Near real-time data analysis with ASAP::O

Tim Schoof

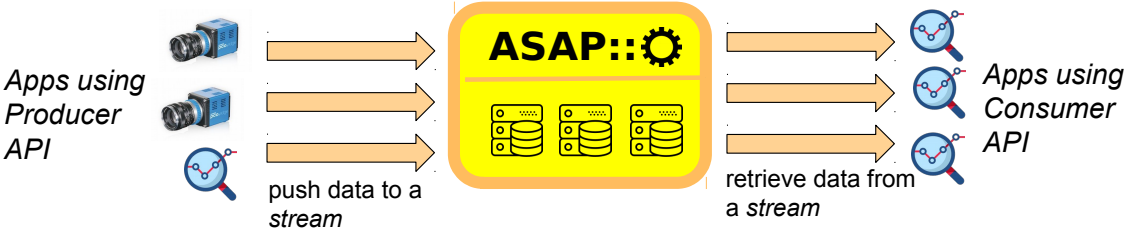
Hamburg, 10.11.2023

Near real-time data analysis

- Analyze all data while it is taken
- Combines the advantages of
  - Online monitoring: fast feedback for timely decisions
  - Automated analysis: ready-to-go pre-analyzed results
- A single workflow for easier configuration
- Analyzing all data increases insight during experiment
- Reduce disk usage by storing only useful data

ASAP::O

- High-performance data transfer
- In-memory buffer for near real-time access
- (Optional) persistent storage of data
- Same API for online and offline access
- Trivial parallelization on a per message basis (even across nodes)



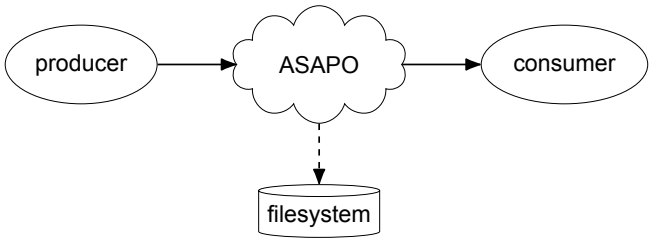
ASAP::O Data transfer

- Online consumer

- Consumer reads data from ASAPO memory buffer

ASAP::O Data transfer

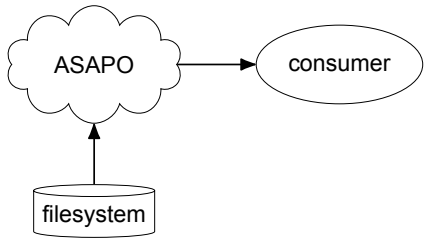
> Online consumer



- Consumer reads data from ASAPO memory buffer
- ASAPO optionally stores data in filesystem for offline access

ASAP::O Data transfer

> Offline consumer

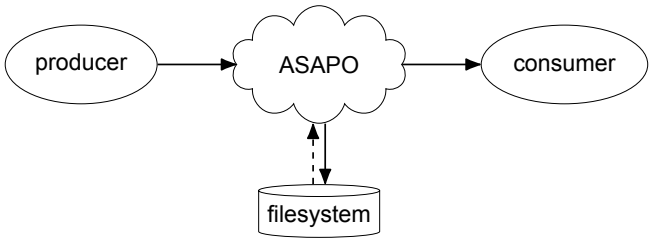


- Consumer or ASAPO reads data from filesystem

> Same API, no change in code or configuration

ASAP::O Data transfer

> Slow consumer

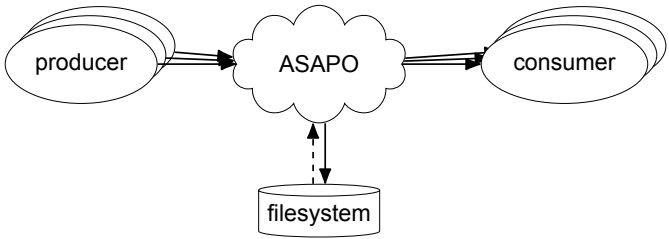


- Consumer switches from online to offline when it falls behind

> Same API, no change in code or configuration

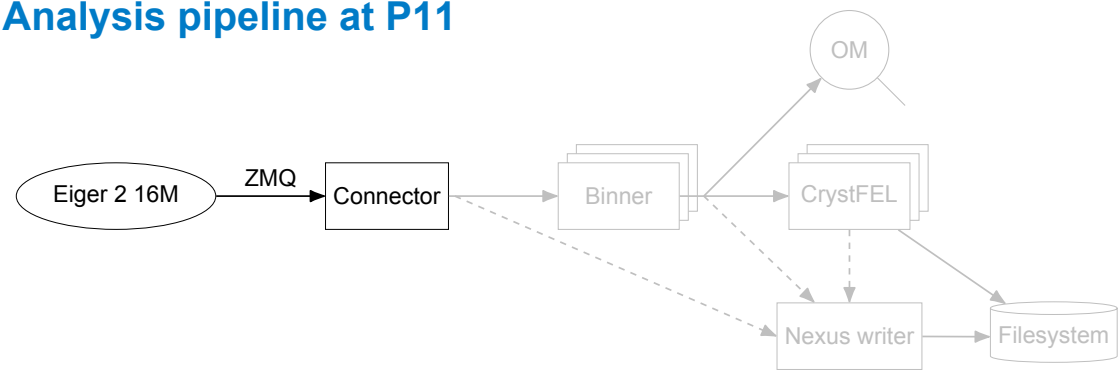
ASAP::O Data transfer

> Parallel procesing



> Consumers of a shared consumer group receive different messages
> Order of messages is preserved even with multiple producers

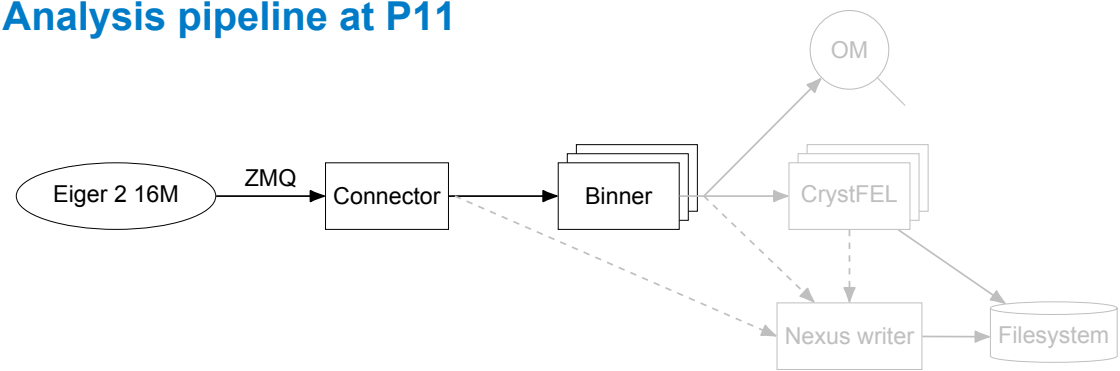
Analysis pipeline at P11



> Asapo-Eiger-Connector

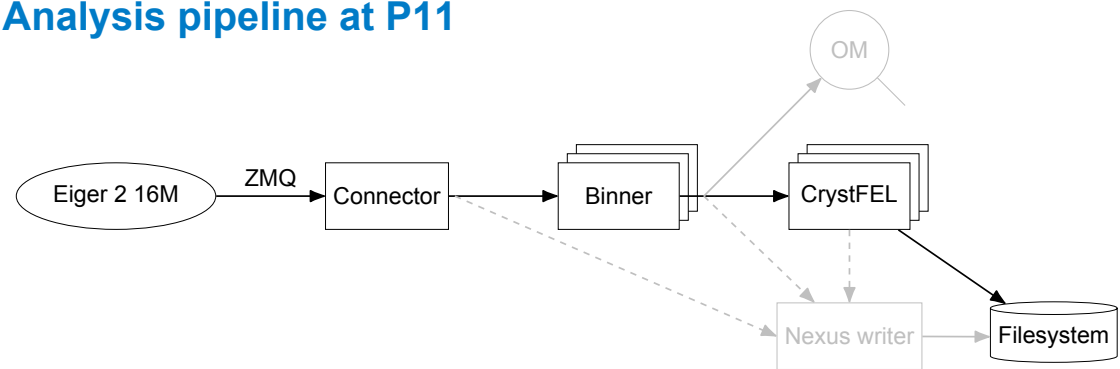
- reads raw data from Eiger ZMQ stream
- maps metadata keys to Nexus-like JSON structure
- converts data to seedee format
- sends converted raw data to ASAPO

Analysis pipeline at P11



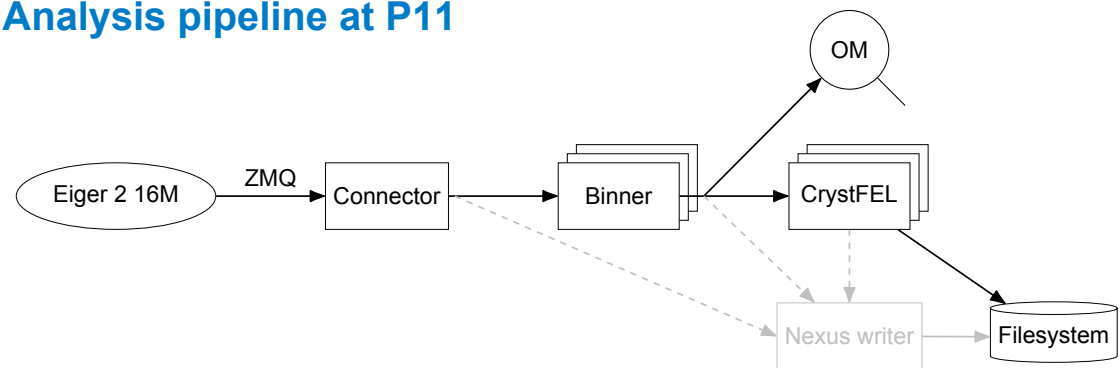
> (Optional) Pixel-Binner reduces image resolution
- reduces data size
- speeds up later processing steps

Analysis pipeline at P11



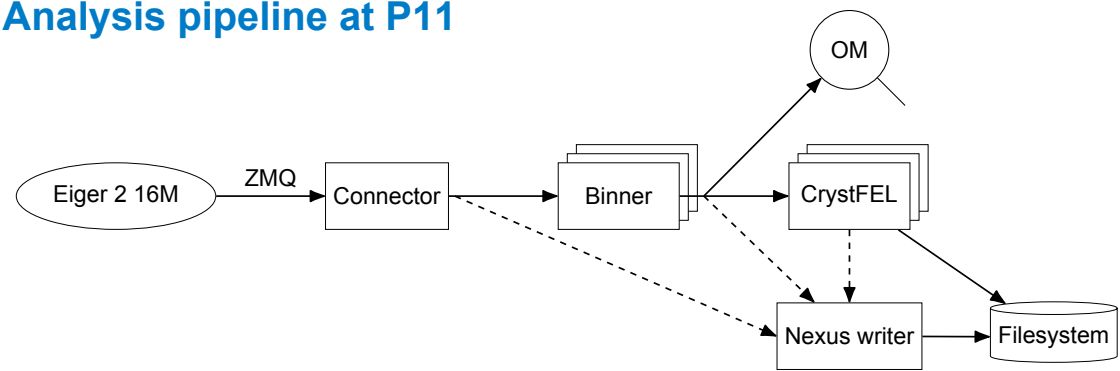
> CrystFEL for peak search, indexing, and integration
- Writes final result to disk
- Creates new ASAPO streams of only “hits”
- Connects directly to raw data stream if Binner is not used

Analysis pipeline at P11



- OM (OnDA Monitor) for live visualization
  - Can also be connected to raw data stream
  - Alternatively uses HTTP live view interface of Eiger

Analysis pipeline at P11



> Nexus-Writer for persistent storage
  - Can be connected to raw, binned, or hits-only stream
  - Writes Nexus files according to JSON structure created by Connector

Summary

- Near real-time analysis enables timely decisions based on all data during the experiment
- ASAP::O provides scalable, reliable, high-performance data streams
- Potential to significantly reduce storage requirements for serial crystallography
  - Only measure as long as necessary
  - Only store useful data

Outlook

- Serial crystallography at P09
  - Uses Pilatus instead of Eiger
  - Replace ASAPO-Eiger-Connector with Hidra and CBF-Converter
  - Rest of the pipeline components can be reused without changes
- ROCK-IT project
  - Involves DESY, HZB, HZDR, and KIT
  - Fully automated catalysis experiments at P65
  - Near real-time analysis for ML to make automated decisions
  - Discussion to use ASAPO as the streaming framework

Thank you!

Main contributors:

- Thomas White (FS-SC)
- Sergey Yakubov (IT)
- Alexandra Tolstikova (FS-SC)
- Philipp Middendorf (CFEL)
- Mikhail Karnevskiy (IT)
- Martin Gasthuber (IT)
- Diana Rueda (FS-SC)
- Marc-Olivier Andrez (FS-SC)

Contact

DESY. Deutsches Elektronen-Synchrotron

www.desy.de

Tim Schoof

FS-SC
tim.schoof@desy.de
+49–40–8998–1861