[Anton Barty](#)

[Sergey Yakubov](#)

[Frank Schluenzen](#)

[Christoph Rosemann](#)

[Tim Schoof](#)

[Mikhail Karnevskiy](#)

(Missing: [Vijay Kartik](#))

Overview by Johanna and Jan

Auto-processing on little cluster at beamline used to work fine, but new detector exceeds the capabilities to process results in time.

Results are crucial in case of short test scans to generate scan parameters for full scan.

Shifting to Maxwell computation is not straightforward (rights, users).

Current cluster runs "XDS" pipeline, timing is about "10s data taking is followed by ~8min processing";

more precisely: certain scans take about 2min to complete and are run in succession, processing takes 4 times longer, thus creating a back log with increasing lag to obtain the processing results.

The expectation is that more and more powerful compute nodes will solve the problem. Two additional nodes have already been acquired and are present in the Maxwell cluster.

Template for such a solution exists at MAX IV, but some obstacles are present. Namely:

- how to start a job (automatically, on Maxwell)
- how to identify/manage the access rights from a particular user
- the container format (EIGER writes HDF5 files instead of single image files) isn't well suited for "grid scans"

Use case summary

1. short data taking needs quick result turnaround
2. feedback needs to be immediate and readily available
3. "regular" processing also needs to be covered

The detector, as well as the used software is fixed; the only available solution is to increase the resources for processing.

Proposed solution

[Sergey Yakubov](#) proposes to use a special beamtime user that grants access to the beamline file system from Maxwell during the duration of the beamtime (i.e. *after* start_beamtime, *before* stop_beamtime).

[[Note: [usually only the core file system is accessible from the compute nodes of the Maxwell cluster](#)]]

The implementation of the solution has to be done in collaboration with IT, since it is governed by the configuration of the Maxwell cluster and the access to the file systems.

Identified steps are:

- aim at a processing speed at least as fast as the data taking; scale resources accordingly
- create clear definition of data flow and processing steps: get data access with sufficient rights, process and return the results in an accessible way
- automate the data and processing flow
- integrate the involved steps into the standard workflows

Issues (known/anticipated)

[Frank Schluenzen](#) raised the issue that in the case of several groups sharing a beamtime (e.g. BAG proposal), the access and accounting is also shared.

Tasking

[Vijay Kartik](#) has implemented a similar (but not identical) workflow for P10 and could help with identifying the different issues involved and outlining the needed steps.

[Sergey Yakubov](#) has agreed to help with the implementation of identified missing requirements on the side of Maxwell.

[Jan Meyer](#) and [Johanna Hakanpää](#) will review the progress and success.