

Data-management at PS

Mikhail Karnevskiy

8 Nov, 2023
DMA ST1 synergy workshop

Overview

Different facilities and even different beamlines within one facility may have different approach to data management. This can be explained by different reasons

- Type of experiments
- Used detectors
- Software stack: Control, DAQ, Storage, Streaming, Computing
- Management decisions

Following cases considered:

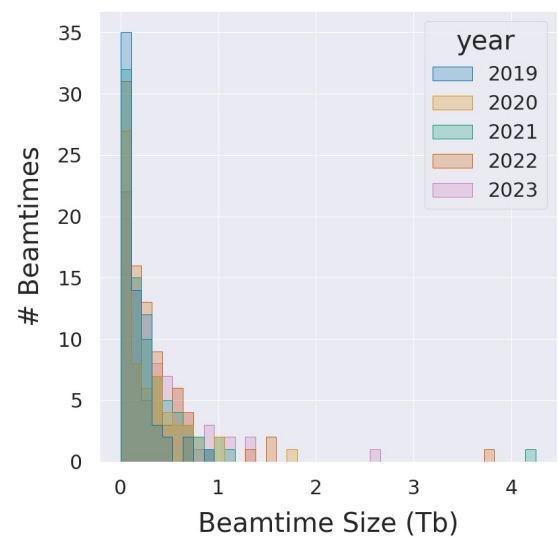
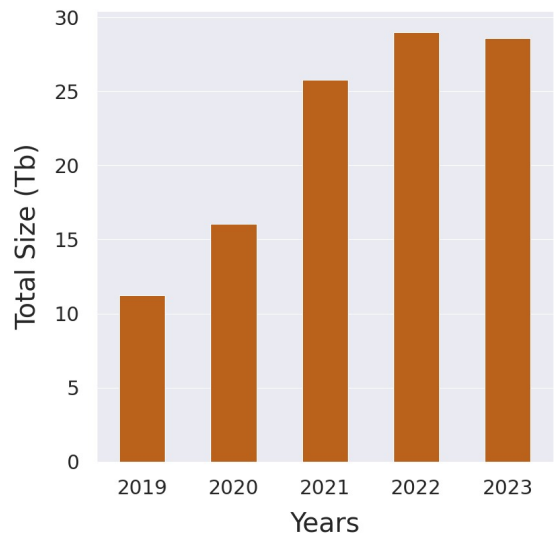
- Petra III: p02.2, p11, p62
- EXFEL-SQS

Infrastructure

- Both considered facilities are located very close to each other and share some resources (Maxwell cluster and other central IT services)
- Both Petra and XFEL facilities are sub-divided to Beamlines that have own computing resources and isolated file system.
- Considered Petra beamlines uses Tango control system. XFEL uses Karabo, which is on-side developed and supported.
- Support infrastructure:
 - Petra: Control, Scientific-Computing, Central IT
 - XFEL: Control, Data-Analysis, IT and Data-Management, Central IT

Data volume

P02.2-Petra III	P11-Petra III	P62-Petra III	EXFEL-SQS
27Gb in 2023 Lambda, PE, PCO H5 for Lambda	100Tb / Week Eiger, Nexus	10s Tb / Week Eiger, Nexus	2-3Pb / Week DCCS, PnCCD, H5



Metadata

Wavelength, beam parameters, sample parameters, environment parameters (Pressure, temperature, motors, mirrors), detector parameters (position, calibration)

Way of storing:

- Files on disc (p02.2, p62)
- Facility supported catalog (ExFEL)
- Amarcord (P11)
- Confluence pages
- Handwritten info

Way of writing metadata depends on control software/hardware (Autoionization).

Data processing

P02.2-Petra III	P11-Petra III	P62-Petra III	EXFEL-SQS
1-2 Nodes	3 Nodes	1-2 Nodes	5-20 Nodes

- Both facilities have a beamline computing resources (Much more at ExFEL) and share maxwell cluster.
- Beamline resources are used for monitoring, pre-processing, format conversion
- Local(facility) processing is always preferable. Exception are in case of smaller datasets
- After-Beamtime processing usually perform on local facility using scientific account.

Summary

To Be Discussed:

- Local vs federated data-processing cluster
- Data format, storage, distribution
- Local vs federated account
- ...