Scientific Data Management and Analysis Software Framework

Yu Hu, Hao Hu, Fazhi Qi

Institute of High Energy Physics, CAS

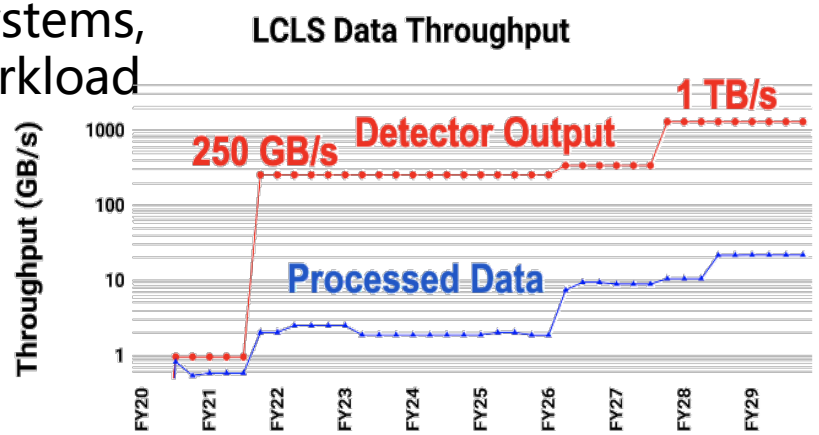
FCPPL2023@Zhuhai 11.06 - 10, 2023

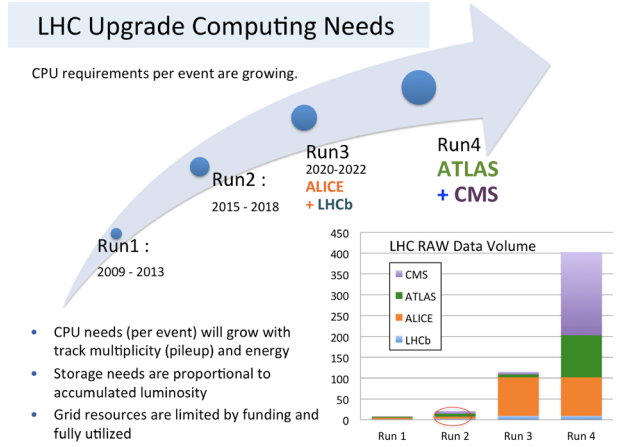
Outline

1. Introduction
2. Demand and Challenges of scientific data and software system
3. The architecture and design of the framework
   a. Data management software framework
   b. Data analysis software framework
4. Summary

Introduction

- The construction and upgrading of advanced scientific facilities have led to the explosion of scientific data
- New opportunities in science, but also challenges
- The existing code has struggled to cope with rising output levels from large-scale experiments
- Scientific discovery under the big data requires advanced mathematics and algorithms
- Efficient storage and processing requires large-scale data management and computing systems, well-designed workflows, and complex workload management software
- Loosely developed academic algorithms and software also need to be integrated into a unified software framework

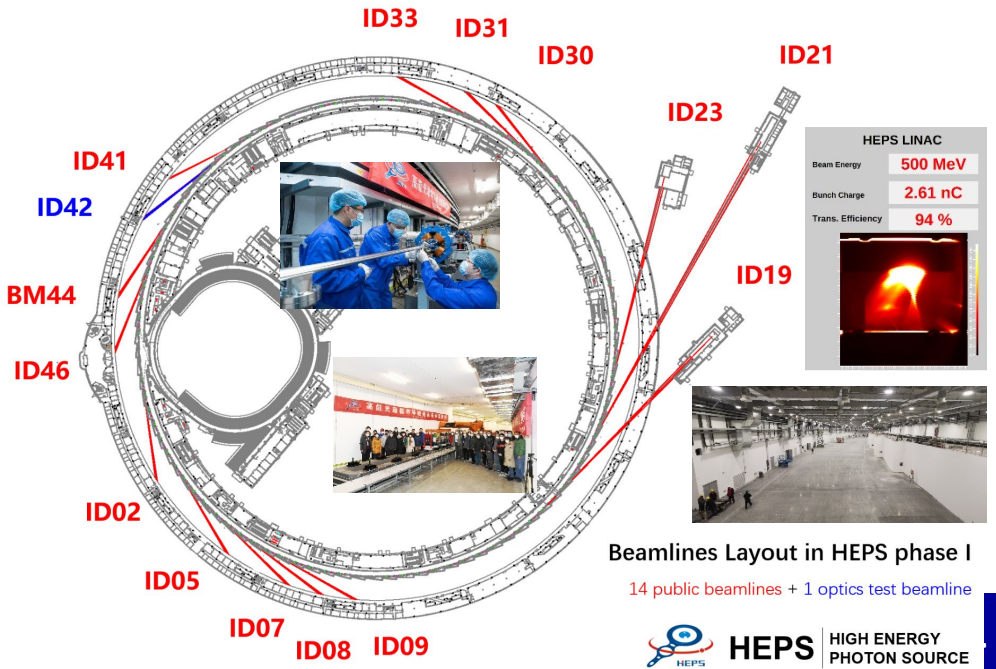




High Energy Photon Source (HEPS)

- New light source in China — High energy, high brightness
- Located in Beijing - about 80KM from IHEP
- The construction was started at the end of 2018, will be completed in mid-2025
- 14 public beamlines + 1 optics test beamline in Phase I
- Can accommodate over 90 beamlines in total

- **Progress**
  - The construction of the civil structure completed. Now at the equipment installation phase
  - 2023.01, HEPS booster installation completed
  - 2023.02, Start installation of storage ring
  - 2023.03, HEPS achieved the first electron beam accelerated to 500 MeV

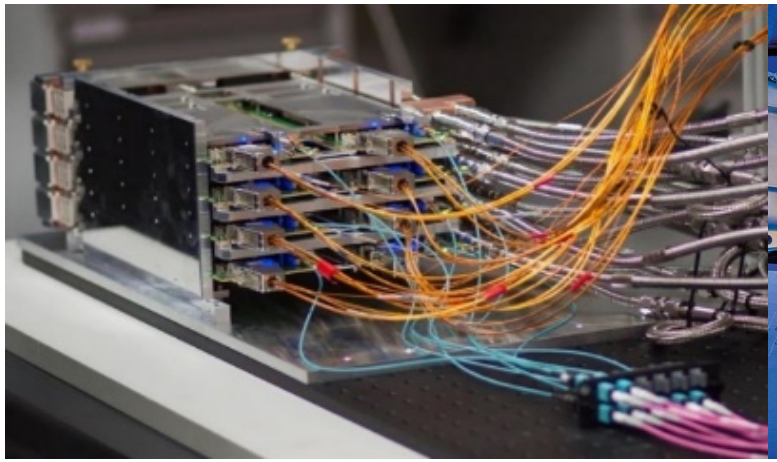


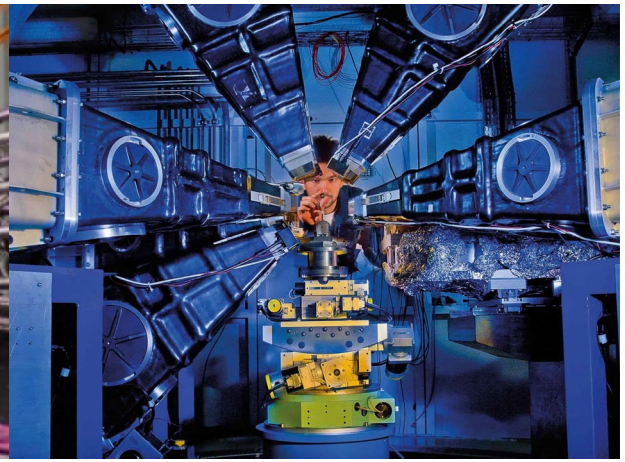
Outline

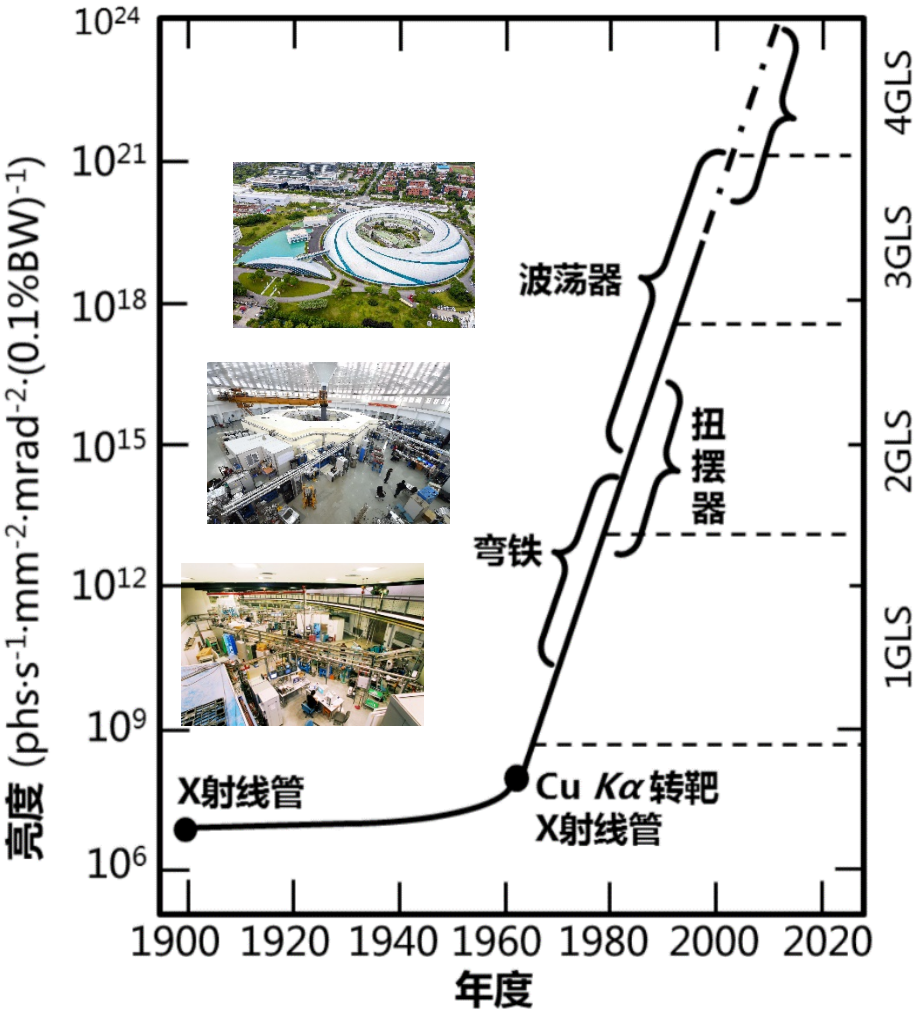
1. Introduction
2. **Demand and Challenges of scientific data and software system**
3. The architecture and design of the framework
   a. Data management software framework
   b. Data analysis software framework
4. Summary

Data Challenges

- Increased source brightness/ luminosity
  - More raw data in greater detail and less time
- Detector capabilities constantly improving:
  - Increased dynamic range, faster readout rates, larger pixel arrays
  - Bigger frames, higher frame rates
  - => more raw data



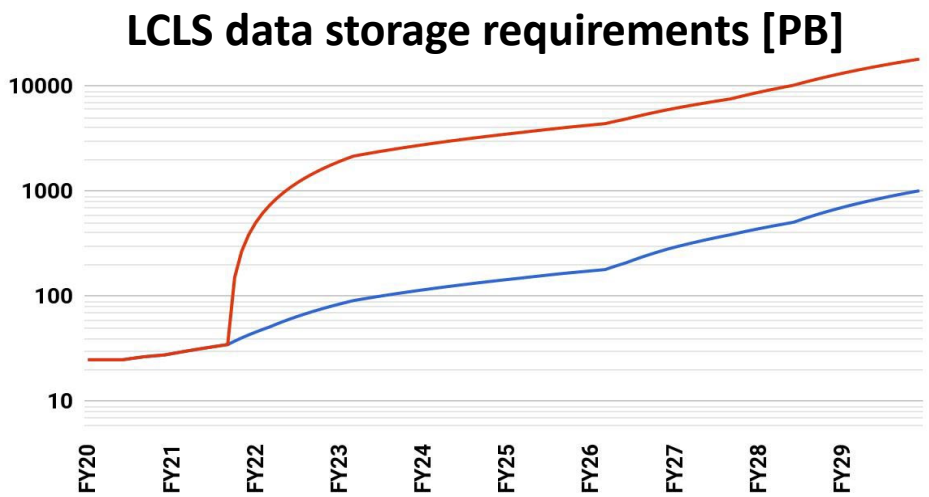




Development of synchrotron radiation light source

Data Challenges

- >200 petabytes of raw data per year for Phase I of HEPS (15 beamlines), more than 90 beamlines volume in total
- Over the next decade, the 5 DOE Light Sources are projected to generate ~1 exabyte of data per year
- The data volumes of the HL-LHC are estimated at roughly an exabyte of new data per year

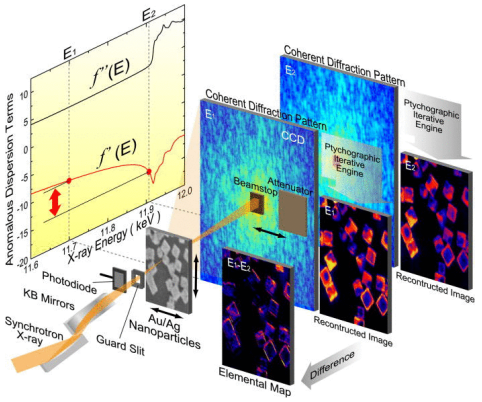


Data volume of HEPS Phase I Beamlines:

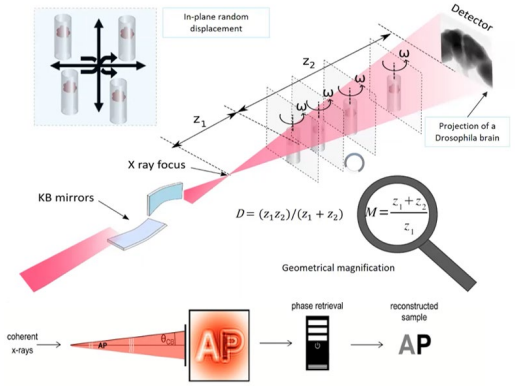
Beamlines	Burst output (TB/day)	Average output (TB/day)
B1 Engineering Materials	600.00	200.00
B2 Hard X-ray Multi-analytical Nanoprobe	500.00	200.00
B3 Structural Dynamics	8.00	3.00
B4 Hard X-ray Coherent Scattering	10.00	3.00
B5 Hard X-ray High Energy Resolution Spectroscopy	10.00	1.00
B6 High Pressure	2.00	1.00
B7 Hard X-Ray Imaging	1000.00	250.00
B8 X-ray Absorption Spectroscopy	80.00	10.00
B9 Low-Dimension Structure Probe (LODISP)	20.00	5.00
BA Biological Macromolecule Microfocus	35.00	10.00
BB pink SAXS	400.00	50.00
BC High Res. Nanoscale Electronic Structure Spec.	1.00	0.20
BD Tender X-ray beamline	10.00	1.00
BE Transmission X-ray Microscope	25.00	11.20
BF Test beamline	1000.00	60.00
Total average:		**805**

Data Challenges

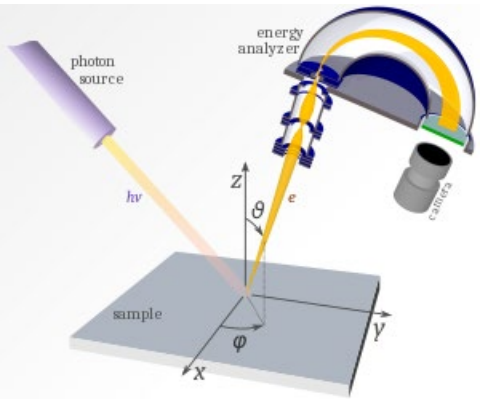
- New and more complex experiments
- Multi-modal experiments require combining data from multiple samples, techniques, and facilities
- In situ and in operando experiments require real-time feedback and autonomous control
- Data throughput and volume vary greatly with experiments and scientific goals
- New users from a wide variety of backgrounds and domains



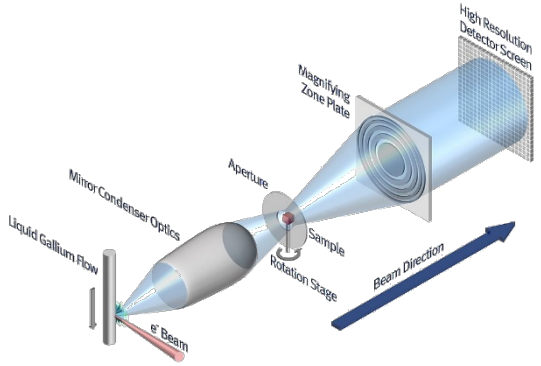
Fluorescence mapping



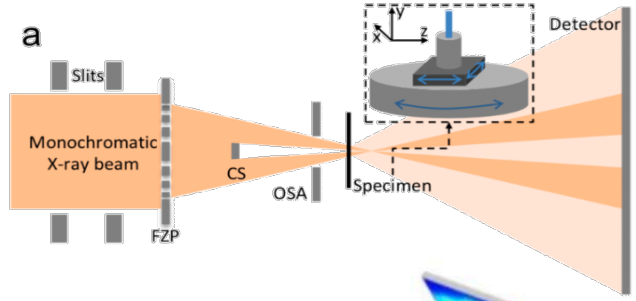
Nanoholotomography



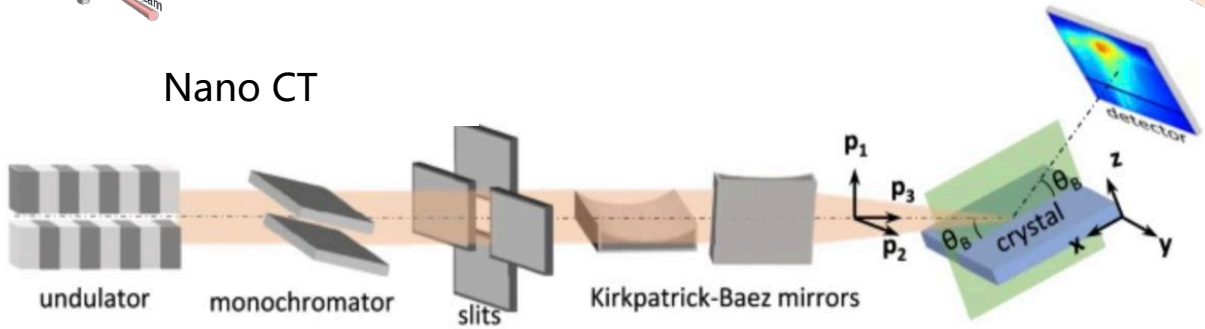
ARPES



Nano CT



Ptychography CT



Bragg ptychography

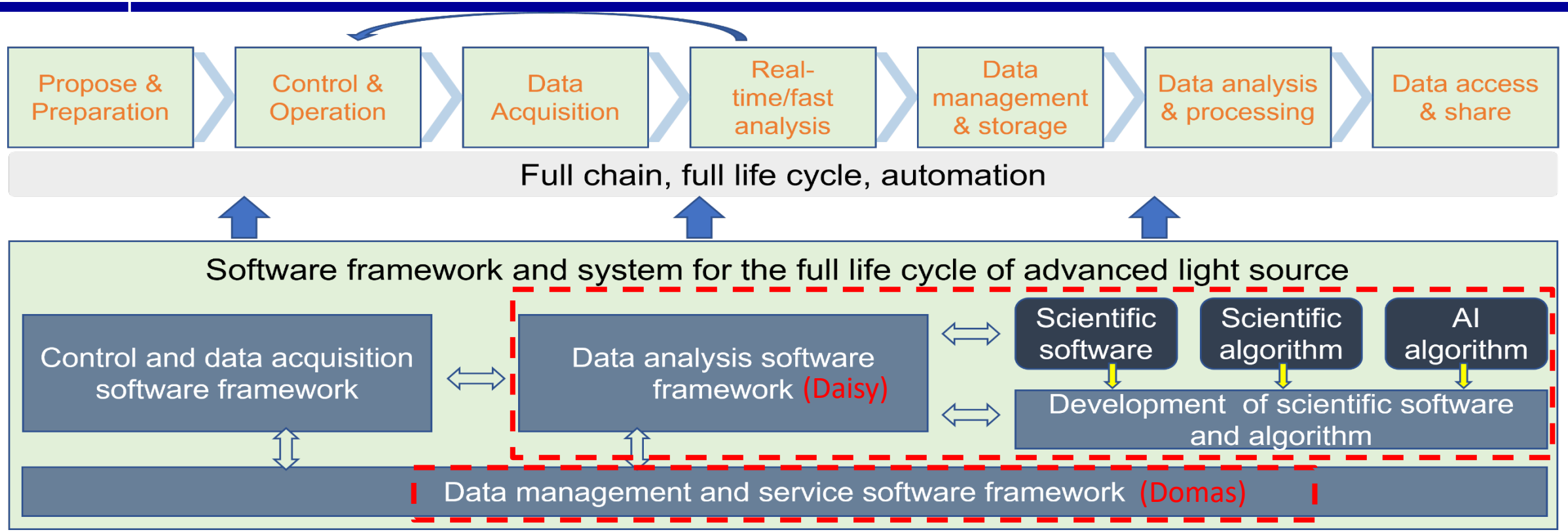
Data Challenges

- Analysis and management of large datasets at advanced scientific facilities is becoming progressively more challenging
- Development and integration of advanced analysis and management tools is needed
  - Provide storage, organization and management of massive scientific data
  - During the experiment, provide real-time analysis and fast feedback to guide the experiment steering and optimize the data acquisition
  - After the experiment, process the massive offline data, accelerate the scientific discovery
  - Provide the scalable distributed heterogeneous computing power, meet the diverse computing requirements of different scientific goals
  - Make things easy and faster for users

Outline

1. Introduction
2. Demand and Challenges of scientific data and software system
3. **The architecture and design of the framework**
   a. Data management software framework
   b. Data analysis software framework
4. Summary

Full data lifecycle software system



- Software framework and system for the full data life cycle of advanced scientific facilities
- Implement the tracking and management of scientific data throughout the entire process
- Support the development of new advanced data analysis methods, as well as the integration of existing software into the framework
- Designed for light source at first, but also fit for other facilities

Outline

1. Introduction
2. Demand and Challenges of scientific data and software system
3. The architecture and design of the framework
   a. **Data management software framework**
   b. Data analysis software framework
4. Summary

Tasks & Goals of Data Management

- **Data policy and Data Format**
  - Establish rules and regulations about data management
  - Design HDF5 data file format for each beamline, follows NeXus conventions
- **Metadata catalogue**
  - Design metadata model, catalogue, metadata database
  - Provide access to metadata and experimental data
- **Metadata acquisition**
  - Acquire metadata from other sub-systems(DAQ, transfer, storage, analysis…)
- **Data transfer**
  - Transfer all the data between beamline storage, central storage and Tape
  - Interact with metadata catalogue when the data storage status changed
- **Data service**
  - Provide a web-based GUI for user to search, access, download, analysis data

HEPS Data Policy

The ownership, curation, archiving and access to scientific data and metadata

- Recommend providing at least 3 months disk storage and permanent tape archive
- Provide permanent storage for raw data
- Provide temporary storage for processed data, calibration data and result data
- Each dataset will have a unique persistent identifier(CSTR/DOI)
- Experimental teams have sole access to the data during the embargo period
- After the embargo, the data will be released with open access to any registered users of the HEPS data portal

Apr. 2021, The Data Policy for HEPS(draft) is finished.

Mar. 2023, The Data Policy for HEPS(draft) is revised after a lot of discussion.


HEPS 科学数据策略（初稿）

一、前言

科学数据策略主要包含：

- 数据所有权。
- 数据管理。
- 数据存储。
- 数据访问。

本数据策略严格遵照国务院办公厅于 2018 年 3 月发布的《国家科学数据管理办法》的相关规定。

二、定义

2.1 HEPS：高能同步辐射光源。

2.2 HEPSCC：高能同步辐射光源计算与网络通讯系统。

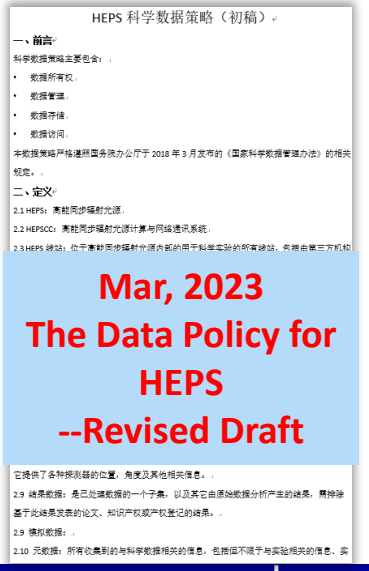
2.3 HEPS 线站：位于高能同步辐射光源内部的用于科学实验的所有线站，包括由第三方机构或群体投资或独资建设的线站。

Apr, 2021
The Data Policy for HEPS
--Draft Version

基于此结果发表的论文、知识产权或产权登记的结果。

2.9 模拟数据：

2.10 元数据：所有收集到的与科学数据相关的信息，包括但不限于与实验相关的信息、实


HEPS 科学数据策略（初稿）

一、前言

科学数据策略主要包含：

- 数据所有权。
- 数据管理。
- 数据存储。
- 数据访问。

本数据策略严格遵照国务院办公厅于 2018 年 3 月发布的《国家科学数据管理办法》的相关规定。

二、定义

2.1 HEPS：高能同步辐射光源。

2.2 HEPSCC：高能同步辐射光源计算与网络通讯系统。

Mar, 2023
The Data Policy for HEPS
--Revised Draft

它提供了各种探测器的位置，角度及其他相关信息。

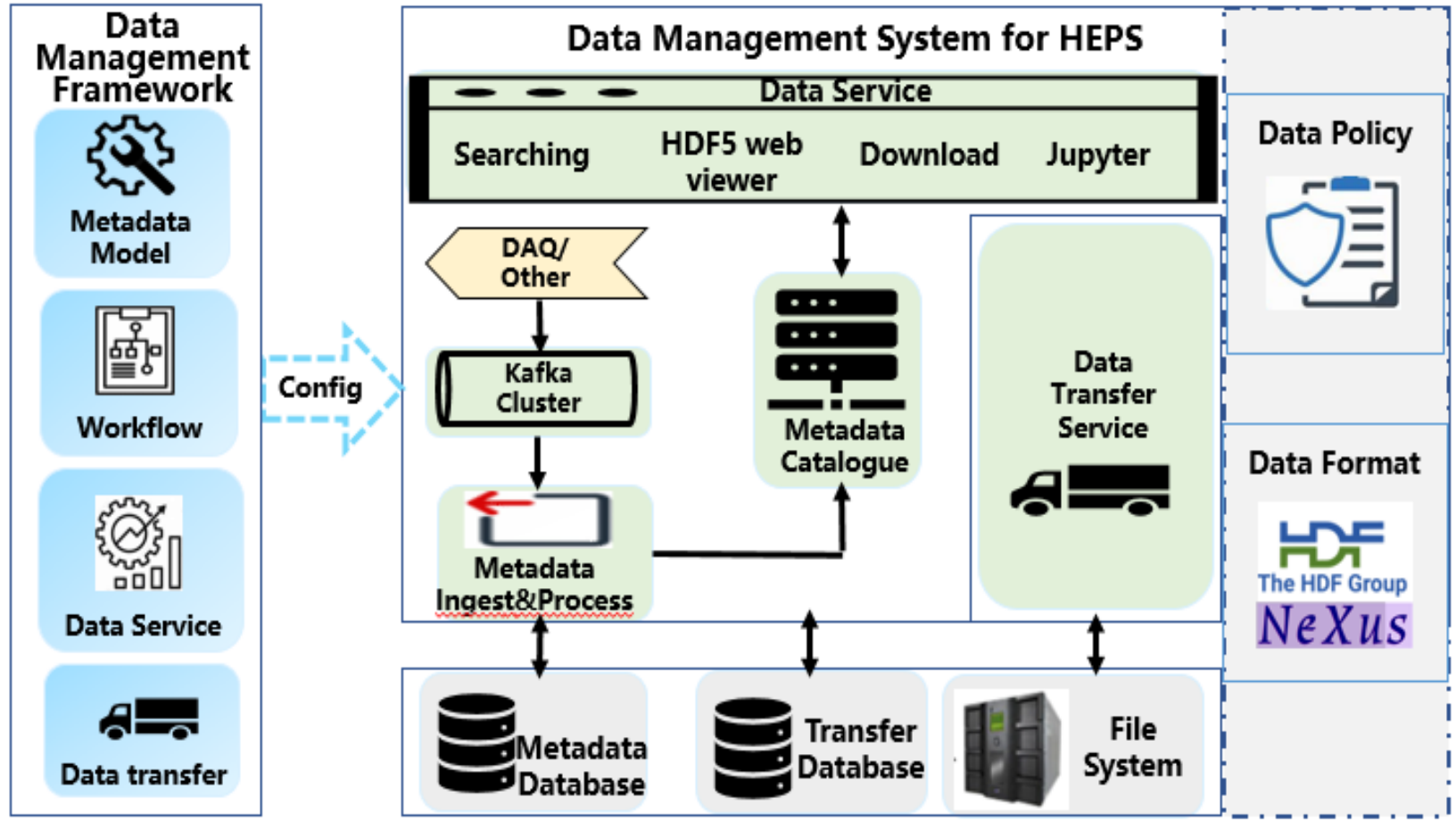
2.9 结果数据：是已处理数据的一个子集，以及其它由原始数据分析产生的结果，排除除基于此结果发表的论文、知识产权或产权登记的结果。

2.9 模拟数据：

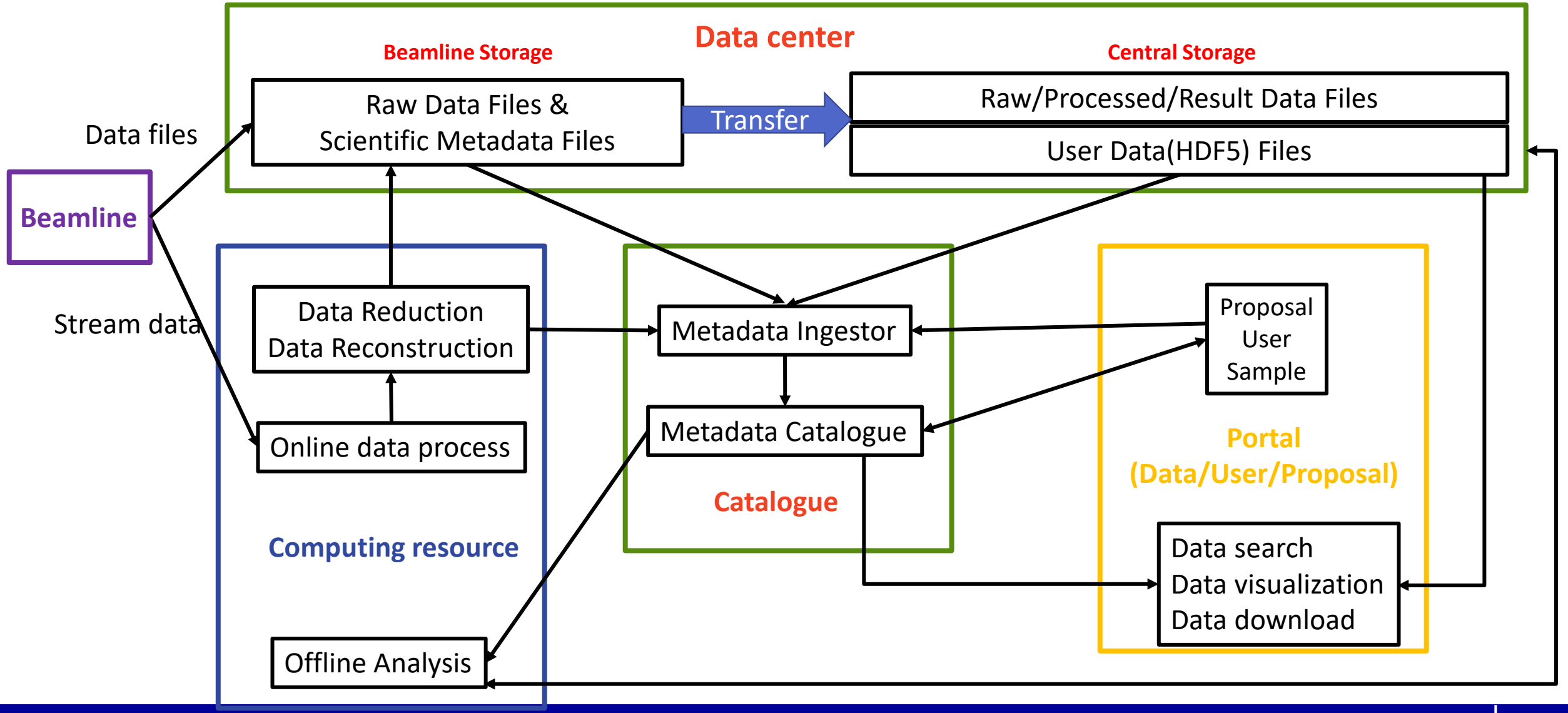
2.10 元数据：所有收集到的与科学数据相关的信息，包括但不限于与实验相关的信息、实

Data management framework

- Common function modules of data management
  - ✓ Metadata Model
  - ✓ Workflow
  - ✓ Data service
  - ✓ Data transfer
- Extensible and standard interface
- Be able to build data management system for facilities/beamlines quickly
- Will be open-sourced progressively



Data Flow of HEPS

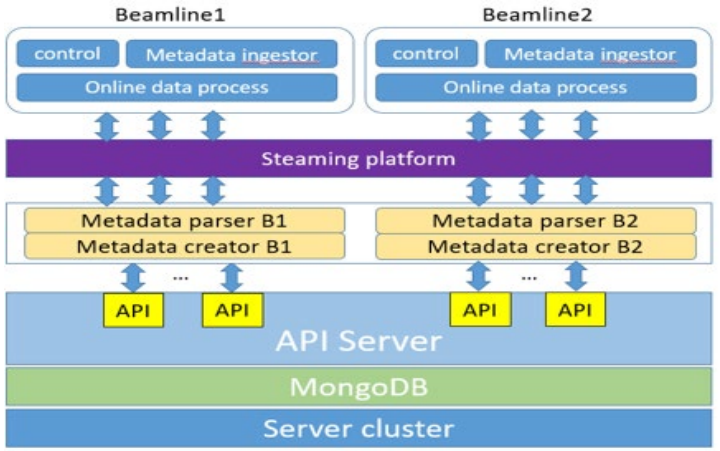


Metadata items to cataloging & Acquisition

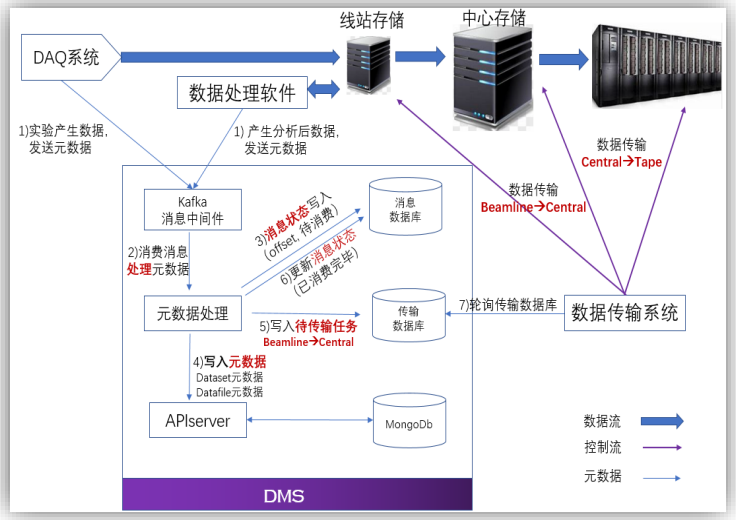
Metadata	Metadata Items	From
◆ **Administrative Metadata**	• Proposal Info, User Info, Exp type, Beamline…	Proposal system, User service system, Transfer system, Storage, Analysis system
	• Data type: raw data, processed data, simulated data, calibration data	
	• Dataset： PID, Path, Data file list, file size, checksum…	
	• Status: disk/tape, transfer status, transfer check value	
	• Analysis software, update time…	
◆ **Scientific Metadata**	• Sample Info	Sample database, Proposal system,
	• Exp environment params： voltage、magnetic field、electric field… Detector Info: scan, x-ray exposure params…	DAQ system, Control system
	• E-log	E-log System

The progress of data management

1. **Finished the core function modules**
   - ✓ Metadata catalogue, metadata ingestion, data transfer, data service
   - ✓ Interfaces with other systems: control system, transfer module, storage system, analysis system
   - ✓ Implement automatic data transfer between hierarchical storage (**beamline storage → central storage → tape**)
2. **Provide plans for data management when network interrupts**
   - ✓ when network interrupts, metadata and data are saved to local disk
   - ✓ After the network recovers, metadata will be sent to be catalogued
3. **HEPS data format design**
   - ✓ Designed and released data format for 7 beamlines
4. **Extend to space astronomy**
   - ✓ Preliminary designed HERD DMS, prototype system validated the solution



Metadata catalogue

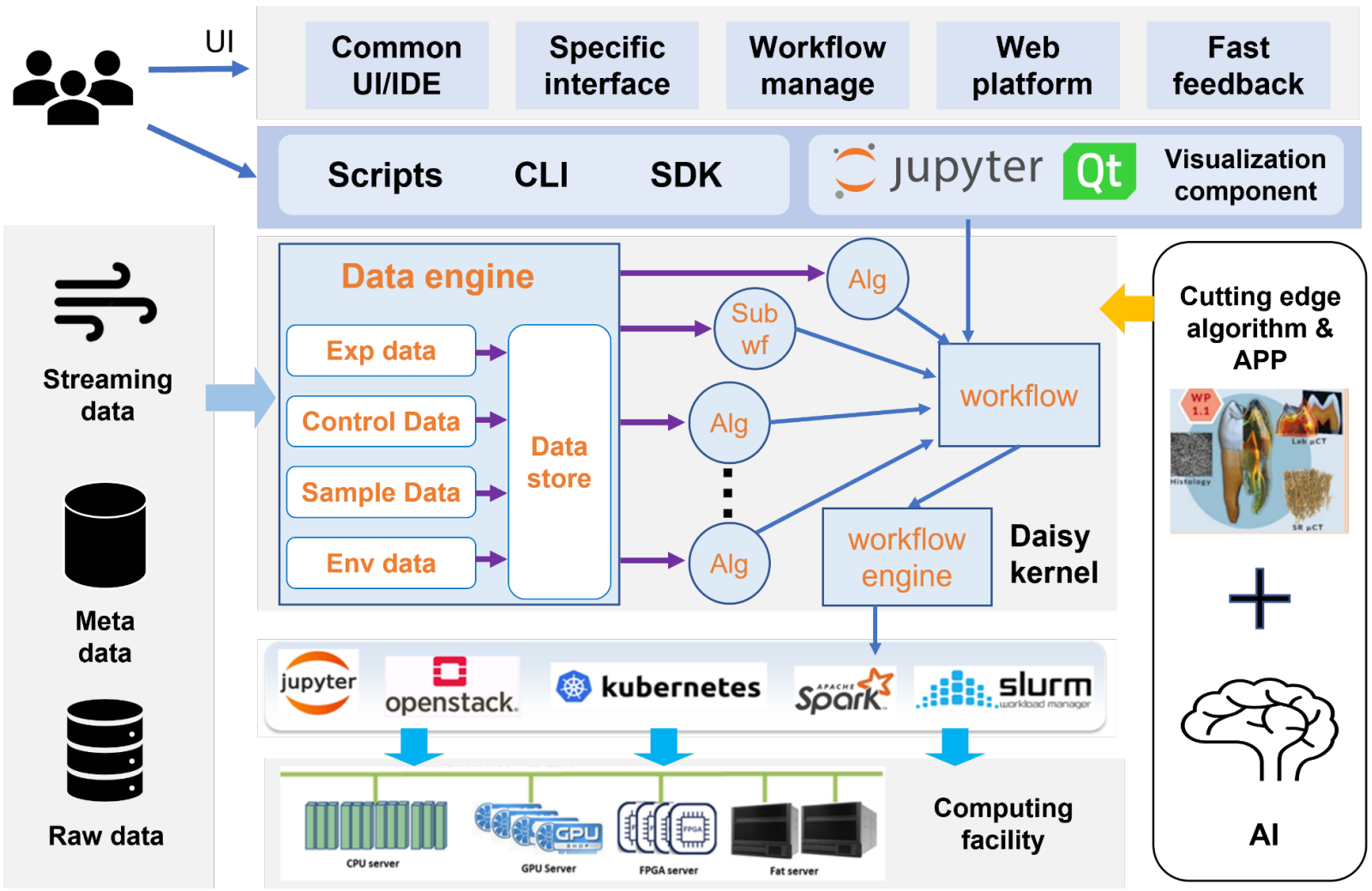


Automatic data management flow

Outline

1. Introduction
2. Demand and Challenges of scientific data and software system
3. The architecture and design of the framework
   a. Data management software framework
   b. **Data analysis software framework**
4. Summary

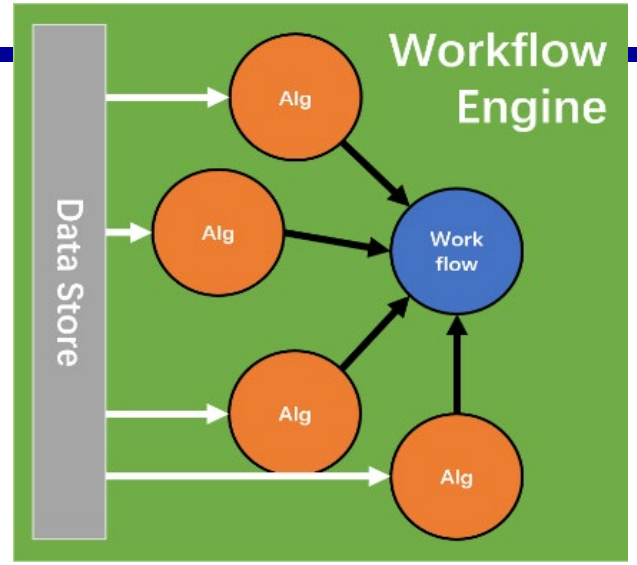
Data analysis software framework—Daisy



- Kernel of the framework
- Derivative technology modules to meet the data processing requirements of advanced scientific facilities
  - Data object management module for high-throughput data I/O, multimodal data exchange, and multi-source data access.
  - Scalable cluster computing power support for data processing with different scales, different throughputs, and low latency
  - Interface and developing environment for scientific software integration and development
- Domain specific App and flexible general workflow management system based on the framework

Kernel of the Daisy framework

Extract domain models independent from technology, and establish relationships between models to form a domain architecture

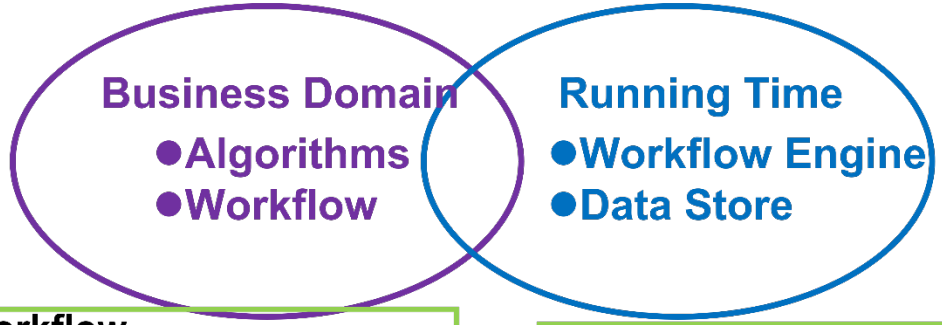


Four core modules are provided:

- **Algorithm:** The smallest unit in framework, defining the domain model, basic data processing module, support integration of third-party libraries.
- **Workflow:** Defines the domain architecture, execute processing tasks by calling a series of algorithms, supporting nesting.
- **Workflow Engine:** Manages the runtime environment and the distribution of the algorithm modules. Uncouple the process task from the computing environment.
- **Datastore:** Manages the creation and transmission of data objects between algorithms.

Algorithms
- Input Data Processing
- Output Data Defined

Workflow Engine
- Handle Data Store
- Running Time Management

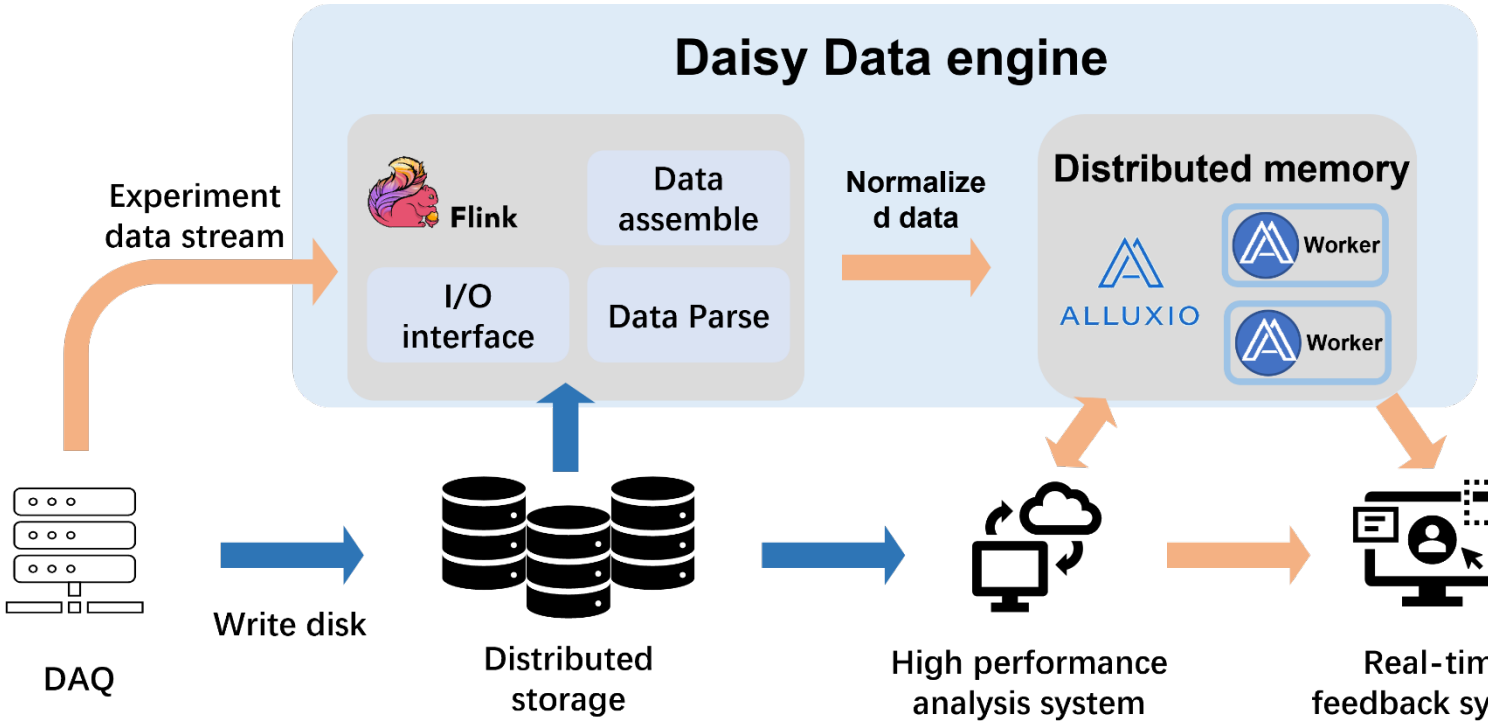


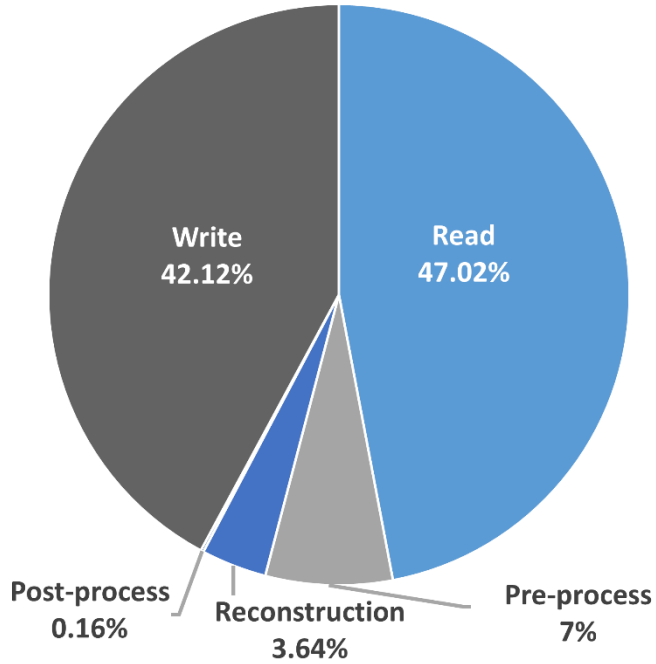
Workflow
- A sequence of Algorithm
- Workflow is also an algorithm

Data Store
- Data Object Management

Data engine

- Unified I/O interface to shield the difference of underlying architecture and data structure
- Support the I/O of streaming data besides disk file, for real time, high throughput data process
- I/O is the bottleneck. Employ asynchronous parallel, distributed memory, adaptive storage parameters and compression to optimize the I/O

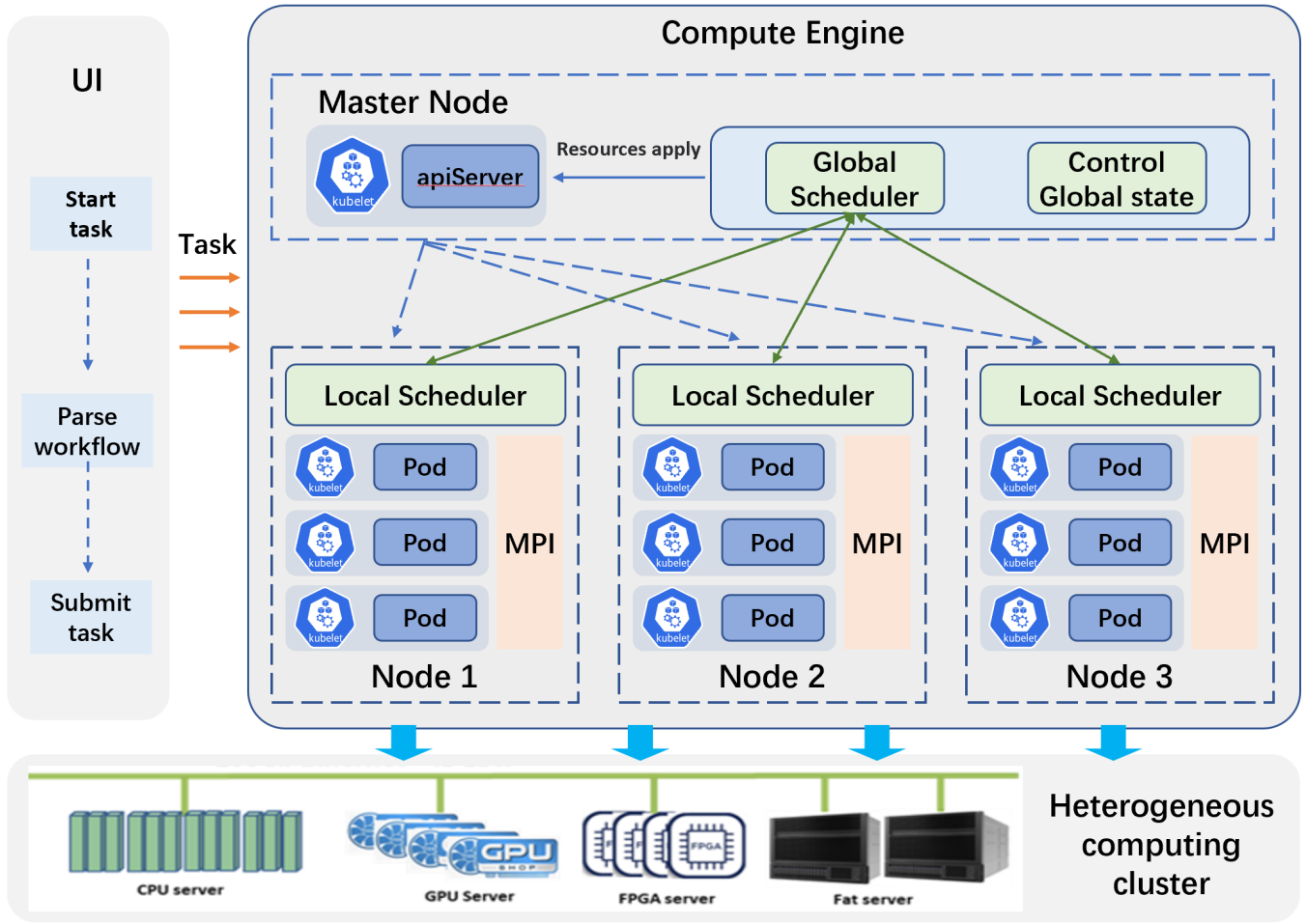




CT data processing time ratio

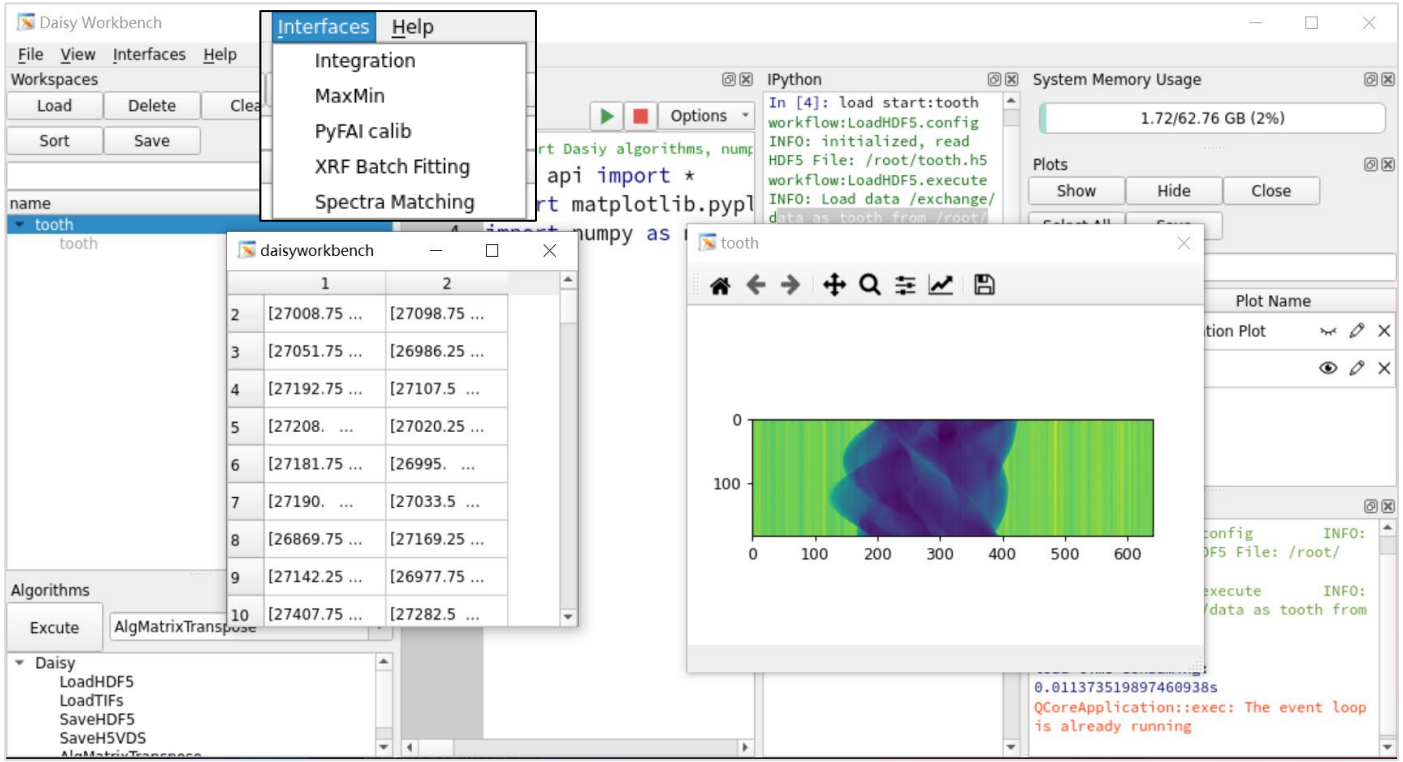
Compute Engine

- Support heterogeneous distributed computing power
- High performance numerical analysis computing library to speed up the computing hot spots
- Provide a unified flexible programming interface API for computing models, to reduce the complexity of parallel programming
- Two layer distributed computing task scheduler to achieve better efficiency



Under development

Daisy graphical user interface



Daisy workbench:

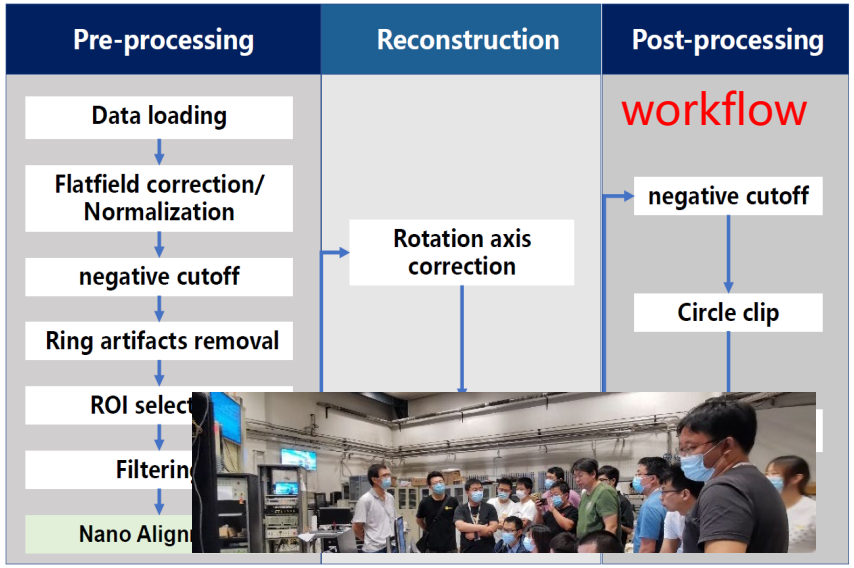
- General-purpose GUI based on PyQt5
- Include data object list, algorithm list, data view/visualization, and IDE for developers
- Interfaces of custom GUIs for a variety of scientific techniques

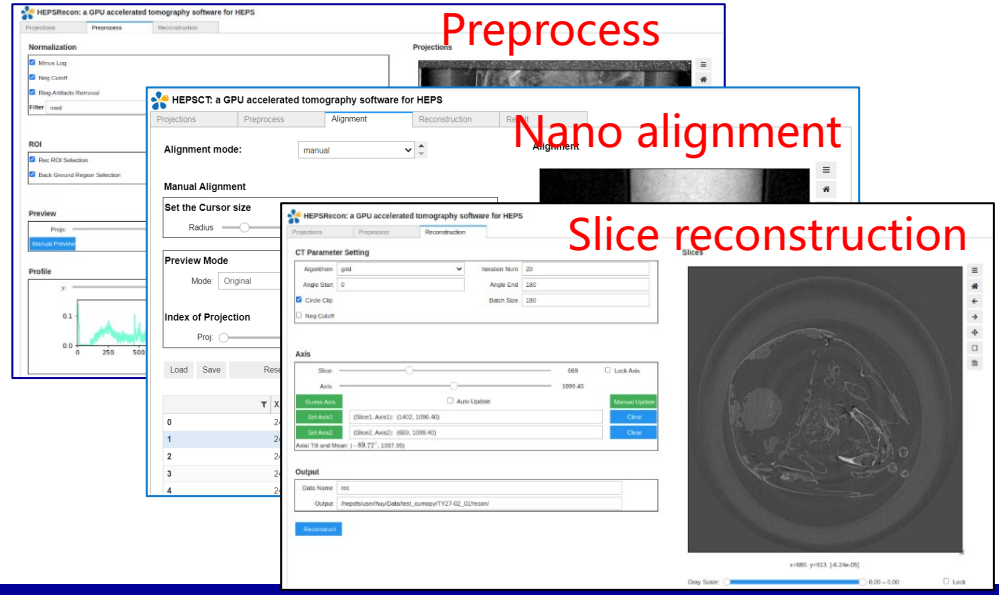
Web data analysis platform:

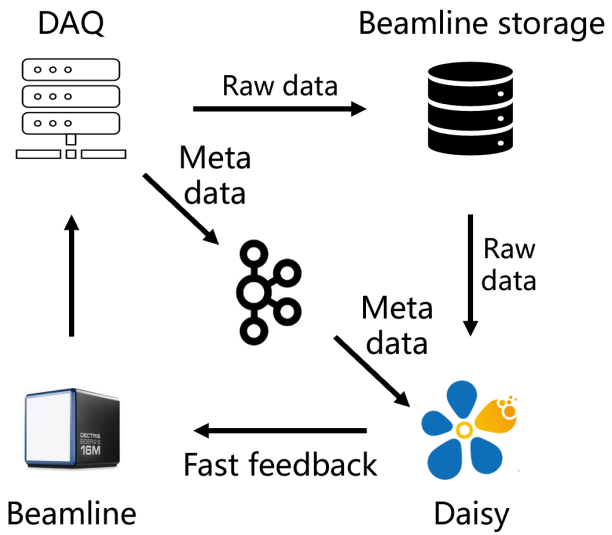
- Based on the jupyterlab ecosystem
- Container encapsulates the computing environment
- Scalable computing resource
- Terminal and web scientific APP

Web based application for X-ray CT

- Web-based interactive data processing interface, integrates self-developed software HEPSCT
- Implement the reconstruction of micro CT and nano CT with different data formats (HDF5，tiff)
- WEB GUI provided for interactive data processing and visualization，will support multiple beamlines of HEPS
- Automated data processing pipeline with DAQ and beamline has been implemented. Validated on the BSRF 3W1A test bed.
- Distributed CT data processing method based on Spark and Ray also has been implemented

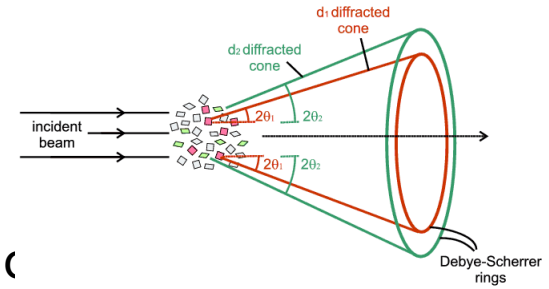


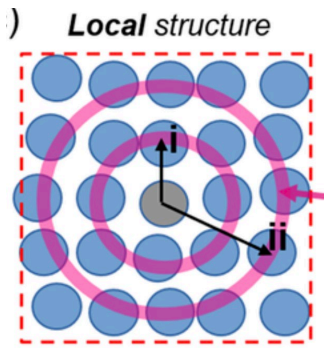


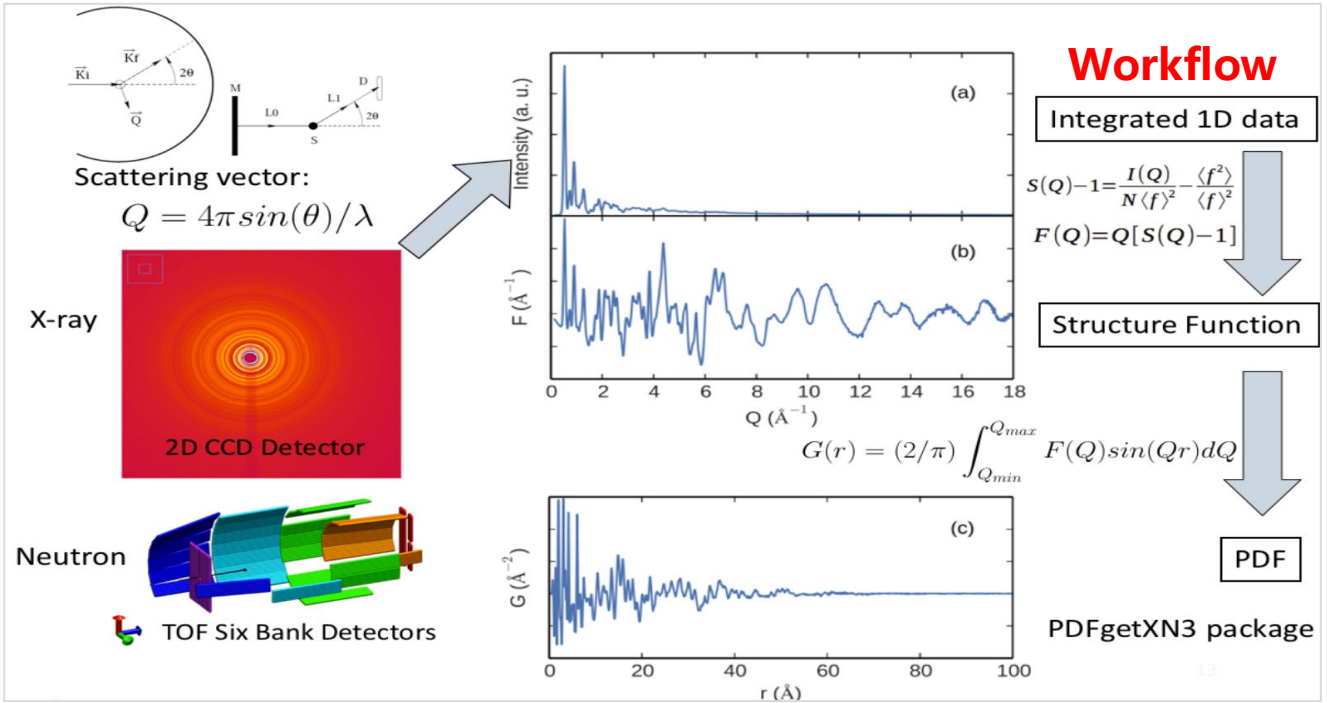


Application for Pair distribution function(PDF)

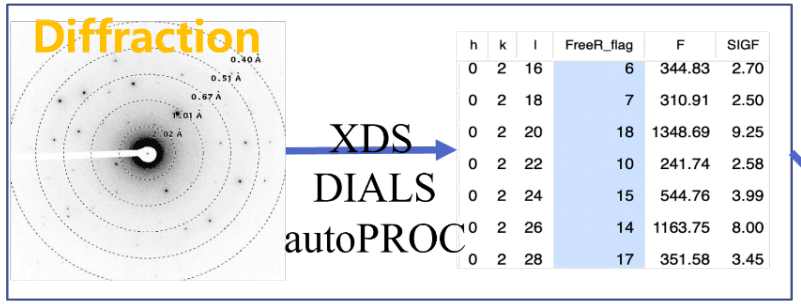
- Serve for total scattering experiment, pipeline from raw data to PDF
- Developed PDFHEPS python package, integrated several X-ray scattering scientific software, such as PyFai, PDFgetX3 and LiquidDiffract
- Web GUI is provided for interactive data processing and visualizatic





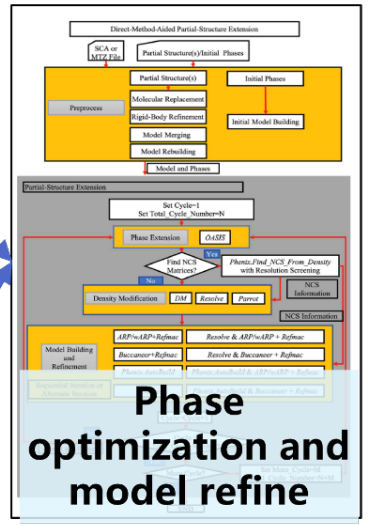


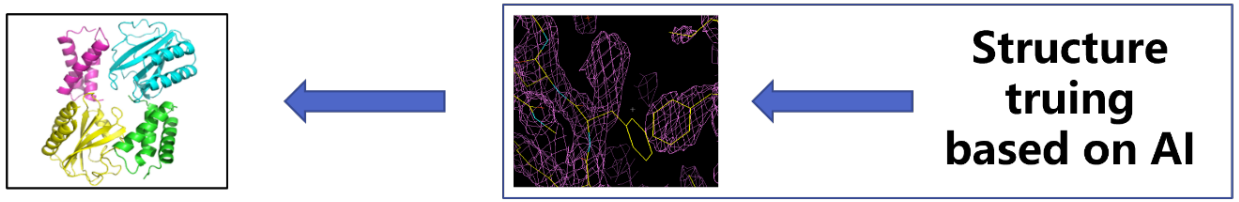
AI-based application for biological macromolecule



Real-time data processing

Structure prediction based on AlphaFold2



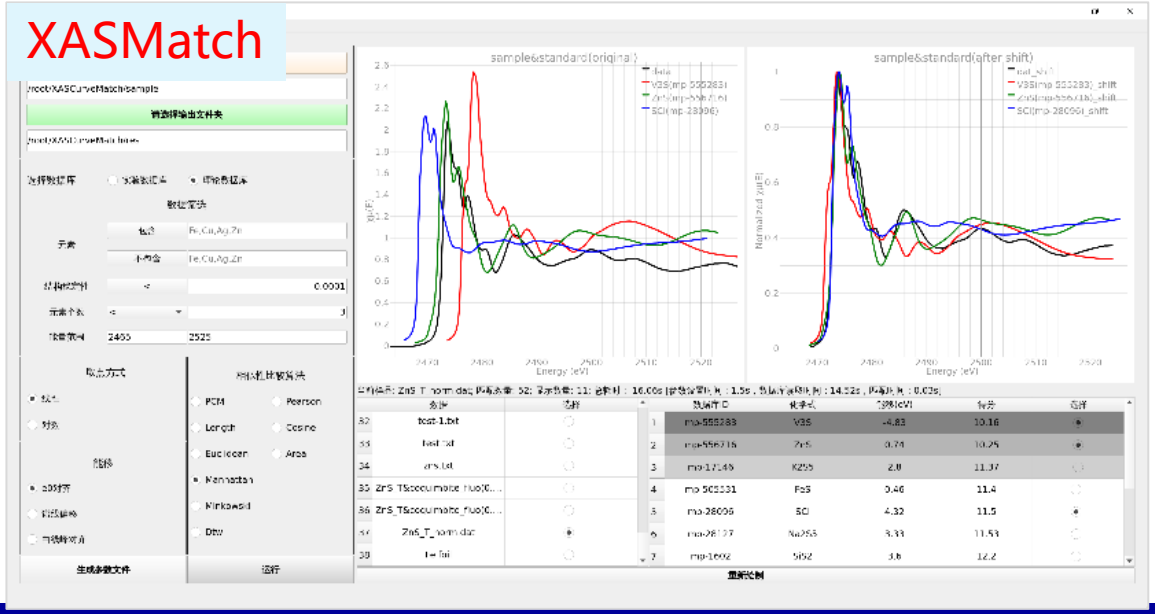


- Automatic pipeline from diffraction to structure
- Web GUI offering real-time data processing status monitoring and result query
- Based on alphafold2, the success rate and accuracy of macromolecular structure reconstruction get improved

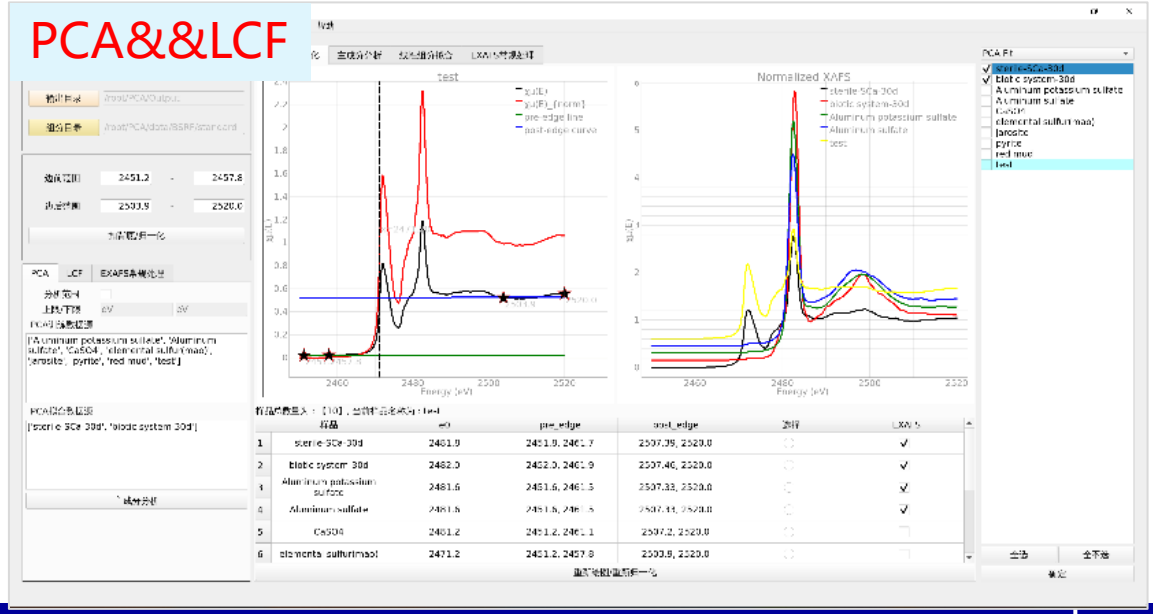
Applications for X-ray absorption spectroscopy

- Used for spectroscopy components analysis, based on PyQt5
- XASMatch
  - Fast matching of experimental spectroscopy from database
  - Automation processing pipeline based on streaming is also in development, demo is ready
- PCA&&LCF
  - Spectroscopy components analysis via PCA and LCF method
  - Automatic pipeline, batch processing, multi-standard spectroscopy input

XASMatch

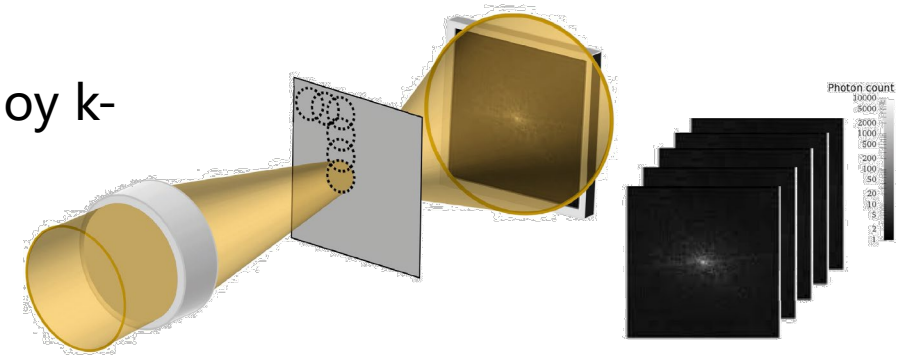


PCA&&LCF



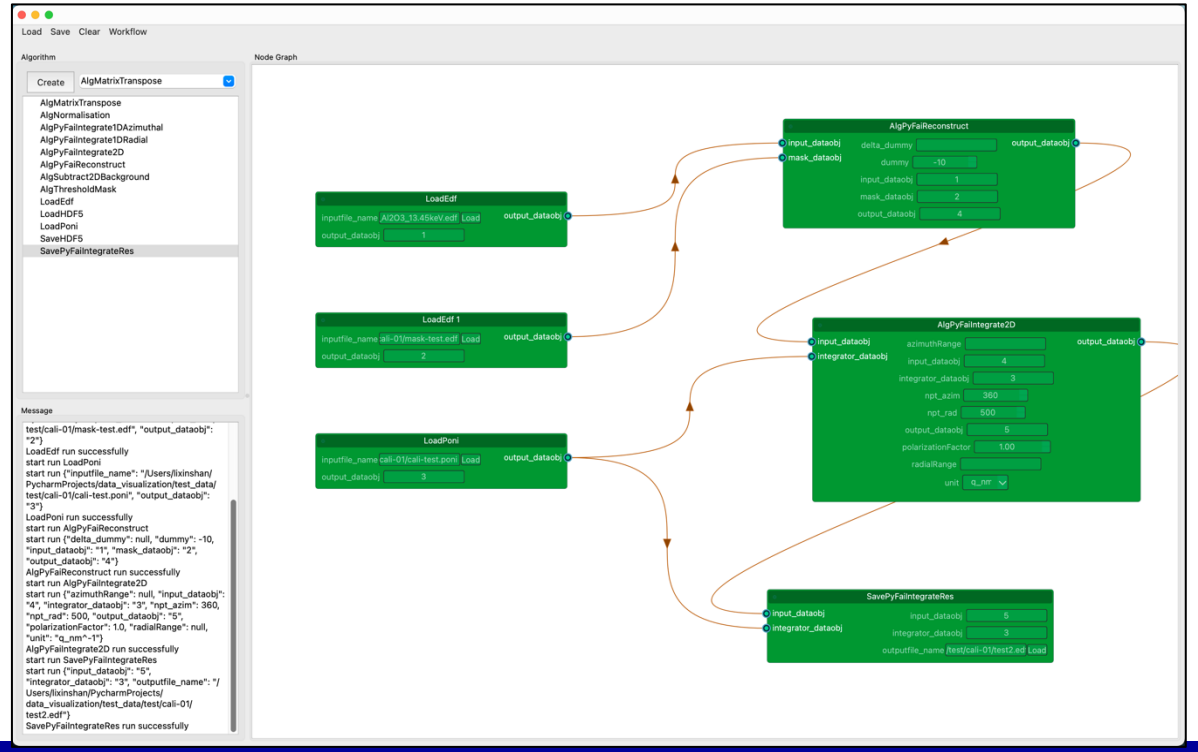
Application for X-ray ptychography

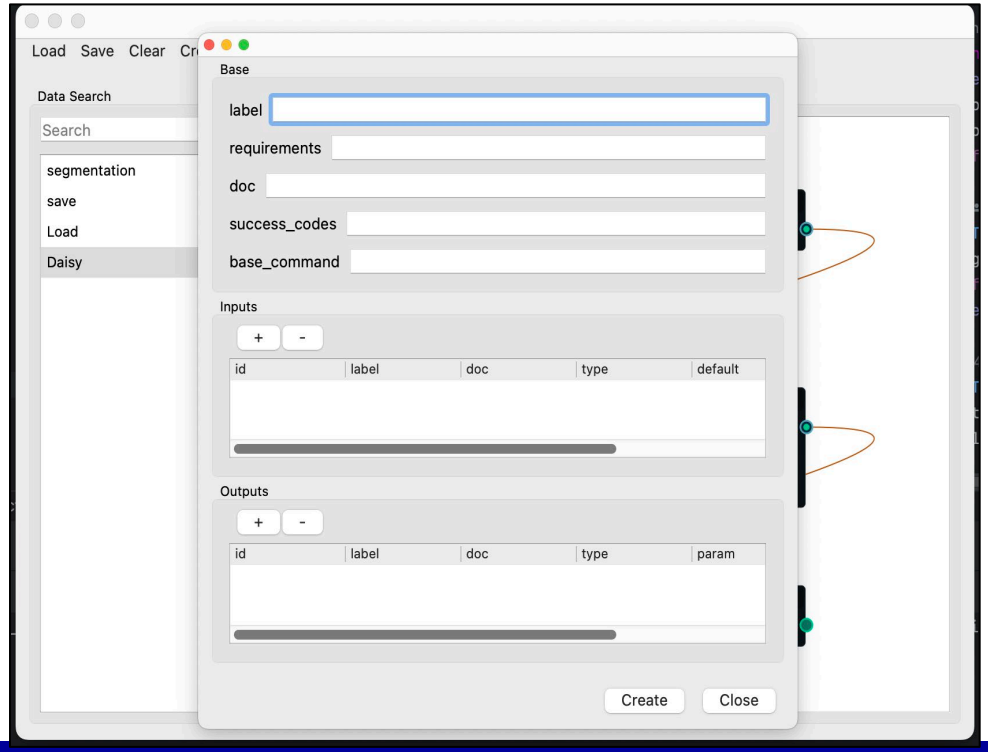
- Supports various phase retrieval algorithms such as ePIE and DM
- Supports multi-GPU parallel processing for large-scale data, employ k-means to partition data
- Migration on Rocm GPU are also in progress
- An AI-based fast online data processing method is also being developed to optimize the DAQ



Workflow management system

- For flexible and general data process task
- GUI support interactive workflow creating, import, export and operation monitoring
- Follow the Common Workflow Language(CWL) standard
- Daisy workflow：automatically parsing algorithms into nodes, execute, monitor, visualization
- Common workflow：create nodes for script commands, can execute on multiple CWL platforms





Support for developers and users

Version control

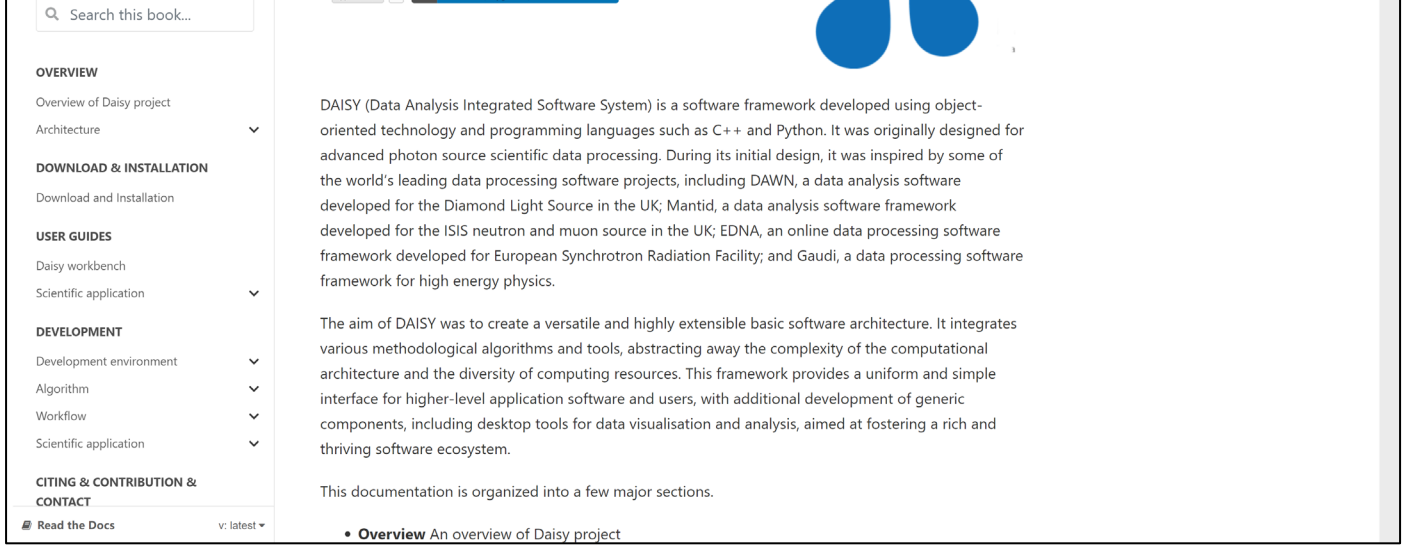
- Git for version control, Gitlab hosted project code, connected CI/CD

Runtime environment

- Container packages the foundational runtime environment
- Harbor manages container images
- CVMFS deploy the pre-compiled software

Documentation

- User documentation，guide for the developer
- Based on Jupyterbook，Sphinx，readthedocs
- Doxygen generate documentation from source code

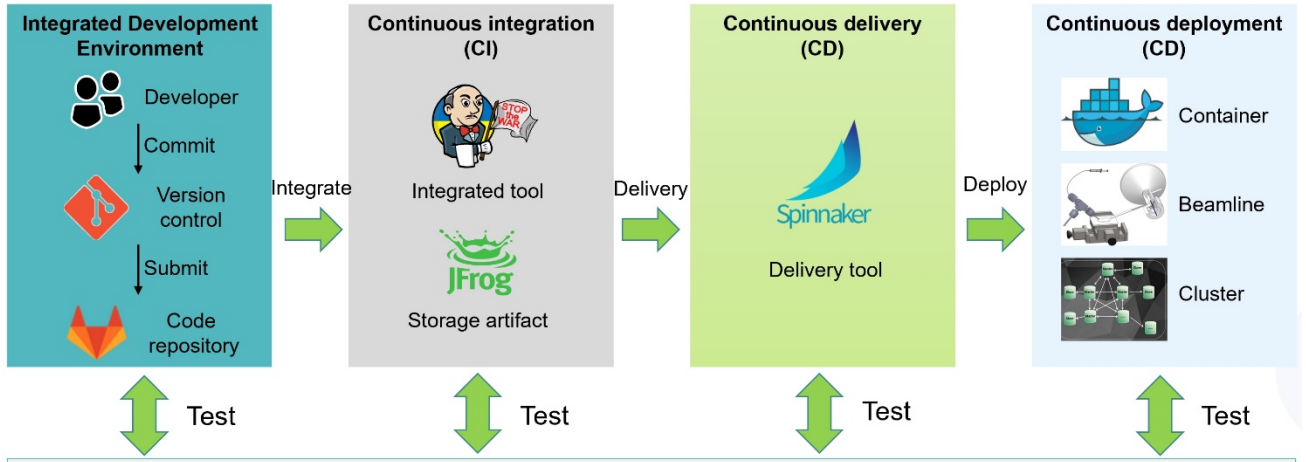


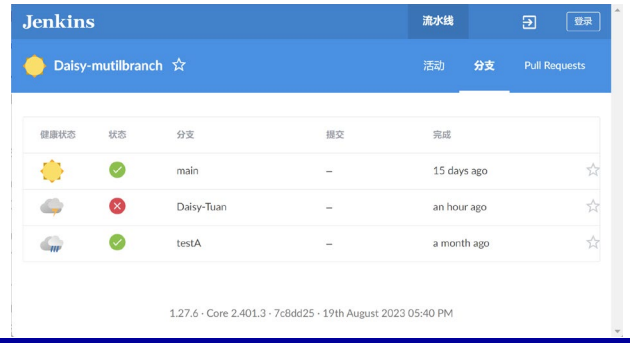
https://daisy.ihep.ac.cn/

Support for developers and users

Continuous integration, delivery, deployment (CI/CD) for software development

- Automated pipeline for software integration, building, test, delivery and deploy
- Continuous monitoring for each stage throughout the lifecycle of software
- Continuous testing to ensure the quality of the codes
- Enables incremental code changes from developers to be delivered quickly and reliably to production
- Based on gitlab, Jeckens, Pytest, PyUnit, and Allure. Some modules are already in production





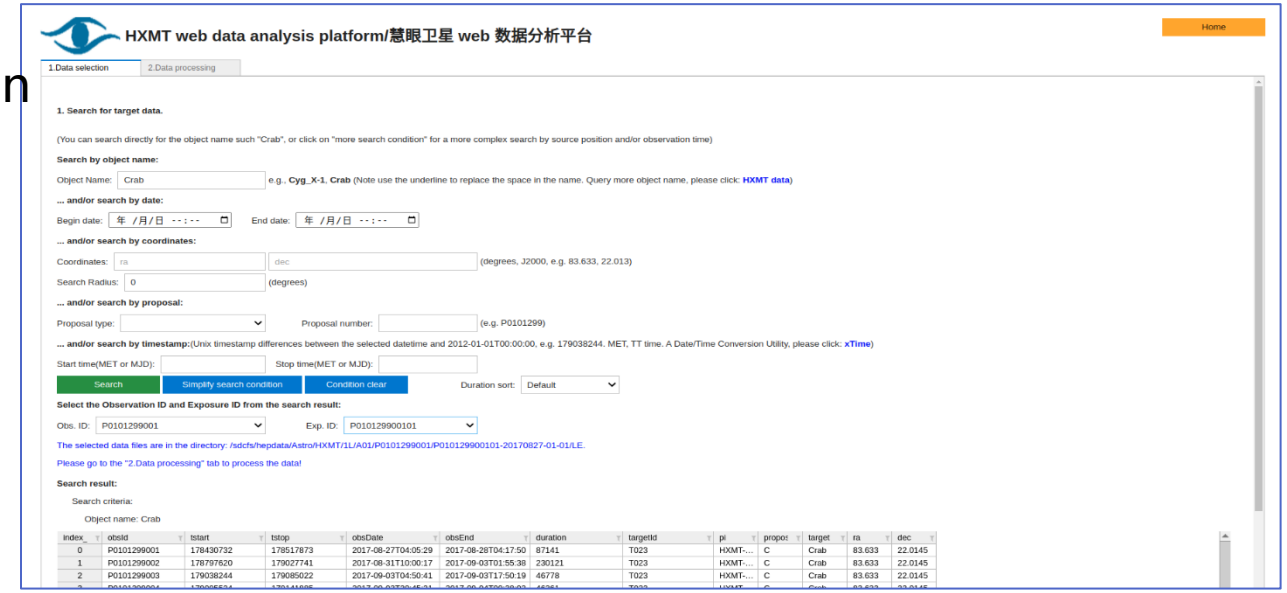
Application of Daisy in space astronomy

Possible application scenarios

- Data processing, analysis, and product generation
- Detector simulation, observation simulation
- Integrate existing software resources to form common software packages

Web based HXMT data processing platform

- Based on Jupyterlab, Docker, K8s
- Provide data processing environment and services via web browser



https://sdccompute.ihep.ac.cn/

Svom and eXTP data processing software based on Daisy are also in processing

- Integrated the I/O algorithms of fits files
- Some data product generation algorithms have been integrated into the Daisy framework

HEPS CC system integration/Test bed/Production

Set up testbed, integrate full data lifecycle software systems to verify the system interfaces, run in the real experimental environment, move to production gradually.

1

Oct, 2020, BSRF 1W1A

Simple verification of the data management system

- Network bandwidth is 1Gb/s
- Beamline storage: **2TB** NAS, Dell EMC NX3240, NFS file system
- Central storage: **80TB** disk array, Lustre file system
- Metadata ingest, catalogue, data transfer, data service

2

July, 2021, BSRF-3W1 test beamline

- Network bandwidth updated to 10Gb/s
- Beamline storage & Central storage: **80TB** disk array, Lustre file system
- Integrate **MAMBA, DMS, Daisy, computing system**

3

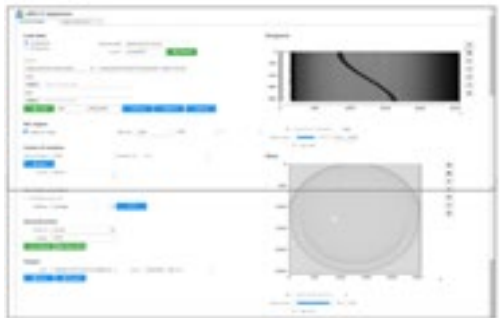
July, 2023, BSRF 4W1B/1W1A/4W1A

Running in production environment

- Network bandwidth updated to 25Gb/s
- Beamline storage: Huawei Ocean Store 9950
- Central storage: 80TB disk array, Lustre file system
- **Follow real experiment process, provide Pymca to do analyzing**

Data acquisition

Analysis framework Interface



CT reconstruction

Integration test at BSRF

Outline

1. Introduction
2. Demand and Challenges of scientific data and software system
3. The architecture and design of the framework
   a. Data management software framework
   b. Data analysis software framework
4. **Summary**

Summary

- The system design has been finished
- Cooperation with other facilities and community is ongoing
- The basic framework has been stable and tested on the test bed
- Based on the framework, scientific software integration and application development are ongoing
- The development of scientific software ecosystem also needs the support and participation of user community

Thanks !

Back up

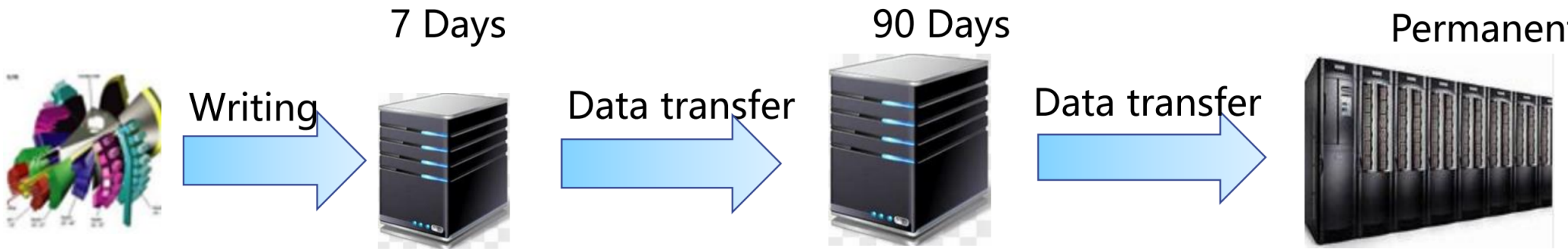
Data storage policy

① **Raw data produced from detector are saved to beamline storage directly，up to 7 days**

② **Data is moved from beamline storage to central storage，data are kept up to 90 days**

③ **Data is moved from central storage to tape for long-term storage**

Data storage policy will be adjusted according to the actual data volume and funding situation



Detector DAQ system

Beamline storage

- SSD
- High speed data IO

Central storage

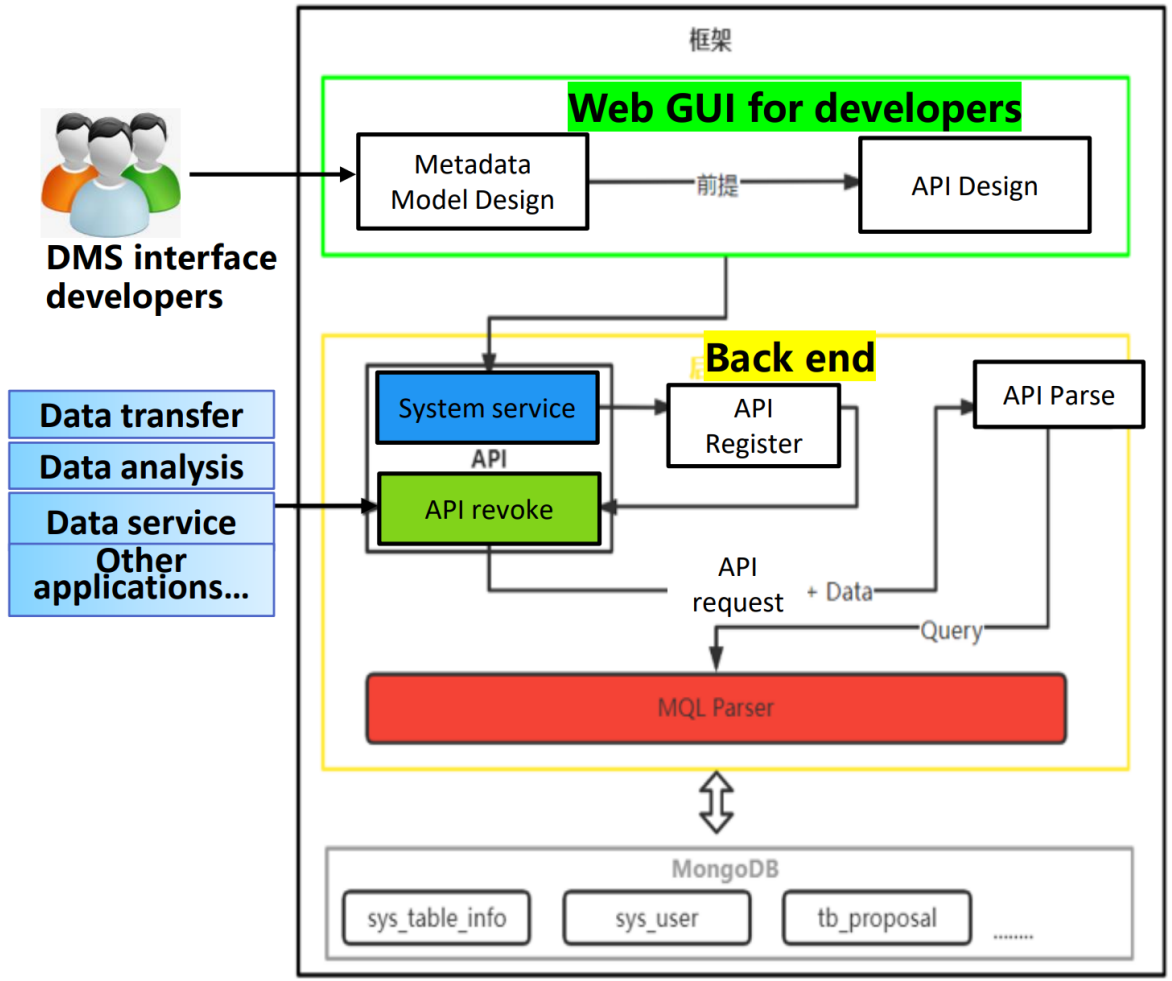
- HDD
- Medium-high speed data IO

Tape

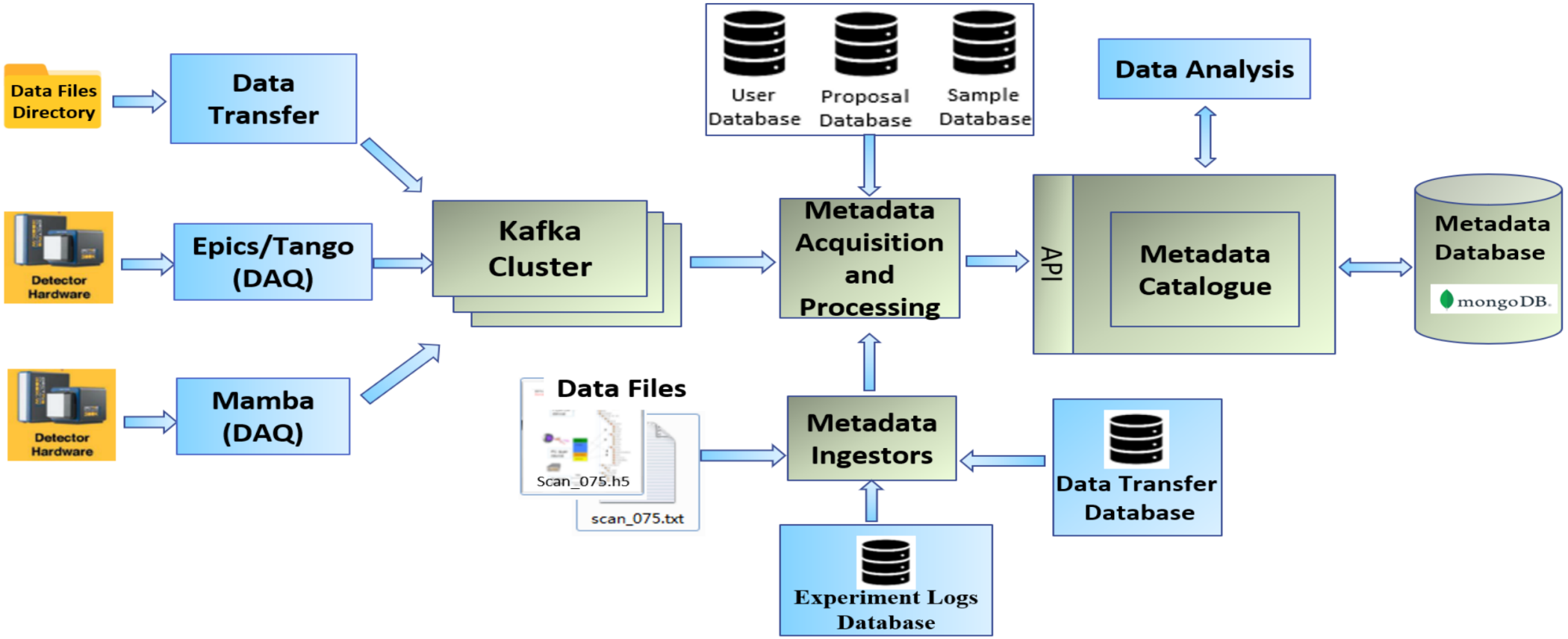
- Permanent data archive
- Prepare the data before retrieve

Metadata catalogue

- **Stores metadata into database and provides APIs to access metadata**
- **Use MongoDB as the database because of the complicated metadata**
- **A tool is developed to generate RESTful API automatically from metadata models**
  1. Interface developer design metadata models and create interfaces from web GUI
  2. The metadata models and interfaces can be parsed, verified and processed
  3. The APIs can be revoked by other system/modules



Metadata acquisition



Kafka cluster

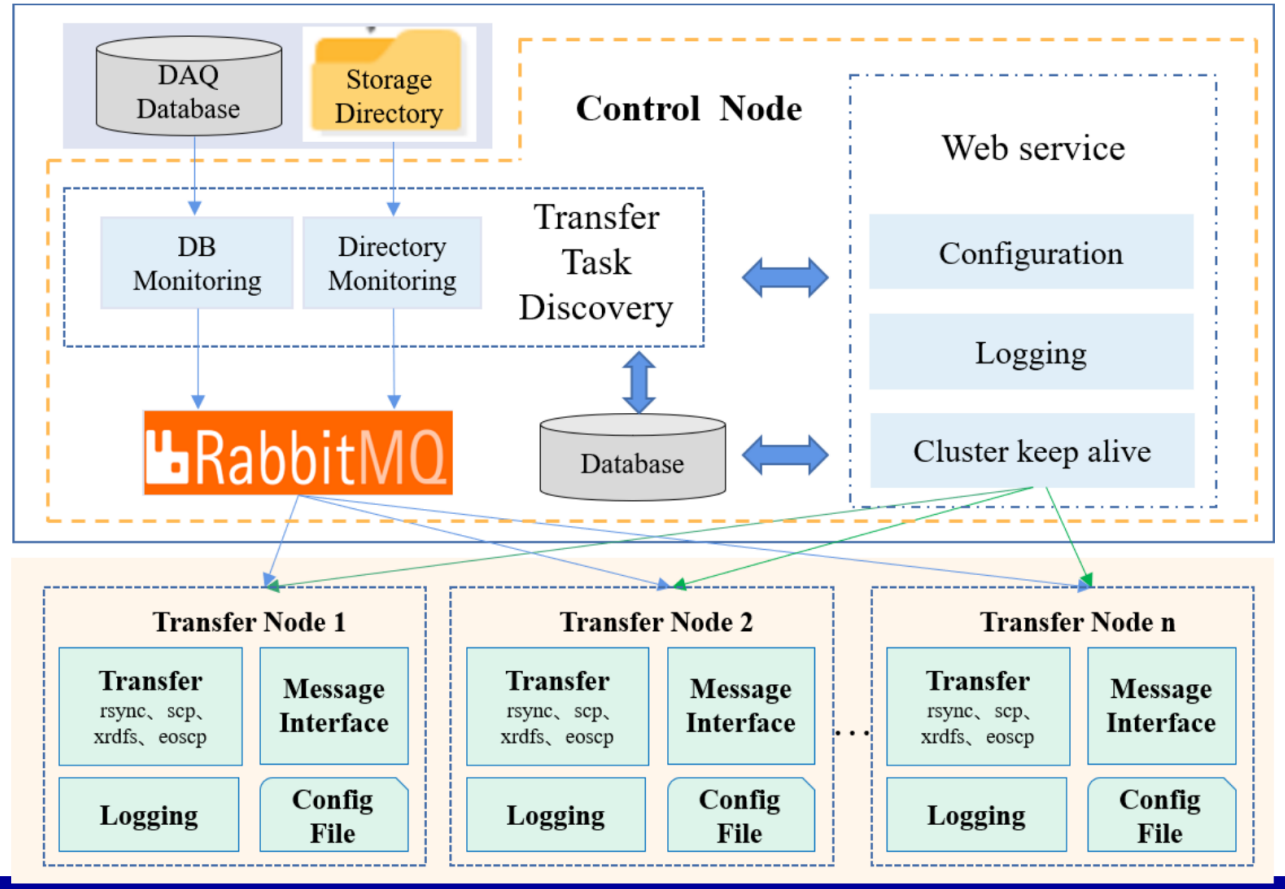
- Acquire metadata from multi-sources
- High reliability for metadata acquisition

Metadata ingestor plugins

- flexible development and deployment
- Make up for lack of interfaces

Data transfer

- Cluster deployment: control node + transfer node
- Control node: transfer task discovery, RabbitMQ, web service for configuration, Logs and cluster management
- Transfer node： transfer, logging, message interface



Features

1) Transfer task discovery

   Directory monitoring, Database polling

2) Transfer

   - Transfer protocols: rsync/scp/xrdfs/eoscp
   - Cluster deployment and multithreaded transfer
   - Checksum validation, retransmission

3) Configuration

   Transfer task discovery, interface, logs, cluster…

4) Logging and monitoring

   log information related to transfer failures and exceptions

An universal data transfer software, is used in other experiments (JUNO, LHAASO…)

Data service (1)-- Dedicated computer terminal

The dedicated computer terminal for data downloading

- placed at user center/user lounge
- suitable for downloading huge volume of data
- supports different storage device interfaces (NAS, disk array, mobile hard disk)

Dedicated computer terminal for on-site data downloading

Data service (2)--Data Web Portal

Data search, download, access, analysis

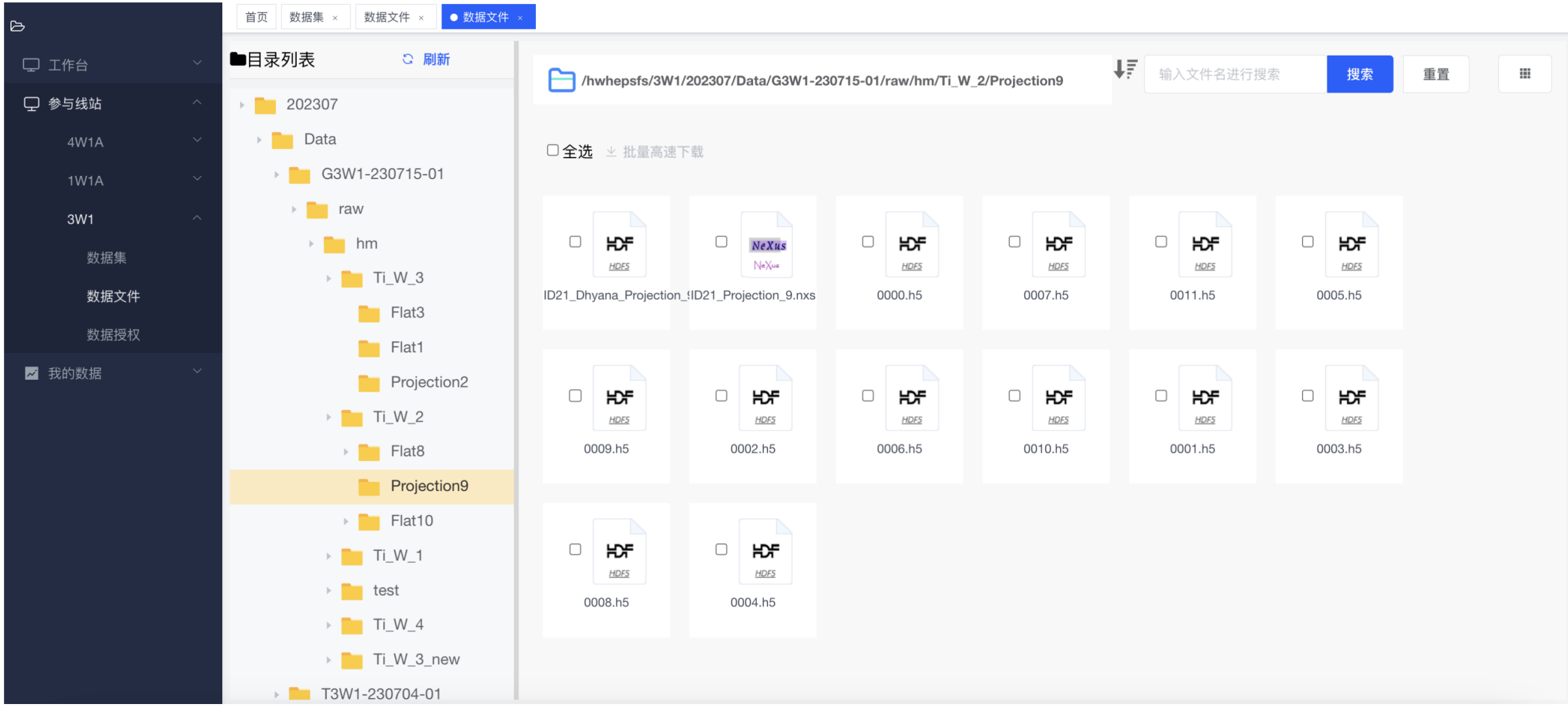
- **Dataset search through metadata retrieval**

Beamtime ID: BeamtimeID | Proposal ID: Proposal ID | Proposal: Proposal Name | PI: PI Name | Start Time: Start Time - End Time | Search | Reset

Beamtime ID	proposal ID	PI	PI Email	Start Time	End Time	Dataset Count	Actions
G4W1A-230723-01	2023-BEPC-PT-008641			2023-07-23 08:00:00	2023-07-25 08:00:00	1	View Dataset
G4W1A-230721-01	2022-BEPC-PT-007389			2023-07-21 08:00:00	2023-07-23 08:00:00	2	View Dataset
G4W1A-230720-01	2023-BEPC-PT-007867			2023-07-20 08:00:00	2023-07-21 08:00:00	5	View Dataset
G4W1A-230718-01	2021-BEPC-PT-006169			2023-07-18 08:00:00	2023-07-20 08:00:00	8	View Dataset
G4W1A-230717-01	2021-BEPC-PT-005200			2023-07-17 08:00:00	2023-07-18 08:00:00	2	View Dataset
G4W1A-230716-01	2023-BEPC-PT-008308			2023-07-16 08:00:00	2023-07-17 08:00:00	2	View Dataset
G4W1A-230714-01	2023-BEPC-PT-008530			2023-07-14 08:00:00	2023-07-16 08:00:00	1	View Dataset
G4W1A-230713-01	2022-BEPC-PT-007658			2023-07-13 08:00:00	2023-07-14 08:00:00	1	View Dataset
G4W1A-230711-01	2023-BEPC-PT-008278			2023-07-11 08:00:00	2023-07-13 08:00:00	1	View Dataset
G4W1A-230710-01	2023-BEPC-JZ-00018			2023-07-10 08:00:00	2023-07-11 08:00:00	1	View Dataset
G4W1A-230709-01				2023-07-09 08:00:00	2023-07-10 08:00:00	1	View Dataset
G4W1A-230707-01	2022-BEPC-PT-006548			2023-07-07 08:00:00	2023-07-09 08:00:00	1	View Dataset
G4W1A-230706-01	2023-BEPC-JZ-00006			2023-07-06 08:00:00	2023-07-07 08:00:00	1	View Dataset
G4W1A-230705-01	2023-BEPC-PT-008244			2023-07-05 08:00:00	2023-07-06 08:00:00	1	View Dataset
G4W1A-230704-01	2023-BEPC-PT-008244			2023-07-04 08:00:00	2023-07-05 08:00:00	0	No Data
G4W1A-230703-01	2023-BEPC-JZ-00006			2023-07-03 08:00:00	2023-07-04 08:00:00	1	View Dataset
G4W1A-230702-01	2023-BEPC-JZ-00018			2023-07-02 08:00:00	2023-07-03 08:00:00	1	View Dataset

Data service (2)-- Data Web Portal

- **Browse and download data files in storage**



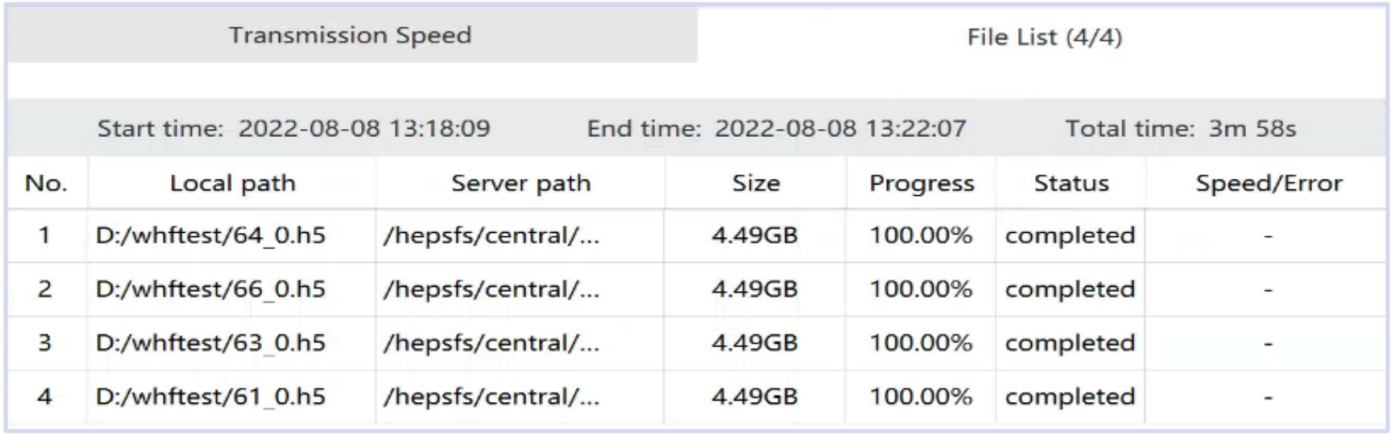
Data service (2)-- Data Web Portal

Integrate a client for downloading

- Maximizes the utilization of network bandwidth
- Greatly improves the download speed

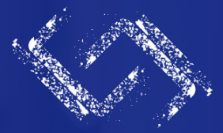
- Download speed test： 4*4.49GB files

	Bandwidth	Duration	Speed
LAN	1000Mbps	2min52sec	839Mbps
WAN	100Mbps	28min	85Mbps



国家高能物理科学数据中心

National HEP Science Data Center



高能所計算中心

IHEP Computing Center

HEPS-CC