

# Dranspose

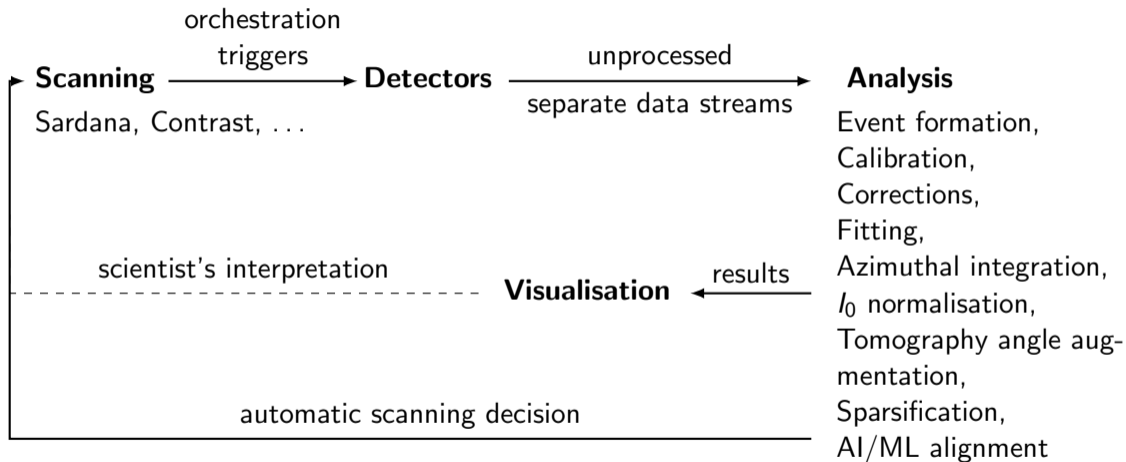
A constrained map-reduce live analysis pipeline for fast experimental feedback

Felix Engelmann, Scientific Data

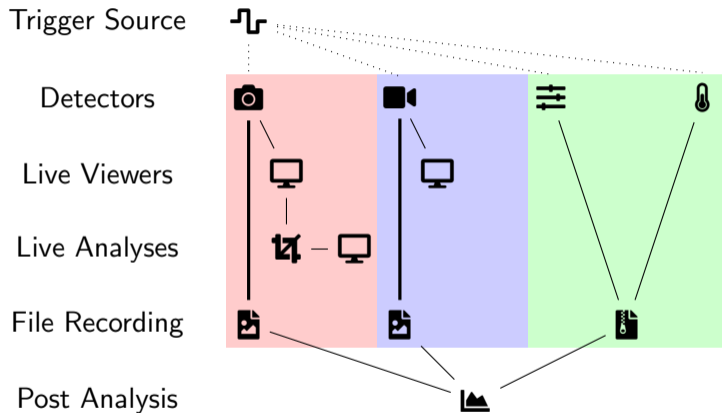
`fe-research@nlogn.org`

25<sup>th</sup> September 2024

# Experimental Feedback Loop



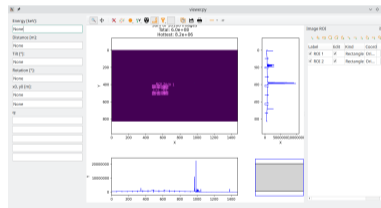
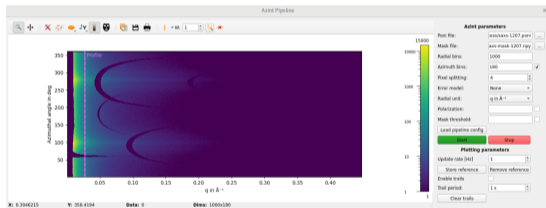
# Most Existing Infrastructure



# Limitations of Live Analysis

## Single Data Stream

- ▶ only one detector data available
- ▶ simple tools, e.g. azint, crop, time integration



## Beyond Limits

normalisation to  $I_0$ , sorting by motor position

# Additional Data Sources

## Devices

PandAbox



XSpres3



Basler

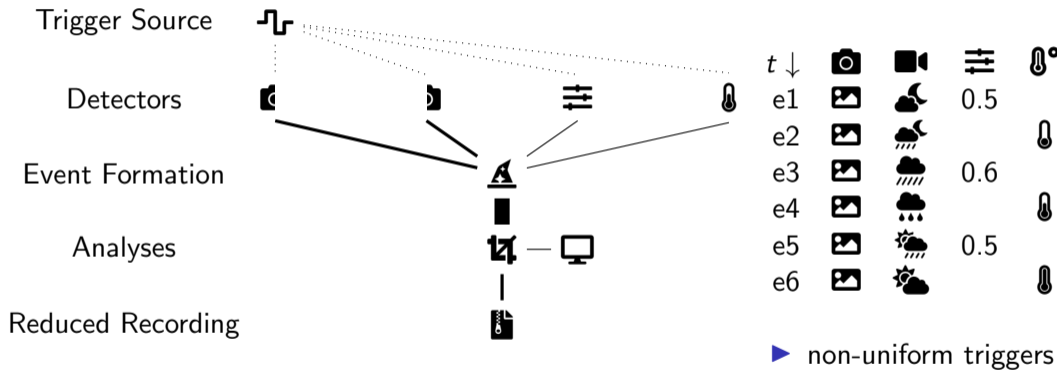


AlbaEM

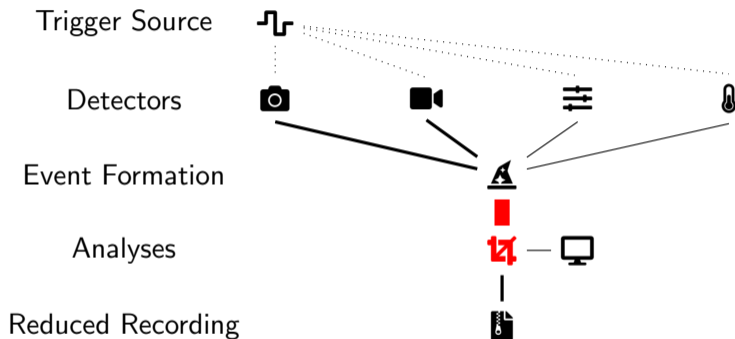
## Software

- ▶ Sardana
  - ▶ Icepap
  - ▶ Meta-data
  - ▶ Filename
  - ▶ Snapshots
- ▶ Contrast
- ▶ Bluesky

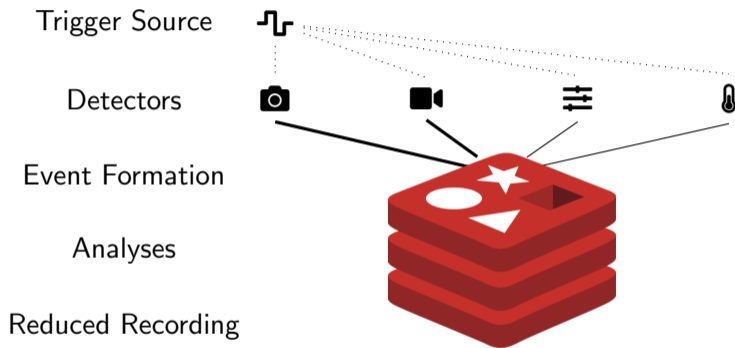
# Event Formation



# Event Formation Bottleneck



# Event Formation Bottleneck



# Existing Solutions

## Apache Flink

- ▶ exactly-once guarantee
- ▶ many small events

## Hummingbird

- ▶ XFEL development
- ▶ tailored to pulsed facilities

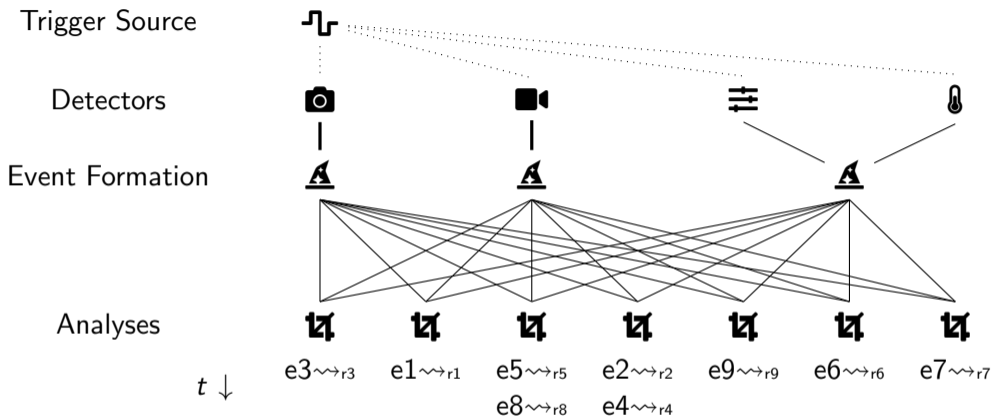
## OnDA / HiDRA

- ▶ DESY developments
- ▶ maintained?

## EWOKS

- ▶ large ecosystem
- ▶ data ingress from blisdata

# Parallelisation by Event, not Stream (map)



## Balancing Constraints

- ▶ event  $2n$  and  $2n + 1$  to same worker
- ▶ event  $m$  to worker debug

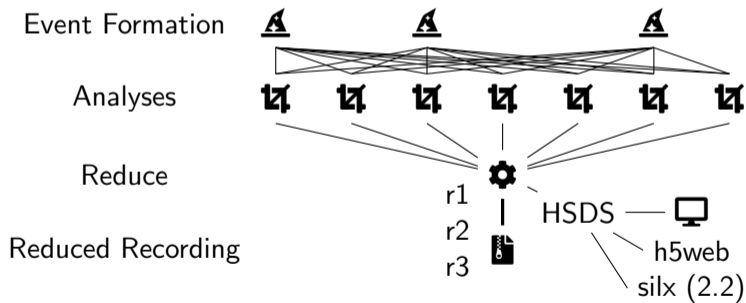
## Missing Inter Event Analysis

- ▶ time integration
- ▶ long temporal correlations

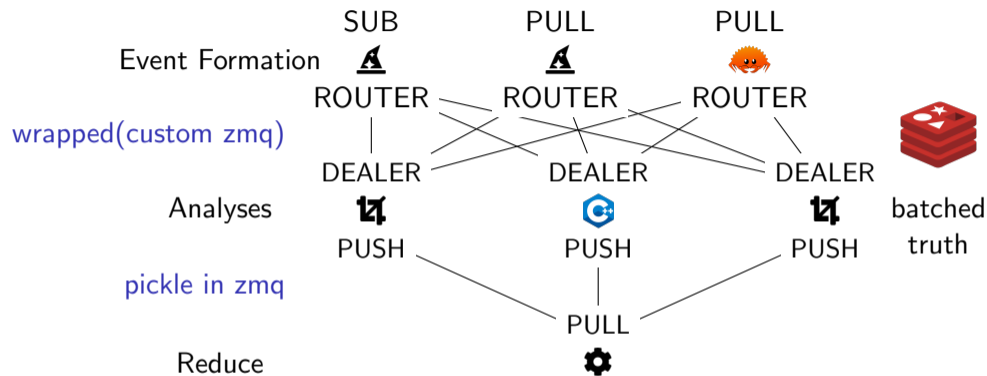
# Sequential Reduce

## Operations at Acquisition Speed

- ▶ append to list
- ▶ sum



# Behind the Scenes



## pytest

- ▶ end-to-end scans
- ▶ includes rust component

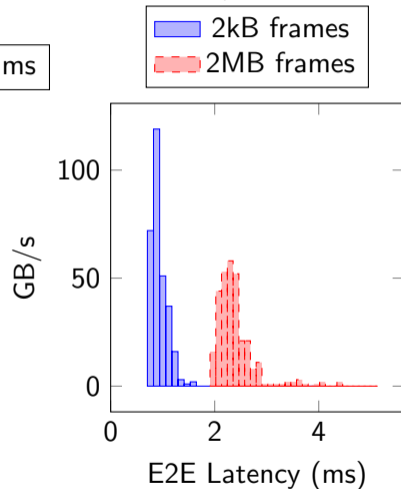
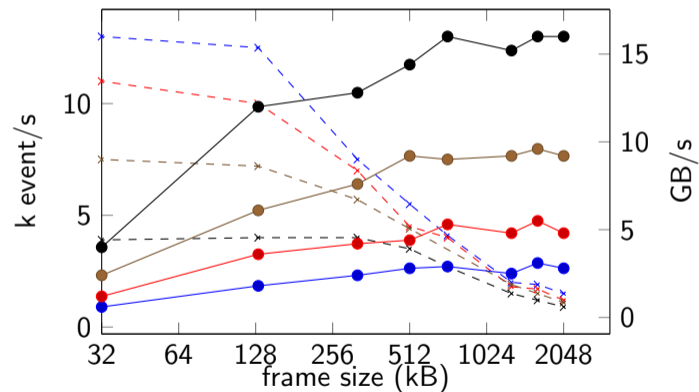
## Typing

- ▶ MyPy
- ▶ Pydantic

## Performance Estimates

K8s cluster of 10: dual Xeon<sup>®</sup> Gold 6326, 100 Gbit/s Mellanox SR-IOV, 465 GB RAM

—●— 1 stream —●— 2 streams —●— 4 streams —●— 8 streams



limits on around **2 GB/s** per stream or **12 kHz** at a **few ms**

# Development: Testing

## Recording

- ▶ ingesters optionally write all streams to disk (sequential cbor dumps)
- ▶ common experimental setup

## Replay

- ▶ from recorded stream dumps
- ▶ from hdf5 files

## Local

run file-based ingesters, workers and reducer locally

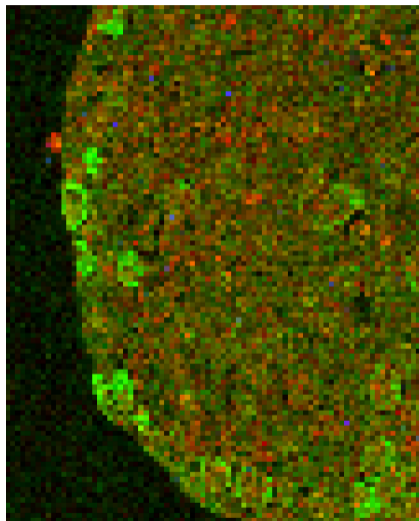
# Case Study: NanoMAX

## Sources

- ▶ XSpres3
- ▶ Contrast (Motor Positions)

## Analysis / Output

- ▶ PyMCA fluorescence fitting
- ▶ Concentration Map



# Case Study: DanMAX

Sardana Basler

Pilatus

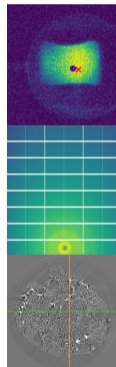
Orca

PandAbox

Beam CoG, Beam Alignment

Pileup File

Encoder angle



```
python -m pip install dranspose  
docker-compose, helm
```

<https://dranspo.se>

or come to Max IV, we are hiring: [paul.bell@maxiv.lu.se](mailto:paul.bell@maxiv.lu.se)