Data analysis support in Karabo at European XFEL

Hans Fangohr
Control and Analysis Software Group
Senior Data Analysis Scientist

ICALEPCS2017 Barcelona, 10 October 2017
Joint effort


European XFEL GmbH, Holzkoppel 4, 22869 Schenefeld, Germany

S. Aplin, A. Barty, M. Kuhn, V. Mariani

Centre for Free Electron Laser Science, DESY, Notkestrasse 85, 22607 Hamburg, Germany

T. Kluyver

University of Southampton, SO17 1BJ Southampton, United Kingdom
Outline

- European XFEL – current status
- Data analysis infrastructure
- Online data analysis
- Offline data analysis
- Summary
Data Analysis support in Karabo at European XFEL

Prof Hans Fangohr, Data Analysis Scientist, 10 October 2017
European XFEL – current status

- European XFEL, Northern Germany
  - Official opening 1 September 2017

- 2 of 6 scientific instruments live

- First experiments started 14 Sept 2017

- X-rays
  - 10 trains per second
  - Up to 2700 pulses (222 ns separation) per train
Data Analysis support in Karabo at European XFEL

Prof Hans Fangohr, Data Analysis Scientist, 10 October 2017
Data analysis infrastructure

- **Hardware: “Online cluster”,**
  - 8 nodes x 20 cores, 256GB RAM dedicated to users
  - Additional nodes for control and XFEL provided calibration and processing

- **Hardware: “Offline cluster” = Maxwell cluster (DESY)**
  - 80 nodes/3200 cores (Intel Xeon E5-2698v4)
  - ~112 TFlops
  - 512GB RAM each node
  - +20 nodes with other spec

- **Software: Karabo [1]**

---

Karabo processing pipeline example

- Data tokens pass through pipeline
- Processing units called “devices”
- Devices can be distributed over hardware
- Example: calibration done for 16 detector panels in 16 pipelines, distributed over 8 nodes
- Different protocols what to do with data if listening device cannot keep up
Online data analysis
Data Analysis support in Karabo at European XFEL

Prof. Hans Fangohr, Data Analysis Scientist, 10 October 2017

Control network
Karabo bridge
- Data stream input adapter
- Data stream output adapter
- Data analysis tool A
- Data analysis tool B

Live scenes

Data files
Sensors
Detectors
Experiment

Data files (HDF5) (XFEL)
Metadata (XFEL)
Data files (XFEL)

Data stream
input adapter

Data stream
output adapter

Online data analysis

European XFEL
Rapid feedback through GUI
Data Analysis support in Karabo at European XFEL

Prof Hans Fangohr, Data Analysis Scientist, 10 October 2017

Control network
Karabo bridge
Data stream input adapter
Data stream output adapter
Data analysis tool A
Data analysis tool B
Data files

Data files
Metadata
Raw data files (HDF5) (XFEL)
Data files (XFEL)

Detection of on the fly control feedback
Live scenes
Pipeline processing
Data Management
Experiment
Detectors
Sensors

Online data analysis
Karabo bridge

- Provide “receiving client code templates”
  - Python
  - C++

- Karabo bridge tested with OnDA [1] during first experiments

- Latency
  - 2.5 seconds corrected
  - 1.5 seconds uncorrected

- Provide test data set

Implementation:

- ZeroMQ as the network connection
- Protocol for serialised data under investigation. Candidates:
  - Pickled Python dictionaries
  - MessagePack
  - ProtoBuf

Offline data analysis
Offline data access

- Preprocessed data files can be requested on demand
- During experiment preprocessed files become available automatically on offline cluster (a few minutes after run stops)
Offline: Two ways of using preprocessed files

- Processing HDF5 files
  - Using European XFEL’s h5tools
  - Or directly

- Sending HDF5 files through the Karabo bridge
  - Imitates online setup
  - Good for re-use of interface
  - Can test in advance of experiment
Reproducible Science and Jupyter Notebook

- Jupyter Notebook
  - Executable document
  - Code, output, interpretation
- Integrate (Python) tools in Notebook
  - h5reading tools [1]
  - Example notebooks [2]
  - Pydetlib [3]
- Hope to grow library of analysis recipes with community

[1, 2] http://github.com/European-XFEL
Summary

- Online: GUI & Karabo bridge (& HDF5)
- Offline: HDF5 file based & Karabo bridge
- User support
  - Online documentation (http://tinyurl.com/ybx29ryt)
  - Growing set of open source tools (http://github.com/European-XFEL)
  - Access to Maxwell cluster (DESY)
  - Jupyter Notebook
  - Docker
  - Support before, during and after experiment
- Collaboration with users and other facilities desired

Acknowledgements:
- All current and past XFEL staff and collaborators
- OpenDreamKit Horizon 2020 European Research Infrastructures project #676541
- Gordon and Betty Moore Foundation through Grant GBMF #4856

Contact
- hans.fangohr@xfel.eu
- http://fangohr.github.io
- @ProfCompMod