Data Handling Environment

Yasumasa Joti

On behalf of SACLA

Contents

- Data acquisition (DAQ) and analysis system at SACLA
  - ✓ Overview
  - ✓ Data analysis from outside SPring-8
  - ✓ Update of the storage system

- Data retention practice
  - ✓ SPring-8 Data and Network Committee
  - ✓ new rules
  - ✓ Environment for taking out data

Requirements for SACLA DAQ and Analysis System

- Reliable pulse-by-pulse recording of image data with data throughput up to 6Gbps.

  # 12 MPCCD sensors, max. rep. rate 60Hz

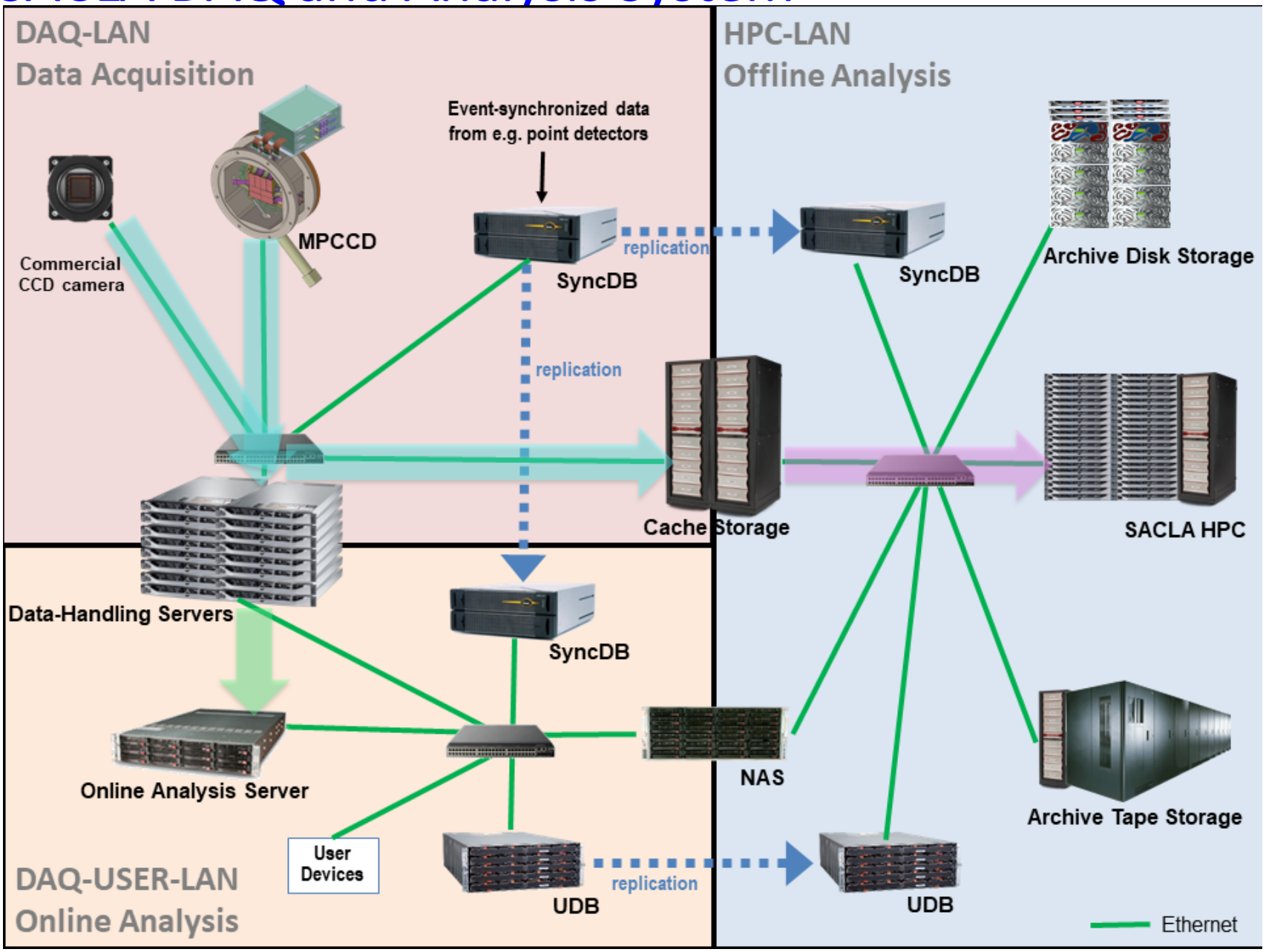
- Pulse-by-pulse recording of XFEL pulse characteristics to analyze a huge image set taking fluctuations of SASE into account.

- Fast handling capability for a huge data set to optimize the experimental conditions during experiments.

- DAQ infrastructure of the CITIUS detectors including 20.2 Mpixel systems is under development.

  => Breakout session A3 (Wednesday)

SACLA DAQ and Analysis System



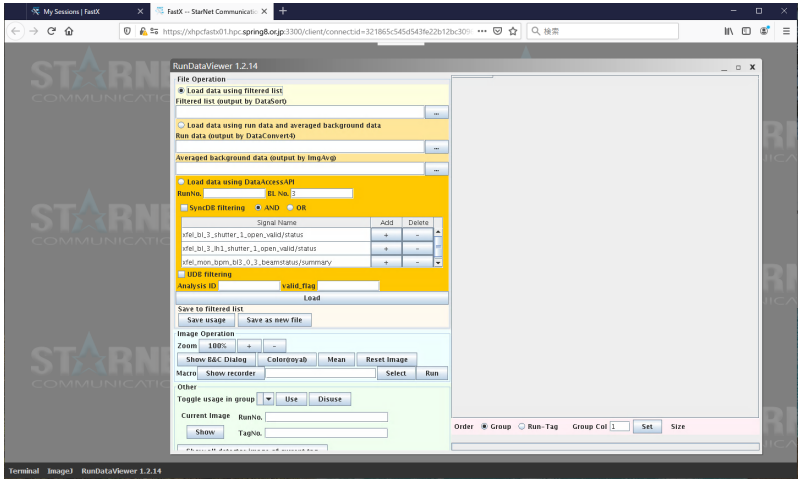
Joti et al. J. Synchrotron Rad. (2015).
Joti et al. Synchrotron Rad. News (2016).

Data analysis from outside SPring-8

The GUI response via VPN (~10Mbps) is very slow, which hinders data analysis from outside SPring-8.

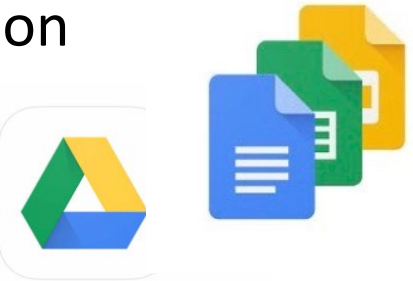
- To facilitate data analysis from outside SPring-8, we have introduced FastX servers in the HPC-LAN.
- FastX displays high-resolution X11 Linux applications with a comfortable speed over VPN.
- Users can use FastX via a web browser.
- 5 users now => 25 users in Mar. 2021.
- The manual will be available on SACLA HPC portal:

  http://xhpcfep.hpc.spring8.or.jp



Communication between on-site and off-site users

- A proxy server has been deployed so that on-site users can access Google docs, spreadsheets and drive.
  - ✓ screenshots and experimental logs taken at the operation consoles can be uploaded.



- A trial service of Nextcloud is launched so that on-site users can send experimental data to off-site users easily.
  - ✓ Nexcloud server has been introduced.
  - ✓ capacity limit will be larger than that of Google drive (free).
    < 1TB

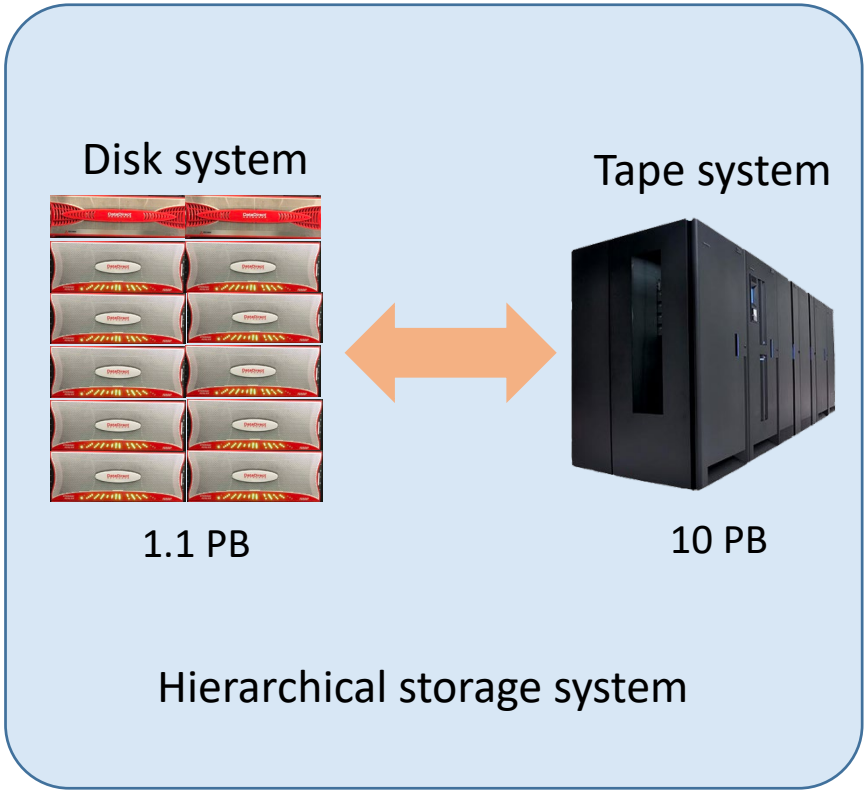
Nextcloud

SACLA data amount

[Raw data]	~ 6PB	(2017- 1.6 PB/year)
[User data]	> 3PB	

Update of the archive storage system

Old system (-2021.3)



- The data on the disk system will be moved automatically to the tape system a few months after it is saved.
- The data on the tape system will be recalled automatically when accessed.

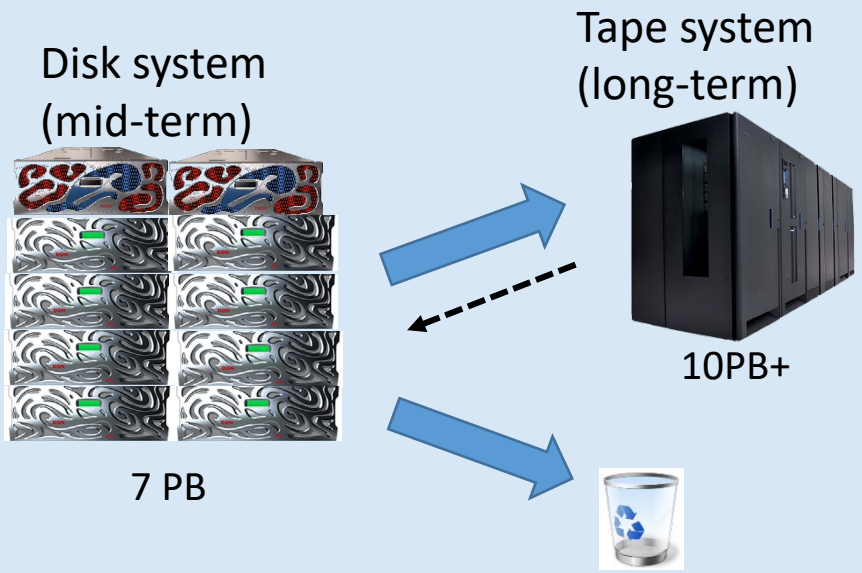
⇒ Data access speed becomes very slow a few months after it is saved.

This period is determined by 1.1 PB of the disk storage.

⇒ We have introduced a new disk system with larger capacity for a mid-term storage.

Update of the archive storage system

New system(2021.3-)



- Raw data will be moved from the disk to the tape 3 years after it is saved.
- User-generated data is retained in the disk system for 3 years after the file creation.

⇒ Data access speed remains fast for 3 years.

- Raw data on the tape system will be recalled when there is a user request.
- Thanks to the cooperation of users, data migration from the old system to the new system is progressing.

http://xhpcfep.hpc.spring8.or.jp/news/news_attachment_2020-08-11-001.pdf
(To access this PDF, a VPN connection is required)

SPring-8 Data and Network Committee

Established in Feb. 2020

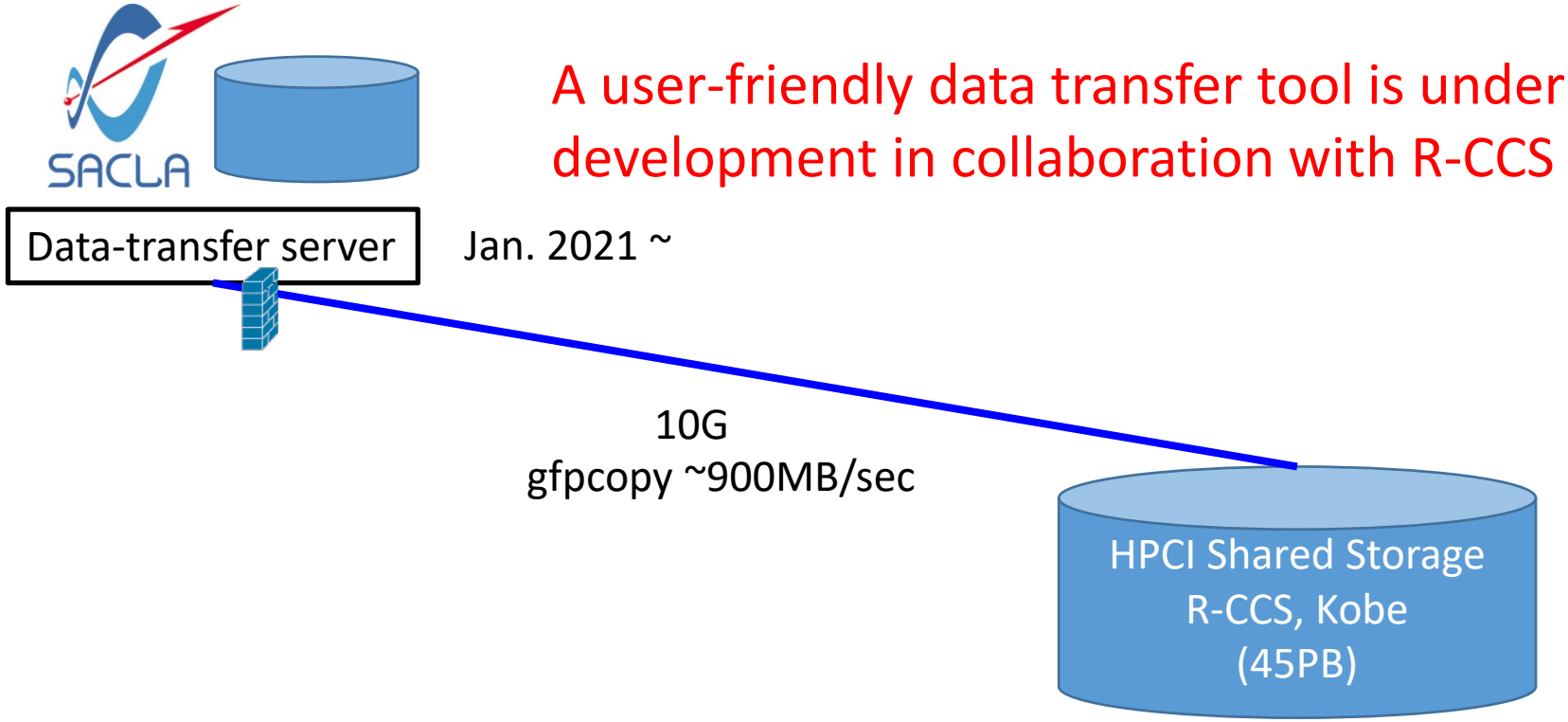
- An organization that formulates policies and rules related to the maintenance and operation of data and network system at SPring-8/SACLA.
- https://dncom.spring8.or.jp/
- Data retention policy was discussed at the 1st committee meeting.
  - ✓ Users of SPring-8/SACLA are responsible for managing the experimental data in accordance with data management policy of the organization to which they belong.
  - ✓ To utilize the experimental data efficiently, the facility is committed to providing users with the data system. Users should keep in mind that the facility do not guarantee permanent data retention.
  - ✓ The facility can set a data retention period, capacity limit, usage period, etc. for each data system.

New data retention practice

Data type	Owner	Cache system	Disk system	Tape system
SyncDB data	Facility	240 days	3 years	+ 7年
Raw data	User	90 days	3 years	+ 5年
User-generated data	User	—	3 years	—

- ✓ Published in July 2020.
  http://xfel.riken.jp/users/pdf/20200716_SACLA_data_retention_e.pdf
- ✓ Enforced from April 2021.
- ✓ We will provide a data take-out environment such as optical disk archive (ODA: 5.5 TB/cartridge, April 2021~) and Linear Tape-Open (LTO: 12 TB/cartridge, this summer~).
- ✓ We have deployed a server for data transfer from SACLA to outside.
  - To use the server, an application is required.
  - The procedure manual will be available on SACLA HPC portal: http://xhpcfep.hpc.spring8.or.jp

Data-transfer from SACLA to HPCI Shared Storage



- The High-Performance Computing Infrastructure (HPCI) is a shared computational environment in Japan. https://www.hpci-office.jp/folders/english
- To use HPCI Shared Storage, users need to make an application by themselves.
- We hope that HPCI Shared Storage will be one of the destinations for data whose retention period has expired.

Summary

- FastX is introduced in the HPC-LAN to improve GUI responses.
- Google docs, spreadsheets and drive are available on the operation consoles at experimental stations.
- The archive storage system has been updated.

http://xhpcfep.hpc.spring8.or.jp/news/news_attachment_2020-08-11-001.pdf

- New data retention practice will be enforced from April 2021.

http://xfel.riken.jp/users/pdf/20200716_SACLA_data_retention_e.pdf

- The data transfer server, ODA drive and LTO drive will help users to take out large amount of data.

Thank you for your attention.