



大科学装置数据技术讨论会 2020.10.19

上海科技大学

SHINE数据系统总体介绍

Ping HUAI (怀平)

SHINE束线站总体·实验辅助分总体

上海科技大学

SHINE

Outlines



- **Overview of SHINE**
- **Data Requirement of SHINE BE**
- **Data System Design**
- **R&D Progress**

Outlines



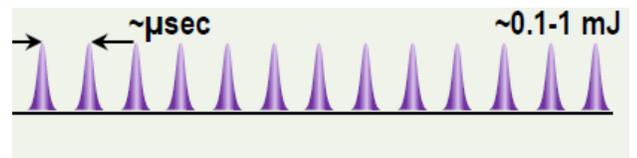
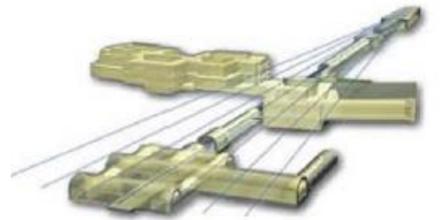
- **Overview of SHINE**
- Data Requirement of SHINE BE
- Data System Design
- R&D Progress

硬X射线自由电子激光装置 (SHINE)



- 国家重大科技基础设施 (建设期2018.4-2025.4) , 总投资: 建设投资95亿; 科技重大专项投资7.7亿
- 建设一台能量8GeV的超导直线加速器、产生0.4-25 keV光子能量范围的3条自由电子激光波荡器线、3条光束线; XFEL脉冲的重复频率可达1MHz、超快脉冲小于10飞秒
- 建设首批10个实验站, 具备纳米级的超高空间分辨能力和飞秒级的超快时间分辨能力

硬 X 射线自由电子激光装置航拍影像图



新的激光技术, **单脉冲可调控的光源**

超快、分子电影...Pathfinder of new technologies

每一个脉冲都是一个实验, **前所未有的操控实验的能力**。全新的实验和数据处理模式。

拍摄时间: 2017年12月

比例尺 1:1000

上海科技大学
嘉兴博海信息科技有限公司 制

SHINE

Main Parameters @ SHINE



	Nominal	Range	Unit
Beam energy	8	4-8.6	GeV
Bunch charge	100	10-300	pC
Max repetition rate	< 1	up to 1	MHz
Electron beam power	0.8	0 - 2.4	MW
Photon energy	0.4-25	0.4-25	keV
Pulse length	20-50	5-200	fs
Peak brightness	5×10^{32}	1×10^{31} - 1×10^{33}	Photons/ $\mu\text{m}^2/\text{rad}^2/\text{s}/0.1\% \text{BW}$
Average brightness	5×10^{25}	1×10^{23} - 1×10^{26}	Photons/ $\mu\text{m}^2/\text{rad}^2/\text{s}/0.1\% \text{BW}$
Total facility length	3.1	3.1	km
Total tunnel length	5.7	5.7	km
Tunnel diameter	5.9	5.9	m
2K Cryogenic power	12	12	kW
RF Power	2.28	3.6	MW

Main FEL Parameters @ SHINE

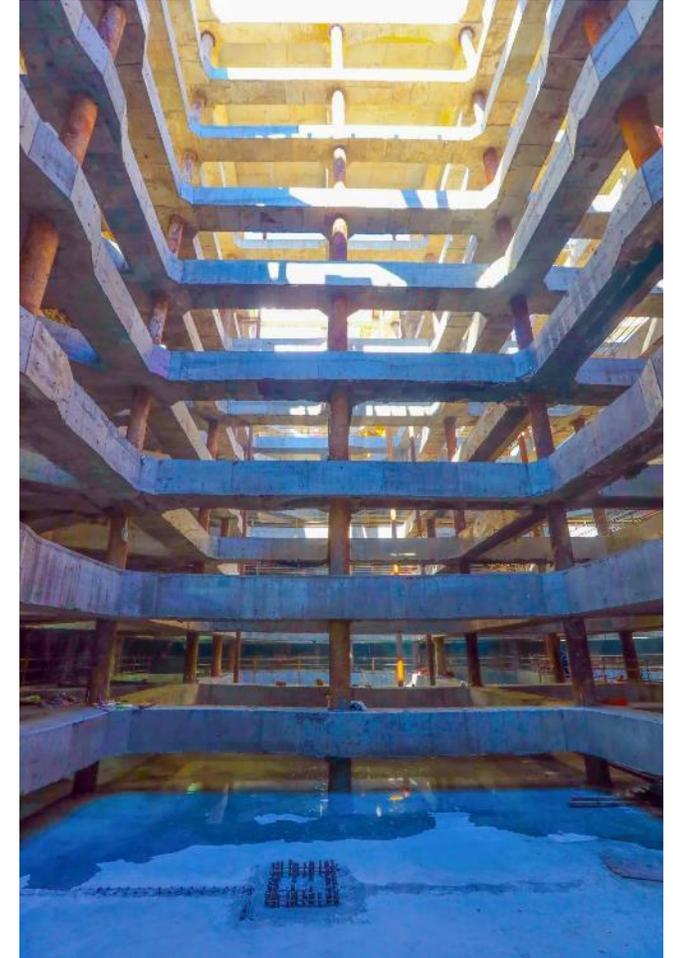


	Nominal	Objective	Unit
FEL-I			
Photon energy	3-15	3-15	keV
Photon number per pulse @12.4keV	$>10^{10}$	$>10^{11}$	
Max pulse repetition rate	0.66	1	MHz
FEL-II			
Photon energy	0.4-3	0.4-3	keV
Photon number per pulse @1.24keV	$>10^{12}$	$>10^{13}$	
Max pulse repetition rate	0.66	1	MHz
FEL-III			
Photon energy	10-25	10-25	keV
Photon number per pulse @15keV	$>10^9$	$>10^{10}$	
Max pulse repetition rate	0.66	1	MHz

Civil Construction (Groundbreaking)



Civil Construction (Shaft 4)



Civil Construction (Shaft 5)

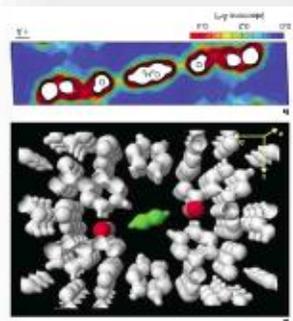


Scientific Motivation

- **探索多学科的核心科学问题**：面向材料、能源、环境、物理与化学、生命及医药等领域的创新研究对高亮度相干X射线光源的需求，利用基于超导技术的XFEL超高的峰值亮度、超短的脉冲和极好的相干性等优异特性，使成像、结构解析与谱学分析这三大类X射线实验技术实现质的飞跃
- **原子尺度上观测体系动态变化**：为科学用户提供前所未有的高分辨成像、先进结构解析、超快过程探索等尖端研究手段；为前沿科学研究在大的时间跨度（秒到亚飞秒）、空间尺度（微米到埃）了解物质体系的构效关系，提供全谱、非线性、多维度、相干的X射线表征物质手段



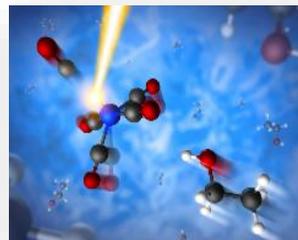
超导材料



纳米材料



燃烧过程



洁净能源



新型药物



蛋白质转变过程



FEL-I Hard X-ray Endstations

- **HSS:** Hard X-ray Scattering and Spectroscopy Endstation
- **CDS:** Coherent Diffraction Endstation for Single Molecules and Particles
- **SEL:** Station of Extreme Light
 - *XFEL + 100 PW Laser System*

FEL-II Soft X-ray Endstations

- **AMO:** Atomic, Molecular, and Optical Science Endstation
- **SES:** Spectrometer for Electronic Structure
- **SSS:** Soft X-ray Scattering and Spectroscopy Endstation

FEL-III Hard X-ray Endstations

- **HXS:** Hard X-ray Spectroscopy Endstation
- **SFX:** Serial Femtosecond Crystallography Endstation
- **CDE:** Coherent Diffraction Imaging Endstation
- **HED:** High Energy Density Science

Main Parameters of SHINE Endstations



实验站名称	设计指标	验收指标	
相干衍射实验站			
分辨率	$\leq 10\text{nm}^*$	20nm @ $> 5 \times 10^{12}$ photons/ μm^2 @ 2keV	
聚焦光斑	1-50 μm	$\leq 5\mu\text{m}$	
原子分子反应成像实验站			
离子动量分辨	$\leq 3\%$	$\leq 5\%$	
单分子/单颗粒相干衍射实验站			
成像分辨率	$\leq 0.5\text{nm}^*$	10nm @ $> 5 \times 10^{13}$ photons/ μm^2 @ 5keV	
聚焦光斑	100nm	$\leq 300\text{nm}$	
物质电子结构实验站			
空间分辨率	$\leq 100\text{nm}$	$\leq 500\text{nm}$	
时间分辨率	$\leq 60\text{fs}$	$\leq 80\text{fs}$	
软X射线高分辨共振散射实验站			
能谱仪能量分辨率	50meV @ 1000eV	100meV @ 1000eV	
串行晶体学实验站			
聚焦光斑尺寸	a. 微米级聚焦串行晶体学系统	$\leq 2\mu\text{m}$	$\leq 4\mu\text{m}$
	b. 纳米级聚焦单颗粒成像与串行晶体学系统	$\leq 0.2\mu\text{m}$	$\leq 0.5\mu\text{m}$

实验站名称	设计指标	验收指标
硬X射线超快谱学实验站		
能谱仪能量分辨率	10^{-4}	10^{-4}
硬X射线高分辨共振散射实验站		
能谱仪能量分辨率	4×10^{-5}	4×10^{-5}

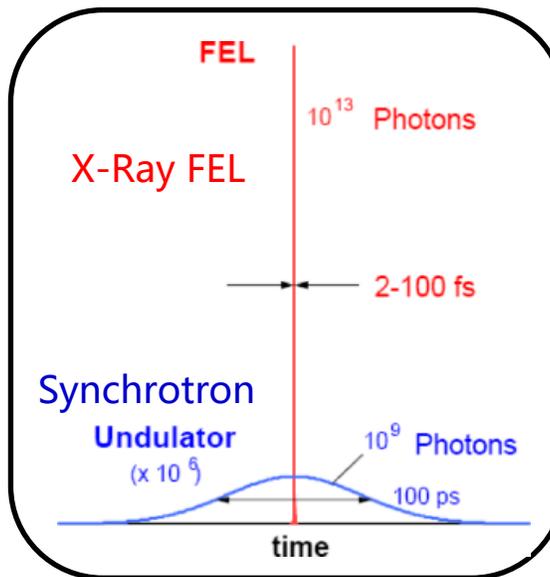
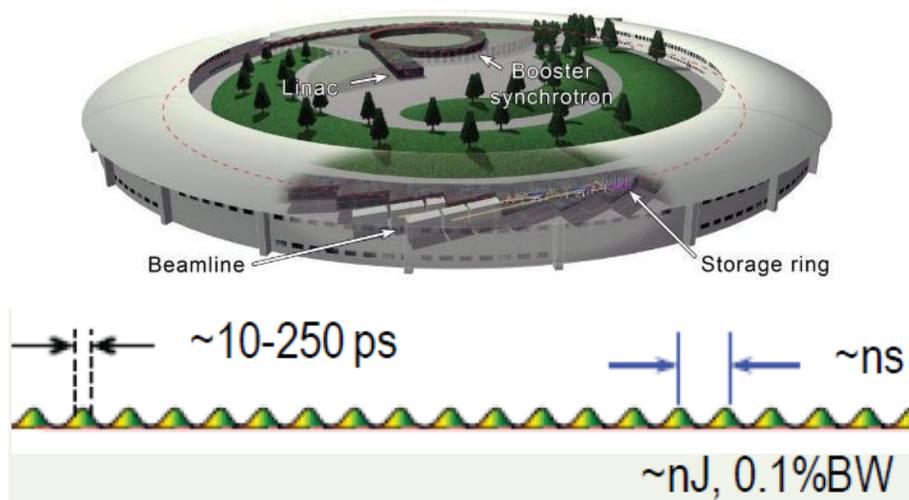
Outlines



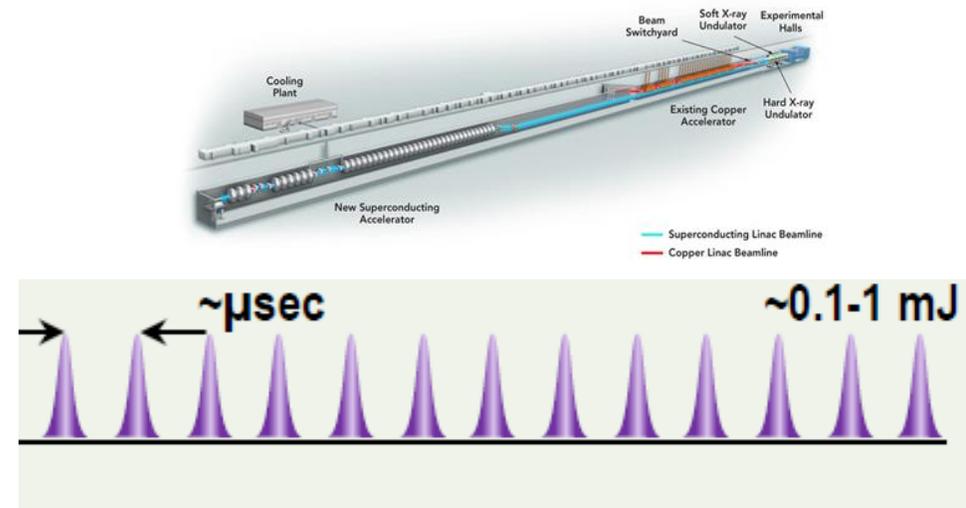
- Overview of SHINE
- Data Requirement of SHINE BE**
- Data System Design
- R&D Progress

Synchrotron Vs. X-Ray FEL

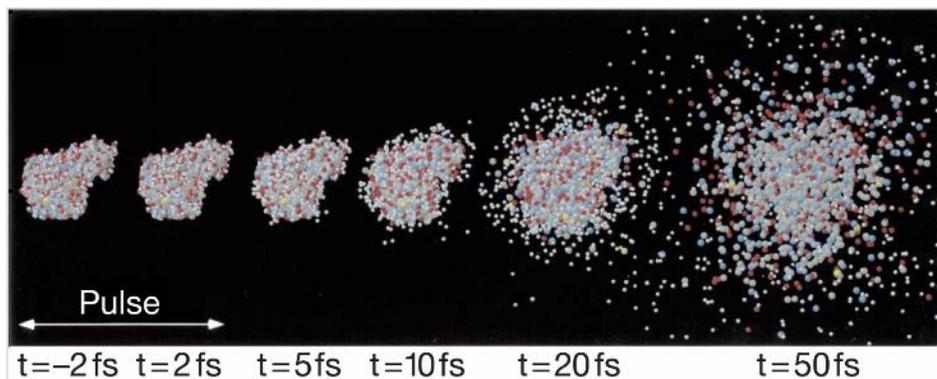
Synchrotron Radiation



X-Ray Free Electron Laser

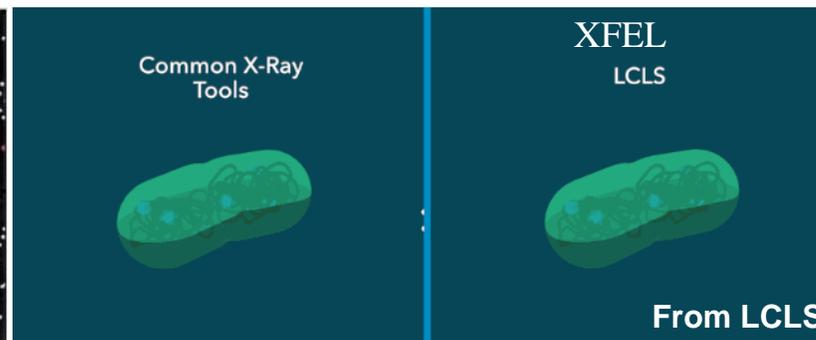


Well Establish Technology-Synchrotron Radiation



SHINE

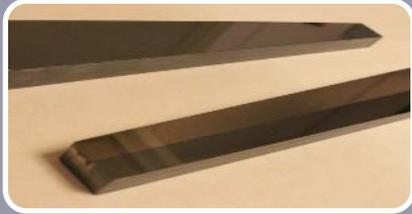
Nature 406, 752-757 (2000)



- ✓ Naval Laser Technology, Single Pulse Control
- ✓ New Experiment Methodology
- ✓ New Data Processing Method

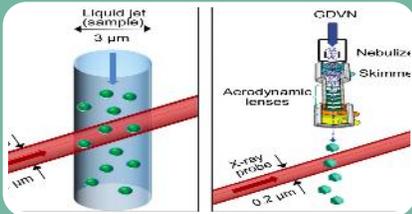
High Brightness + Ultrafast Detection Before Destruction 14

Challenges of SHINE Beamline and Endstation (BE)



X-Ray Optics

withstand high repetition rate; exhibit high accuracy



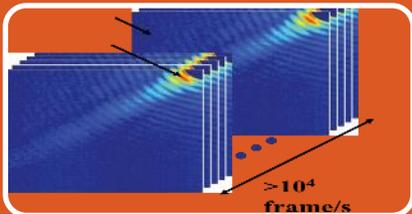
Sample Transport

match high repetition rate; positioning



Detectors

match high repetition rate; high spatial resolution; radiation resistance



Big Data

extreme high data rate; big data management

Big Data Challenge in Photon Science



光子科学大数据的5V特征



科研大数据3H特征



High Throughput Data Processing in LCLS-II



LCLS-II Data Analysis Pipelines: Single Particle Imaging Example

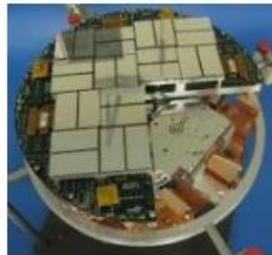
SLAC

Experiment Description

4 μm liquid jet
3 μm X-ray beam
200 μm
Gas dynamic virtual nozzle
LCLS beam

- Individual particles are injected into the focused LCLS pulses
- Scattering patterns are collected on a pulse-by-pulse basis
- Particle concentration dictates “hit” rate

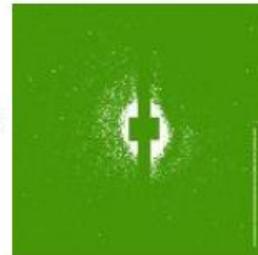
Multi-megapixel detector



60 GB/s
1 TB/s

- 8 kHz in 2024 (4 MP)
- 40 kHz in 2027 (16 MP)

Coherent scattering image



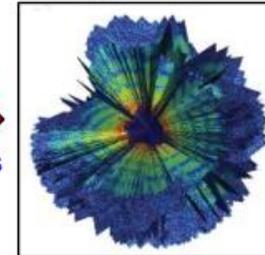
6 GB/s
100 GB/s

Data Reduction

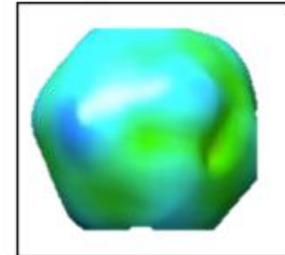
- Remove “no hits”
- >10x reduction

3 TFlops
16 TFlops

Intensity map from multiple pulses



Interpretation of system structure / dynamics

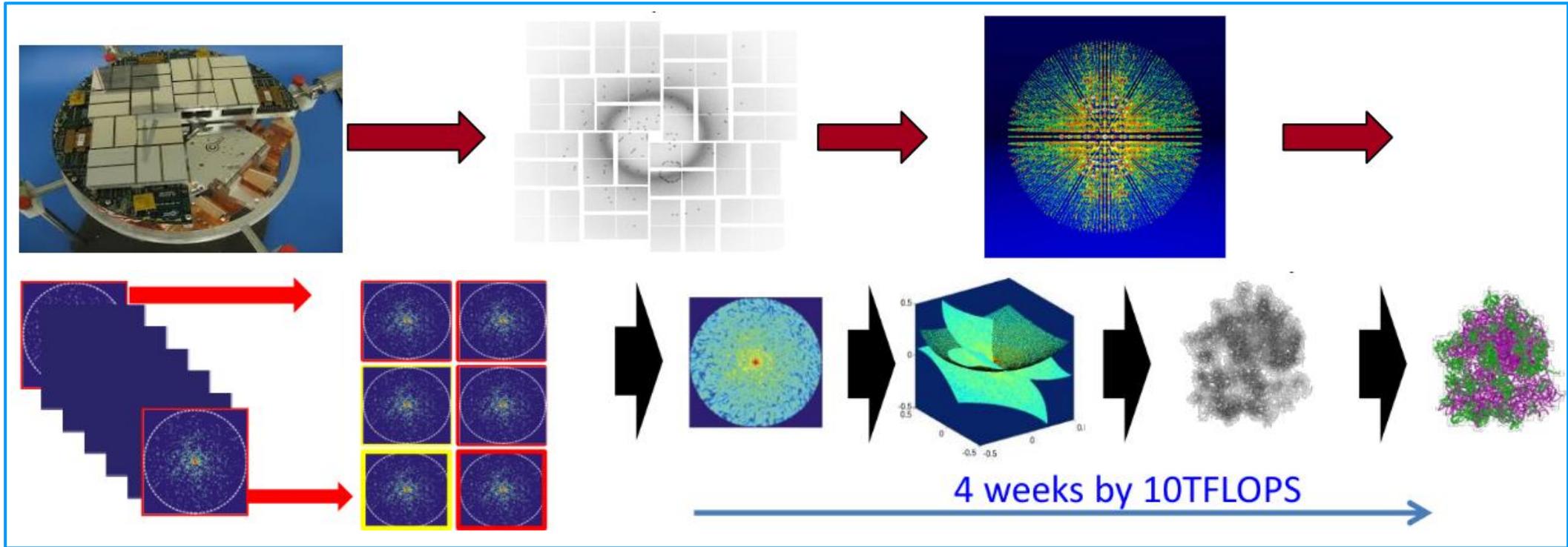


Data Analysis

- Orient patterns
- Average
- 3D intensity map
- Reconstruction

270 PFlops
1340 PFlops

Data Requirement of Coherent Diffraction Imaging (CDI)



重构出一个结构，需要10,000张高清照片，如硅像素面阵探测器的分辨率 $4k \times 4k$ 计算，每个像素点占2 byte，每张照片是32MB，按照0.1%命中率计算，实验需要采集10,000,000张照片，其原始实验数据是320TB。

数据量：320TB/实验，10PB/年

Data Challenge of SHINE



从常温的低重复频率软X射线自由电子激光用户装置演化到高重复频率硬X射线自由电子激光装置，对高速数据采集、传输与大数据存储提出了更高的要求

参数	软线装置	硬线装置@5kHz	硬线装置@100kHz
平均数据率	0.1-1GB/s	2-20GB/s	2GB/s-1.2TB/s
峰值数据率	5GB/s	100GB/s	4.8TB/s
数据存储	3PB	100PB	5EB
数据分析能力	50Tflops/s	1-10Pflops/s	50-100Pflops/s

计算密集型



数据密集型

Data Requirement of FEL-III Endstations



Pulse mode	
1	Burst mode
2	CW mode

devices in DAQ	
1	Fast ADC
2	FPGA
3	...

Main steps in online data analysis	
1	Detector Correction
2	Calibration
3	Event Building

Estimated data throughput for CDI in CDE

Frequency	Recent		Future	
	Front detector	Back detector	Front Detector	Back detector
	1.2 K × 1.2 K	0.5 K × 0.5 K	4 K × 4 K	1 K × 1 K
1 KHz	3 GB/s	0.5 GB/s	32 GB/s	2 GB/s
10 KHz	30 GB/s	5 GB/s	320 GB/s	20 GB/s
100 KHz	300 GB/s	50 GB/s	3.2 TB/s	0.2 TB/s
1 MHz	3 TB/s	0.5 TB/s	32 TB/s	2 TB/s

Estimated data throughput for XPCS in CDE

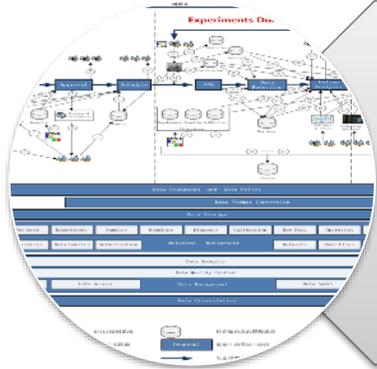
Frequency	5 KHz	100 KHz
Average rate	1 GB/s	20 GB/s
Peak rate	10 GB/s	200 GB/s
Data storage	8 TB/d	160 TB/d
Computing	1 Pflops	20 Pflops
Pixels	1 k × 1 k	

Outlines



- Overview of SHINE
- Data Requirement of SHINE BE
- Data System Design**
- R&D Progress

Mission of SHINE Data System



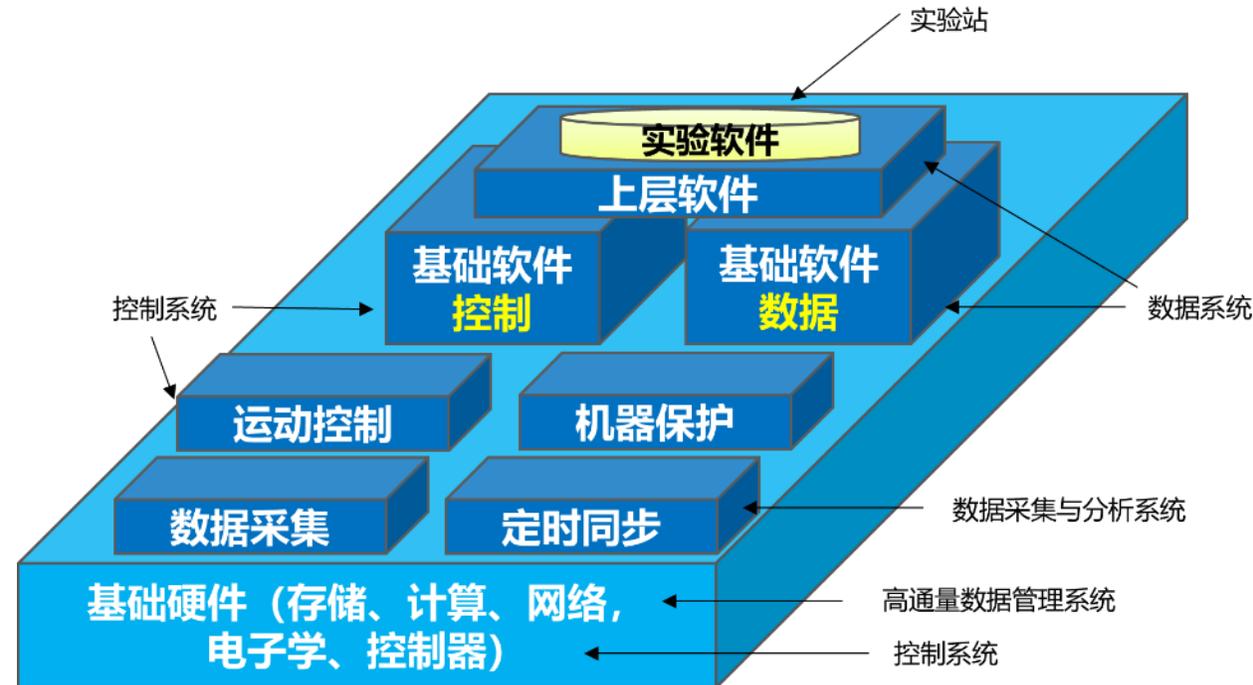
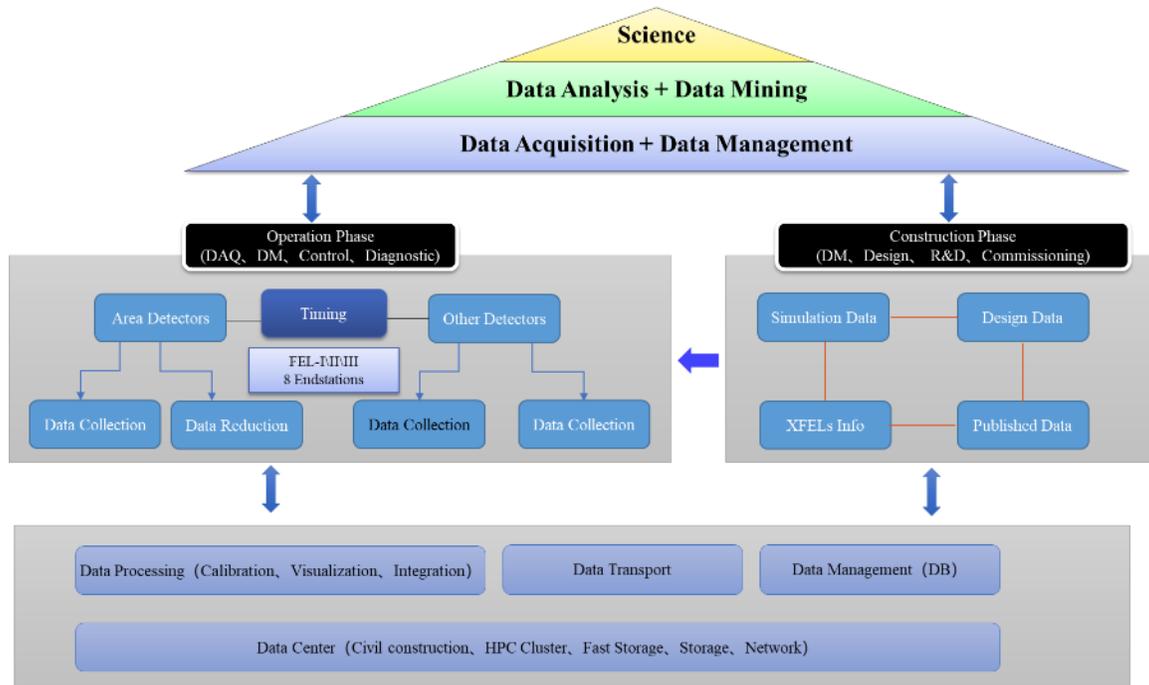
全生命周期数据管理

- 工程设计阶段
- 工程实施阶段
- 安装调试阶段
- 装置运行阶段

数据支持的
工程建设



数据驱动的
光子科学



SHINE Data System Overview



数据触发/筛选系统

高速IT架构

基于AI的数据筛选

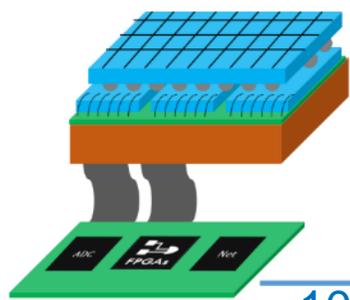
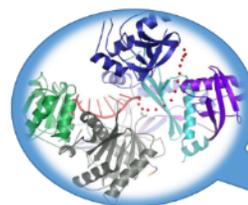
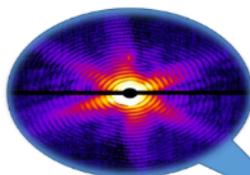
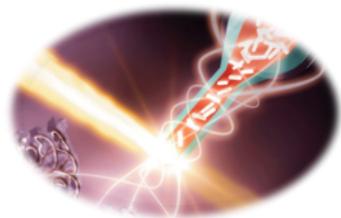
高性能数据库

Detector Calibration

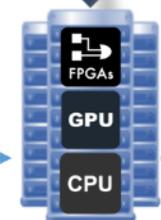
Online Reconstruction

Offline Analysis

Data Sharing



IB
10GbE



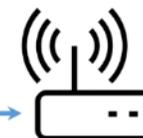
DAQ



Fast feedback



Offline data center



Tape storage

IB network

SHINE Detector

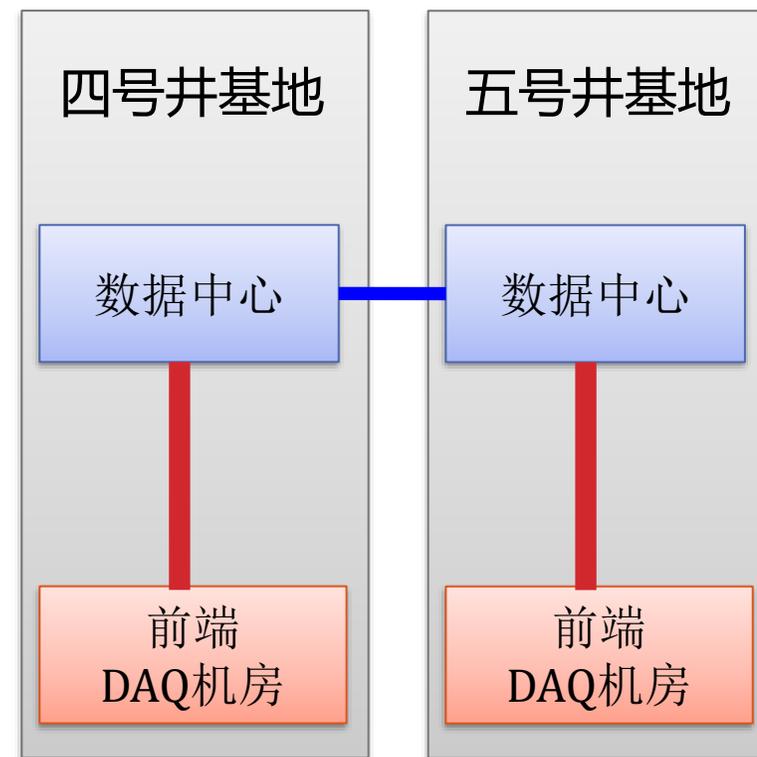
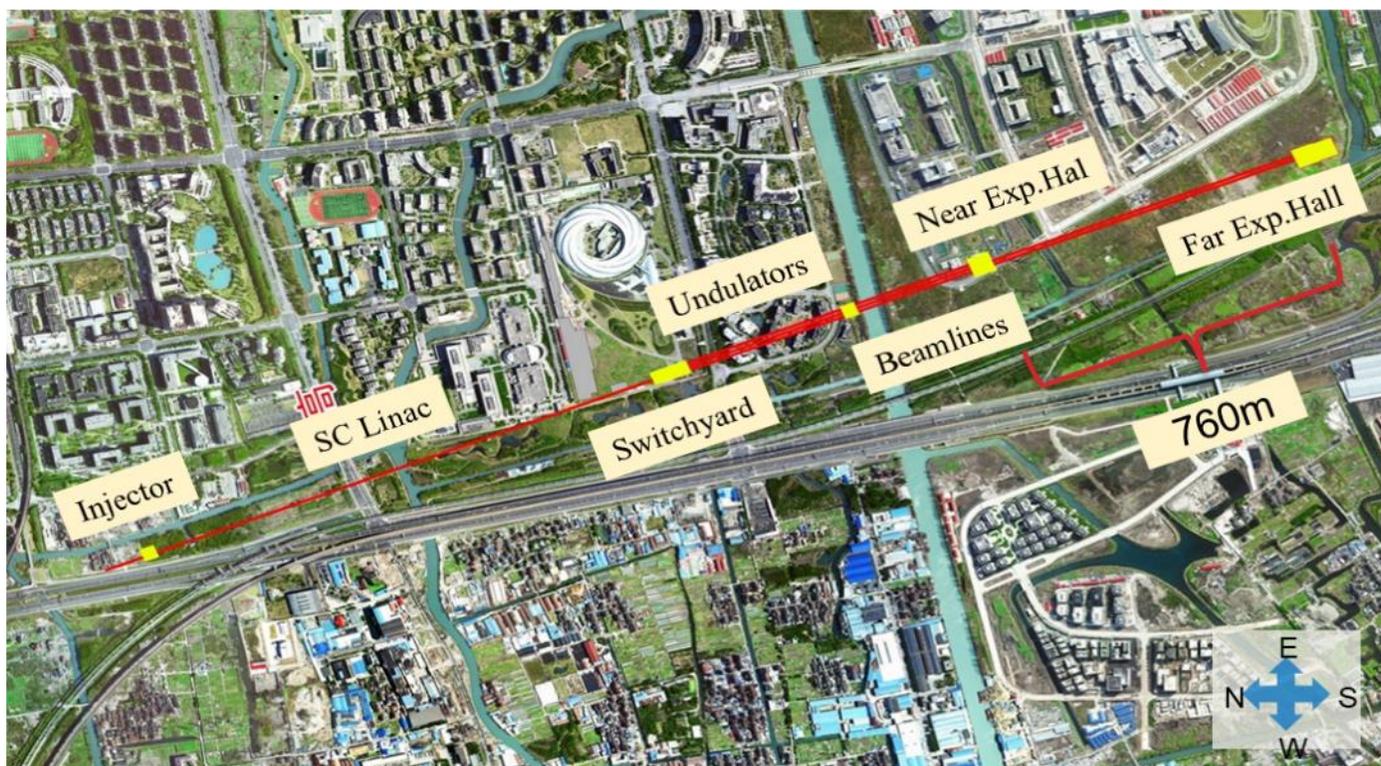
Overview of Data Center @ SHINE

- 能量8GeV的超导直线加速器、0.4-25 keV光子能量、重复频率可达1MHz、超快脉冲小于10飞秒
- 建设首批10个实验站，纳米级空间分辨能力和飞秒级时间分辨能力

总体规划(2017立项时):

存储: 6PB

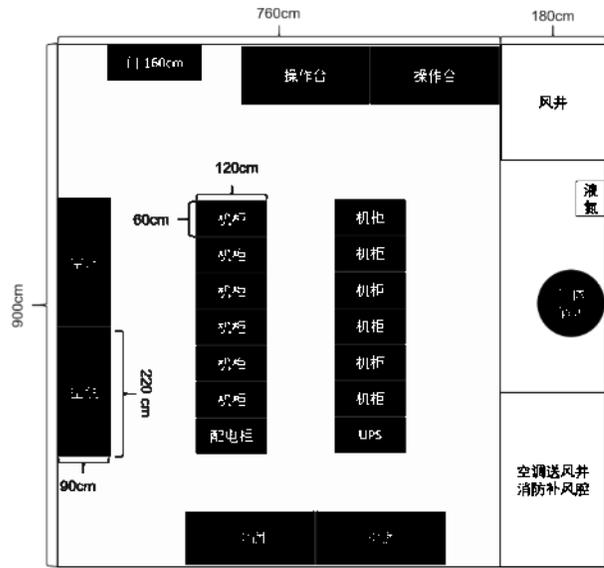
计算能力: 200TFLOPS



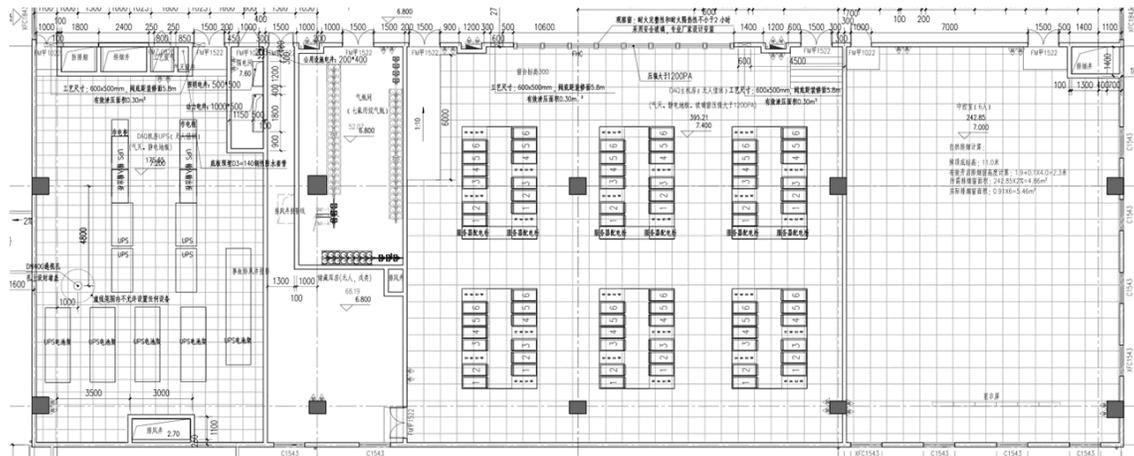
整体规划充分参考CERN、Eu-XFEL、LCLS的建设经验，采用前端高通量在线采集筛选处理和后端离线处理相结合的方式。两个数据中心互为备份，并分别处理两个实验大厅实验的数据处理需求。

Data Center Layout

前端DAQ机房

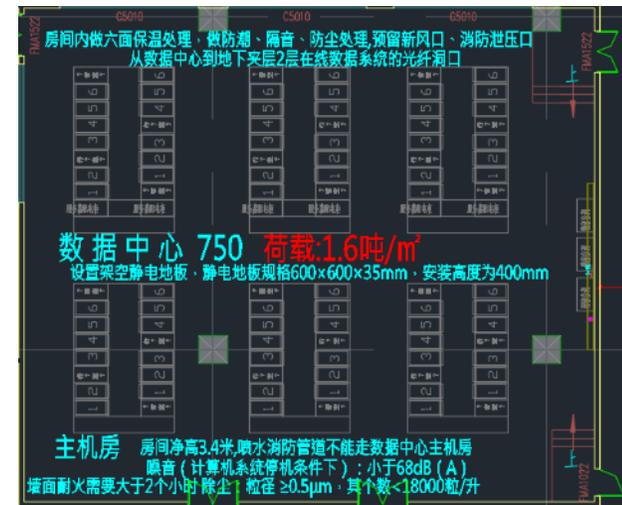


四号井数据中心机房



6个冷池共72台机柜

五号井数据中心机房



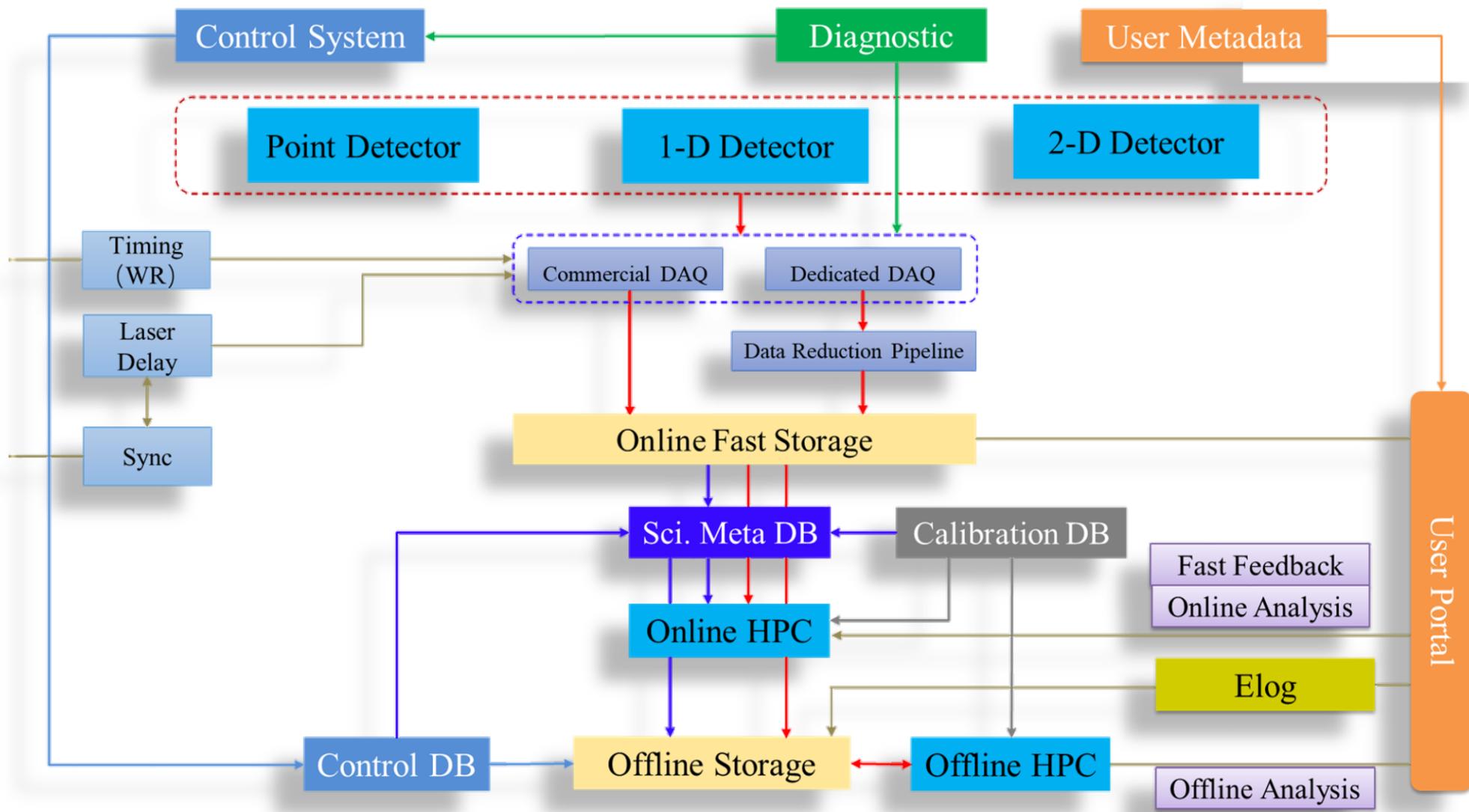
6个冷池共84个机柜

实验数据快速缓存、在线分析、快速反馈

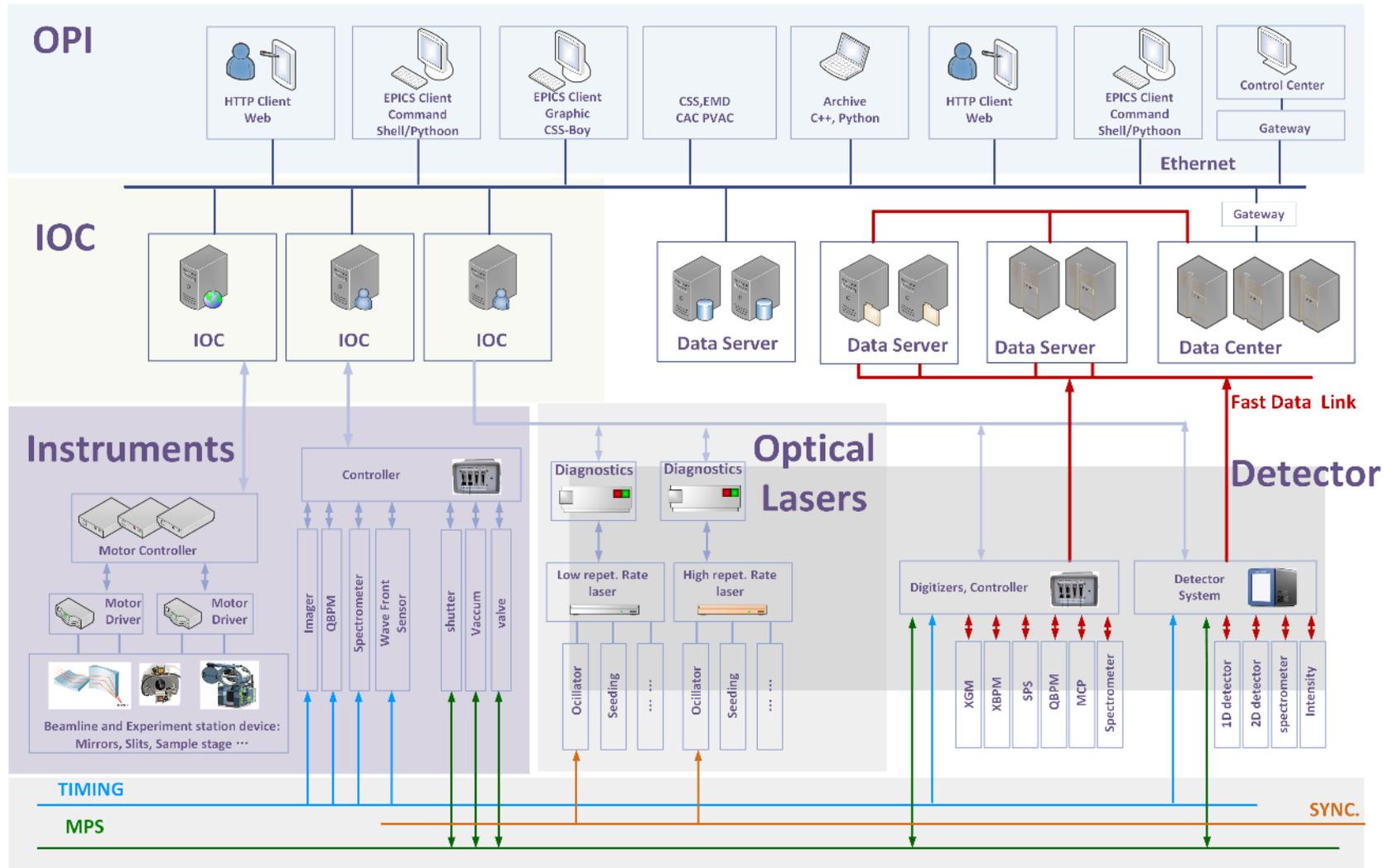
实验数据长期存储、离线分析、共享等

机房的面积、空调、电力均考虑了二期建设的需求!

Data Sources@ SHINE BE



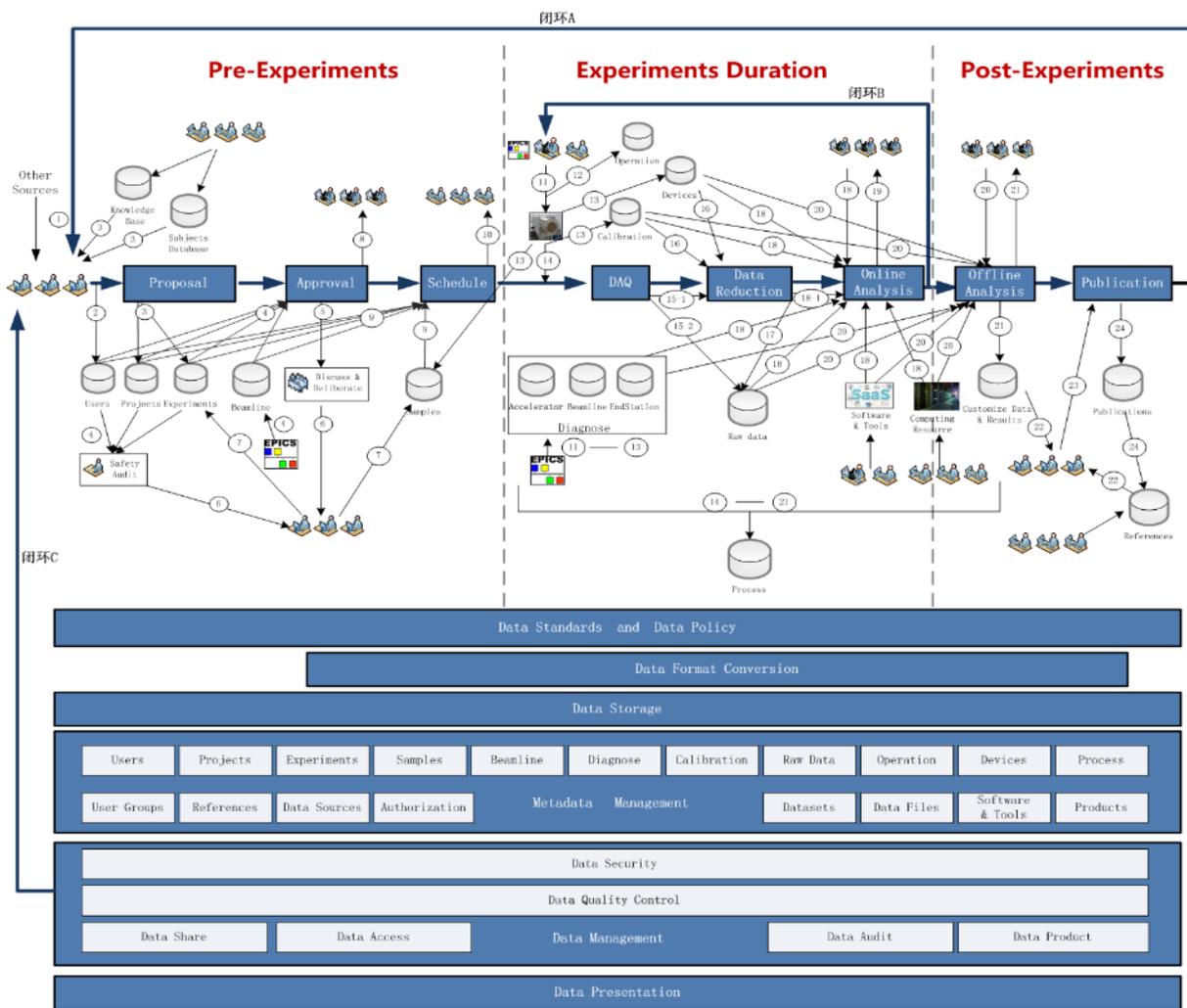
Control System@ SHINE BE



Life Cycle Data Management @ SHINE



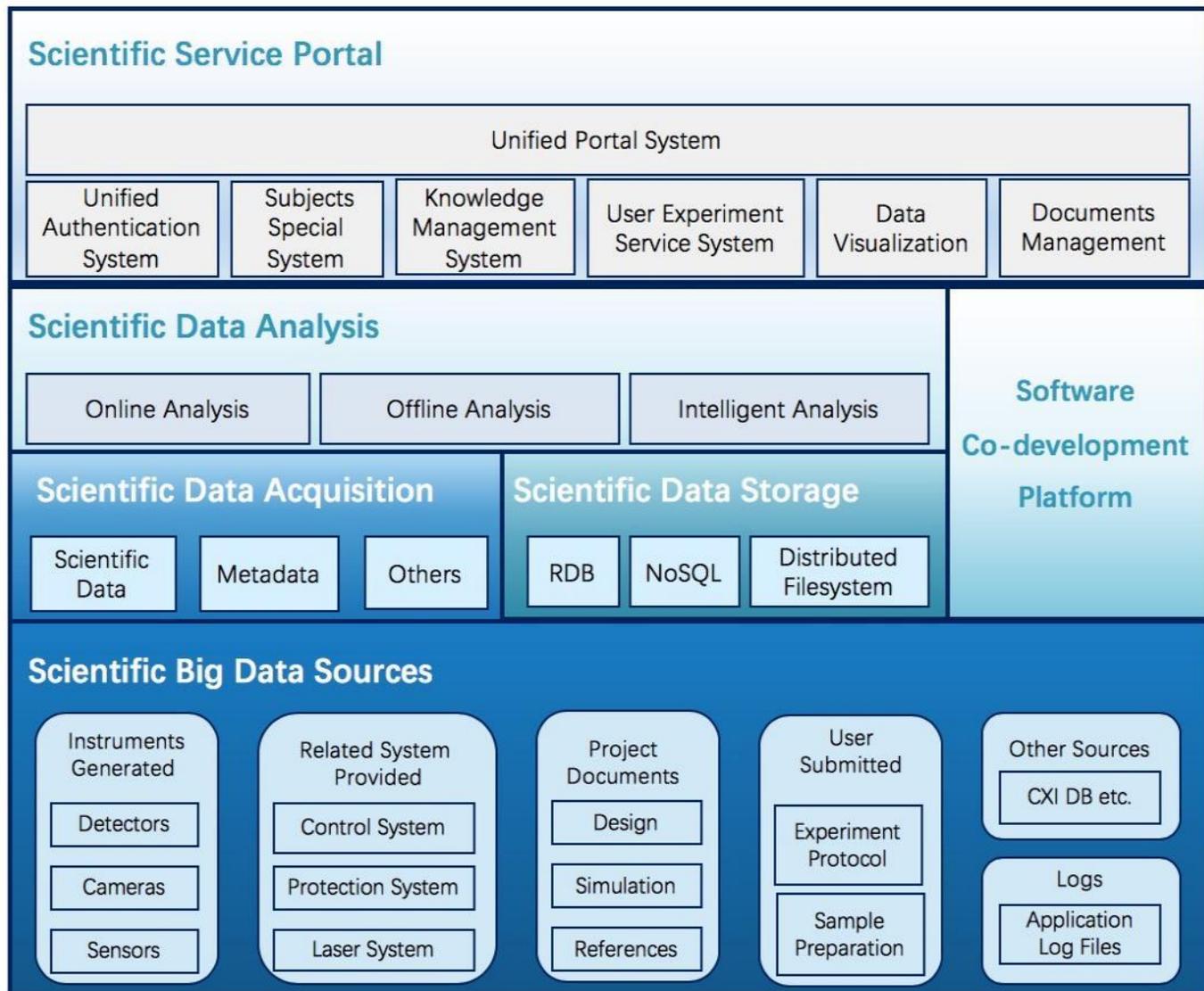
基于全生命周期的科研数据管理与分析系统



序号	组成	描述
1	Proposal	用户在统一认证系统注册认证后，登录门户网站，提交实验相关资料，包括项目相关信息和实验计划，并申请预约机时。
2	Approval	束线站科研技术人员对实验安全进行审计，并根据束线站运行情况和用户需求安排机时，复杂情况需多位科学家共同商议，对用户的申请进行讨论与评议。用户根据反馈结果准备样品，并在门户中提交样品的详细信息。同时安排实验操作者进行实验操作培训。
3	Schedule	束线站科研技术人员根据上一阶段的结果和用户提交的信息，确认用户机时，并安排相应的实验技术支持人员。
4	DAQ	实验开始，采集探测器输出数据。若实验不需要数据约简，则将实验原始数据保存下来，供后续在线分析。
5	Data Reduction	对于需要数据约简的实验，根据用户设置的约简比重进行数据约简，约简后的原始数据保存下来，供后续分析。
6	Online Analysis	用户根据保存下来的原始数据、装置诊断信息、探测器刻度数据以及数据集采集探测器的状态，利用软件平台和超算平台进行在线分析，得到快速反馈结果。
7	Offline Analysis	用户根据保存下来的原始数据、装置诊断信息、探测器刻度数据以及数据集采集探测器的状态，利用软件平台和超算平台进行离线分析，得到结果。
8	Publication	用户根据分析结果编写论文，投稿发表。
9	Data Standards and Data Policy	1-8 的过程中遵循相应的数据标准和政策。
10	Data Format Conversion	1-8 的过程中部分数据需要进行格式转换，保证存储下来的是平台定义的标准格式。
11	Data Storage	1-8 的过程中产生的所有数据都需要存储。
12	Metadata Management	1-8 的过程中的元数据管理。
13	Data Management	1-8 的过程中的科研数据管理，包括科研数据安全、科研数据质量管控、科研数据共享管理、科研数据访问控制、科研数据合规审计、科研成果产出管理。
14	Data Presentation	1-8 的过程中的数据展现。
15	闭环A	8 Publication 阶段产生的文献可以作为其他科研项目或实验的来源。
16	闭环B	根据 6 Online Analysis 在线分析结果，可能需要重复多次实验过程。
17	闭环C	根据 13 Data Management 的统计结果，会给用户或用户组织反馈。

- 实验用户或用户组织机构
- EPICS控制系统
- 待存储内容的逻辑表示
- 子模块
- 实验操作者
- 实验站探测器
- 数据生命周期阶段
- 子模块
- 实验支持技术人员或数据平台运维人员
- 软件工具服务平台
- 生命周期阶段流向
- 数据流向

SHINE Software Architecture



Data Acquisition & Analysis

Data Transport & Storage

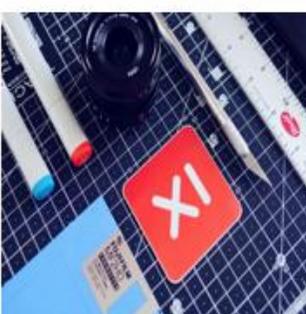
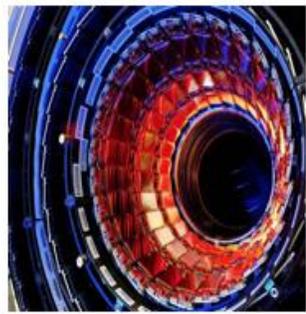
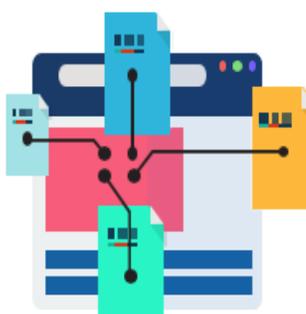
Data Service Portal

Software R&D Platform

- 围绕光子大科学装置的科学需求，开发研发、调试、运行阶段的数据软件
- 分层次、分阶段实现光子大科学装置大数据的融合交汇、科学管理与智能分析能力，并在整体平台进行集成贯通
- 可为高质量的光子大科学研究提供重要的数据支撑与服务

Software R&D Plan



统一标准管理		<ul style="list-style-type: none">● 解决非结构性科研大数据的标准化、统一化管理问题● 基于数据全生命周期管理，构建科研信息化服务平台
构建研发平台		<ul style="list-style-type: none">● 建立面向XFEL的数据分析软件研发平台● 提供高效便捷的科研专业性分析平台支撑能力● 实现科研分析过程的标准化和过程化管理
高通量数据引擎		<ul style="list-style-type: none">● 建立一套面向XFEL装置的通用式科研高通量数据流式处理模式● 形成原型，摸索解决高通量数据处理体系的瓶颈问题
汇交挖掘分析		<ul style="list-style-type: none">● 实现SHINE装置海量数据的汇交融合● 进行数据挖掘、智能分析、科学管理● 为高质量的科学研究提供不可或缺的基础支撑条件

- **研发**科研高通量数据流式处理系统，**比肩**国内外同类装置
- **实施**数据全生命周期管理，**突破**大科学装置数据管理规模和模式
- **保障**数据质量和安全，让每个决策都有数据支撑，实现数据**可回溯**、科研结果**可重现**
- **建设**先进的科研信息化环境和能力，成为未来科学的“**显微镜**”和“**望远镜**”

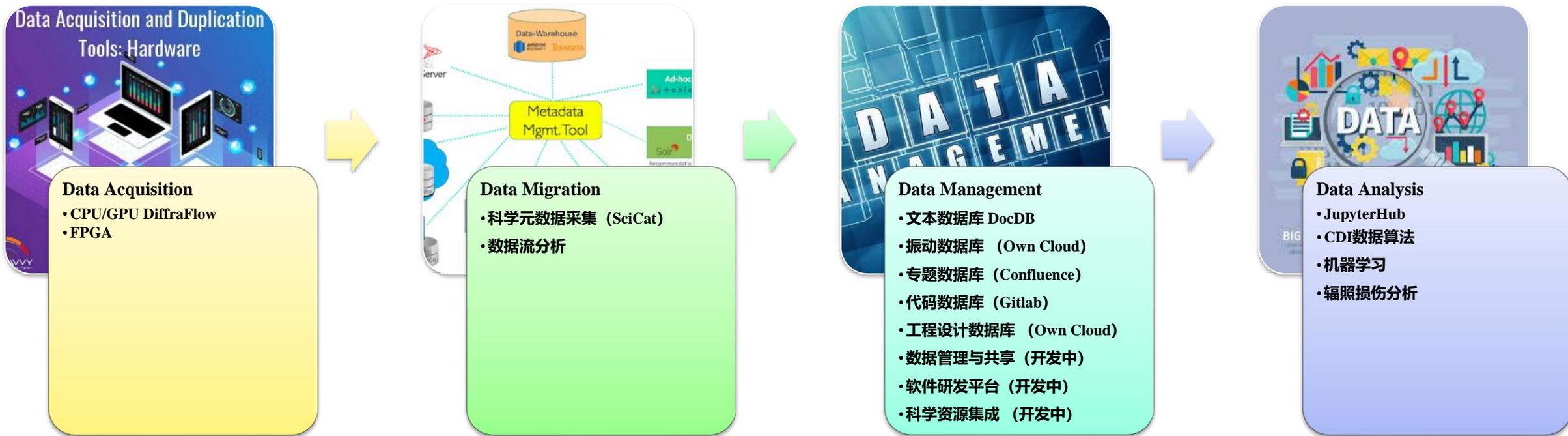


Outlines



- Overview of SHINE
- Data Requirement of SHINE BE
- Data System Design
- R&D Progress**

R&D of SHINE Data System



Data Center + HPC + Storage + Network

High Throughput Data Processing Prototype



借鉴CERN、LCLS-II国内外大科学装置数据平台设计，建设一套低成本、小规模、可扩展的原型设备。

- 验证、测试高通量数据的传输、存储的硬件和软件系统
- 为软件系统研制提供基础存储和计算分析环境

测试项目	指标	实测
存储容量	不小于300TB	864TB
顺序写入速度	5~10GB/s	15.23GB/s
CPU节点浮点计算能力	1TFLOPS	6.77TFlops
GPU节点单精度浮点计算能力	5TFLOPS	43.57TFlops



SHIN

高通量数据处理样机

```
T/W      N  NB  P  Q      Time      GFlops
-----
WC00C2R2 300591 384 2 2      2674.34    6.77053e+03
#PL_pdgsv() start time Fri Dec 20 10:02:48 2019
#PL_pdgsv() end time   Fri Dec 20 10:47:22 2019      CPU:6.77TFlops
||Ax-b||_oo/(eps*(||A||_oo*||x||_oo+||b||_oo)*N)= 0.0033139 ..... PASSED
Finished 1 tests with the following results:
1 tests completed and passed residual checks.
0 tests completed and failed residual checks.
0 tests skipped because of illegal input values.
End of Tests.

2019-12-20 09:56:29.583
T/W      N  NB  P  Q      Time      GFlops
-----
#R02L2L2 175851 512 4 2      4.357e+04    GPU:43.57TFlops
||Ax-b||_oo/(eps*(||A||_oo*||x||_oo+||b||_oo)*N)= 0.0000227 ..... PASSED
Per-Process Host Memory Estimate: 32.28 GB (MAX) 31.47 GB (MIN)
```

CPU/GPU浮点测试



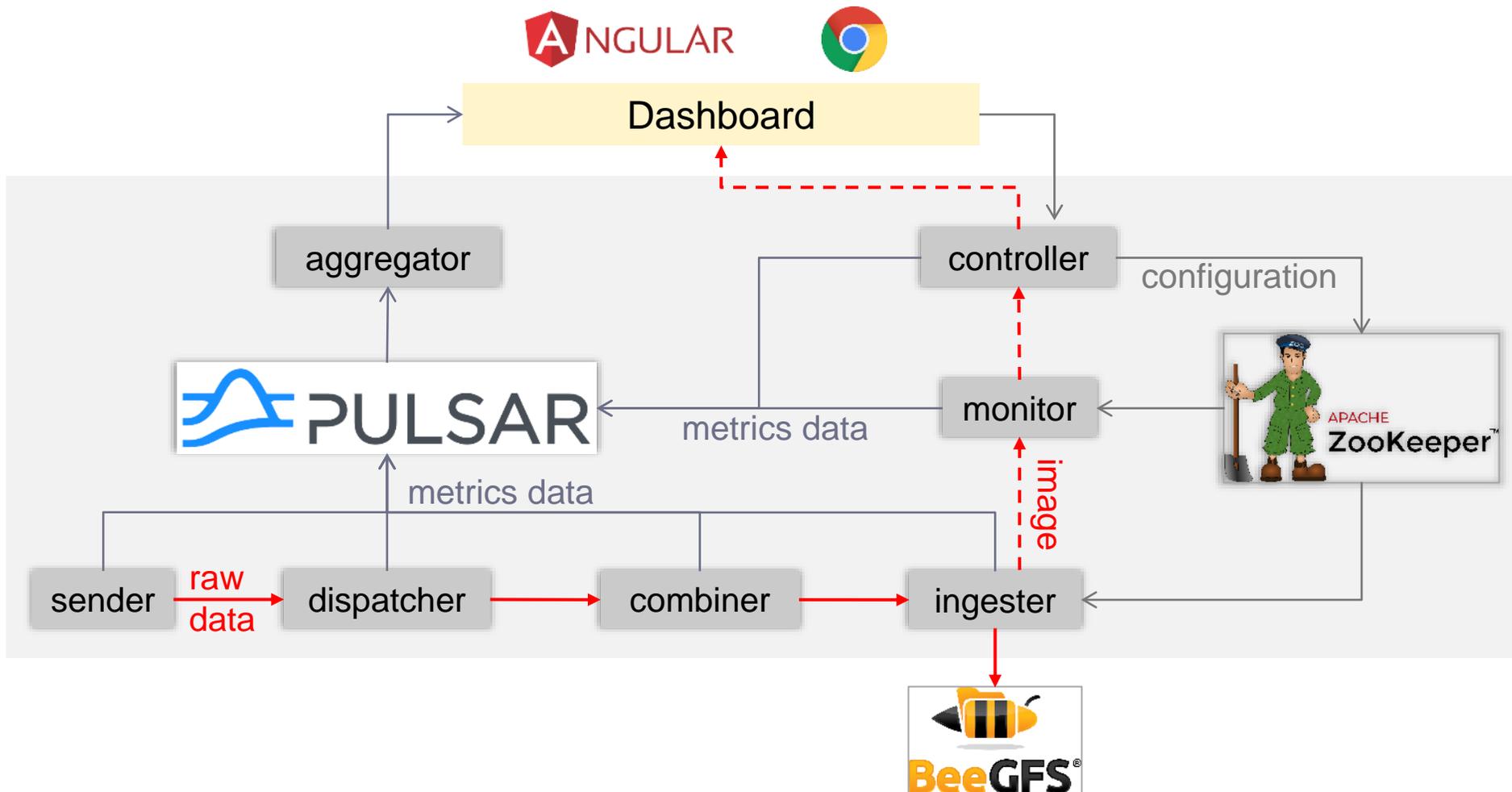
写入测试

High-Throughput Streaming Data Software

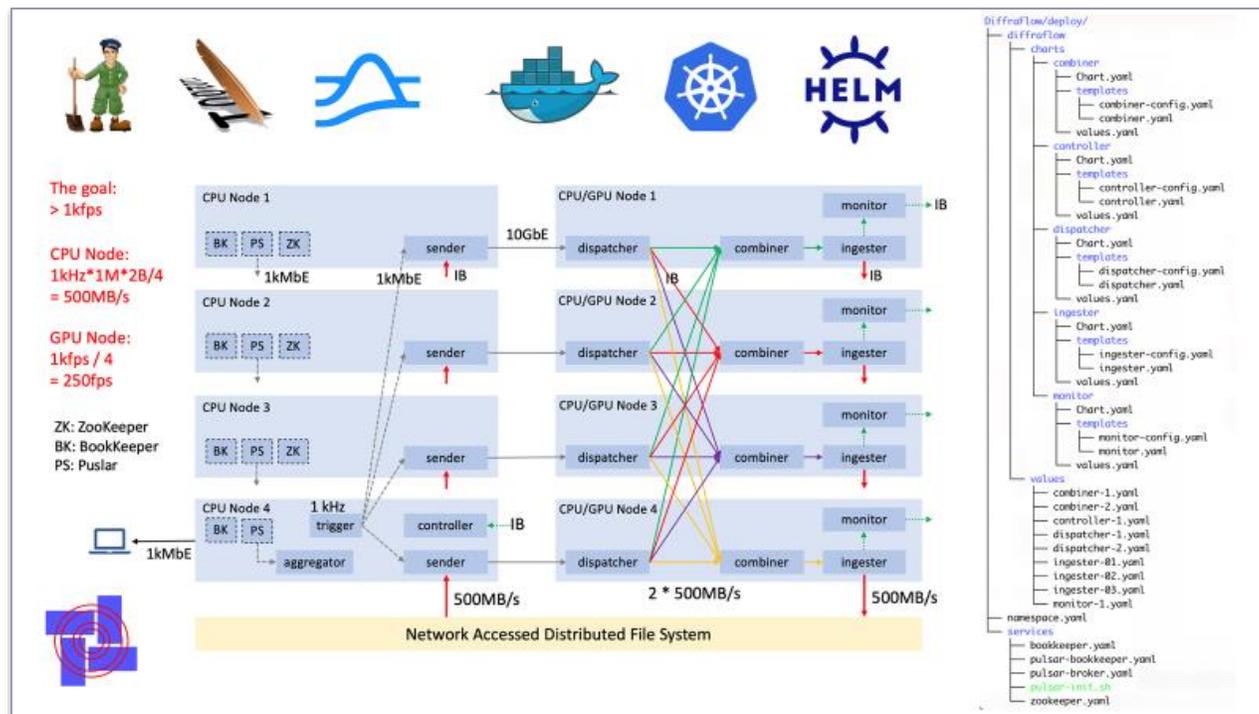
面阵探测器的通用高速数据处理软件：DiffraFlow

 DiffraFlow v0.1

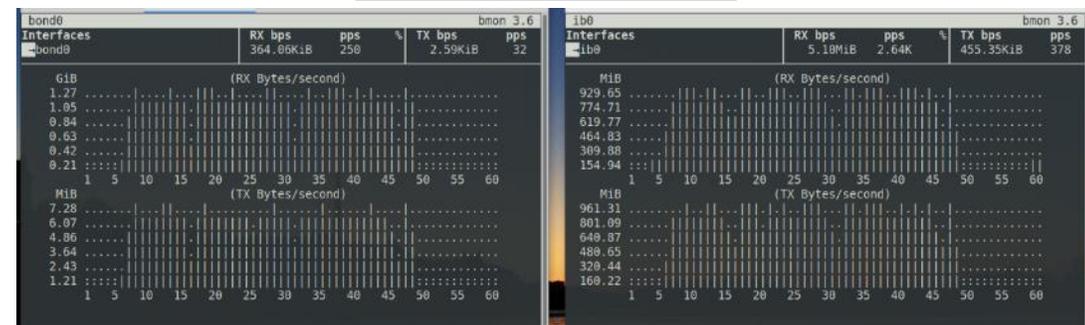
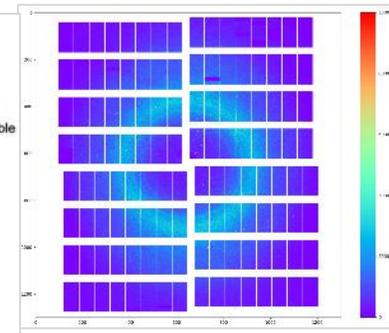
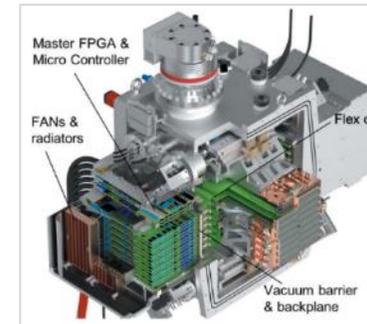
All components are fully tested on Data Processing Prototype,



High-Throughput Streaming Data Software (DiffraFlow)



ODS ID 83	
Deployment Summary	
Deployment:	diffraflow
Contact:	info@odis.ch
Deployment date:	2019-08-13
Last modified:	2019-08-13
URL:	https://odis.ch/odis83
Related entries:	ODS ID 80
Publication Details	
Title:	Neutron's wide community
Authors:	Max D. Pflaum et al.
Journal:	Nature Communications
Year:	2019
DOI:	10.1038/s41467-019-14047-7
Experimental Conditions	
Method:	Serial Femtosecond Crystallography
Sample:	β-lactamase
Wavelength:	1.33 Å (9.69 eV)
Lightsource:	European XFEL
Beamline:	SFX-SLS
Data Files	
Raw Data:	14 KB
Generated Data:	14 KB



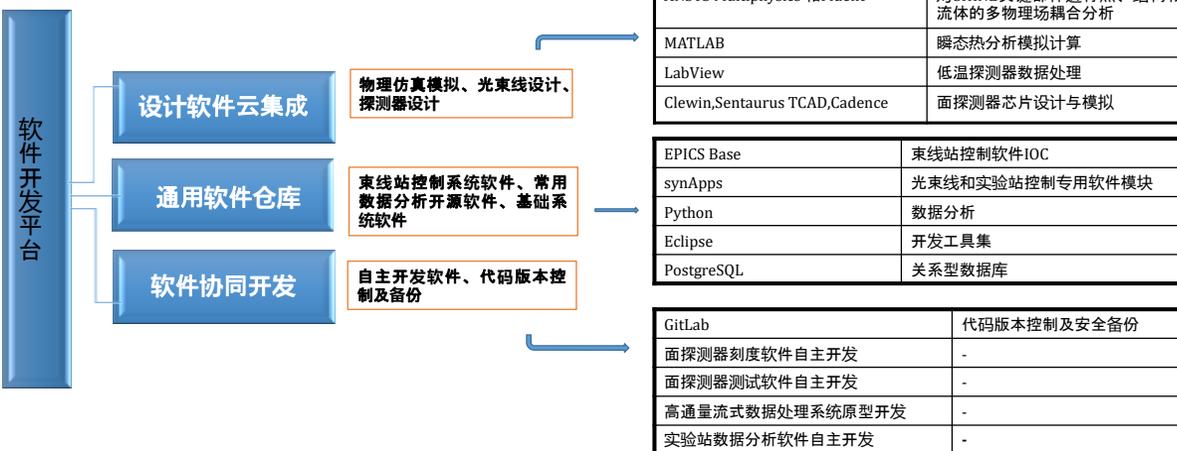
数据接收速率 (万兆网)

事例重建中数据交换 (IB网)

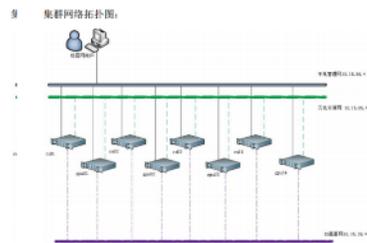
- ✓ 已经完成原型系统软件方案的内部评审会
- ✓ 已经完成核心代码开发、样机软件环境部署
- ✓ 完成样机上的发送触发测试和数据传输测试, 百万像素图像数据接收与事例重建速率达到2kfps

Software R&D Platform

光子科学数据分析软件研发平台上线服务：
整合了硬件资源、软件资源，形成了工程设计数据平台，提出了管理规则、软件测试方案



软件资源

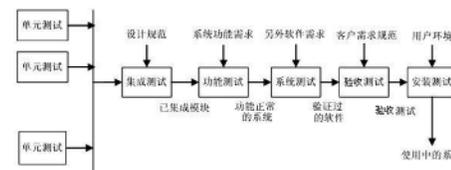
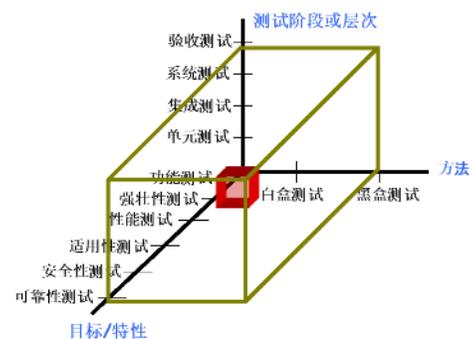


1.3 集群软件资源配置情况

集群	主机名	CPU 配置	内存配置	GPU 配置
CPU 节点	cu01	Intel xeon 6132 * 2 进程: 28	192G	未提供
	cu02	Intel xeon 6132 * 2 进程: 28	192G	未提供
	cu03	Intel xeon 6132 * 2 进程: 28	192G	未提供
	cu04	Intel xeon 6132 * 2 进程: 28	192G	未提供
GPU 节点	gpu01	Intel xeon 6132 * 2 进程: 28	192G	NV_3XGH_TESLA-V100*2
	gpu02	Intel xeon 6132 * 2 进程: 28	192G	NV_3XGH_TESLA-V100*2
	gpu03	Intel xeon 6132 * 2 进程: 28	192G	NV_3XGH_TESLA-V100*2
	gpu04	Intel xeon 6132 * 2 进程: 28	192G	NV_3XGH_TESLA-V100*2

1.4 集群软件资源配置情况

类型	软件	版本	备注
操作系统	CentOS	7.6	
	gcc	intel2018	/opt/intel
编译器	icc	intel2018	/opt/intel
	gpc		
并行环境	intel-mpi	intel2018	/opt/intel
	mpich	3.2.1	/opt
	mvapich2	2.3	/opt
	openmpi	3.1.0	/opt
CUDA 环境	cuda	10.0	/usr/local
应用软件	Matlab	2019a	/data-ibapps/Matlab-2019a



涉及 涉及到的所有计算服务器网络配置信息如下表所示:

集群	主机名	万兆网络 掩码: 24 位	IB 网络 掩码: 24 位	千兆管理网 网掩码: 24 位
CPU 节点	cu01	10.15.85.28	10.15.39.20	10.15.86.19
	cu02	10.15.85.29	10.15.39.21	10.15.86.20
	cu03	10.15.85.30	10.15.39.22	10.15.86.21
	cu04	10.15.85.31	10.15.39.23	10.15.86.22
GPU 节点	gpu01	10.15.85.32	10.15.39.24	10.15.86.23
	gpu02	10.15.85.33	10.15.39.25	10.15.86.24
	gpu03	10.15.85.34	10.15.39.26	10.15.86.25
	gpu04	10.15.85.35	10.15.39.27	10.15.86.26

1.1 集群网络环境

1.2 集群共享目录

1. /opt 集群共享目录, 用于安装公共应用软件
2. /data-ib 集群共享目录, 用于存放用户家目录数据



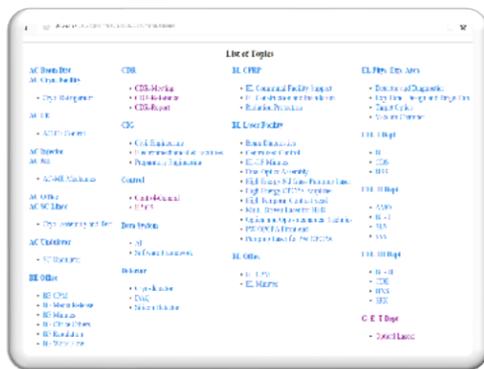
硬件资源

软件测试工程管理方案

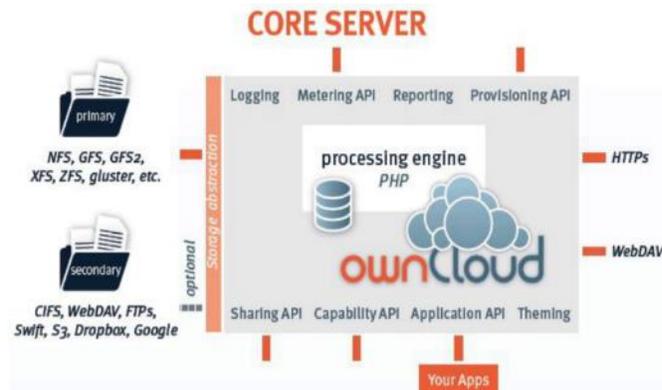
Databases of SHINE BE



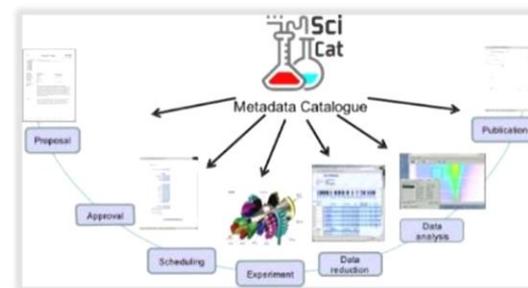
建设包含多样化信息来源的光子科学数据库群



SHINE DocDB



BE Design DB



SciCat



E-learning DB

- ❑ 已完成文档数据库SHINE DocDB的部署、年度维护测试
- ❑ 已搭建束线站工程设计数据库
- ❑ 已经部署了振动数据专题数据库
- ❑ 正在进行SciCat的部署与测试
- ❑ 已建立以CDI为主题的E-Learning数据库 (Confluence)

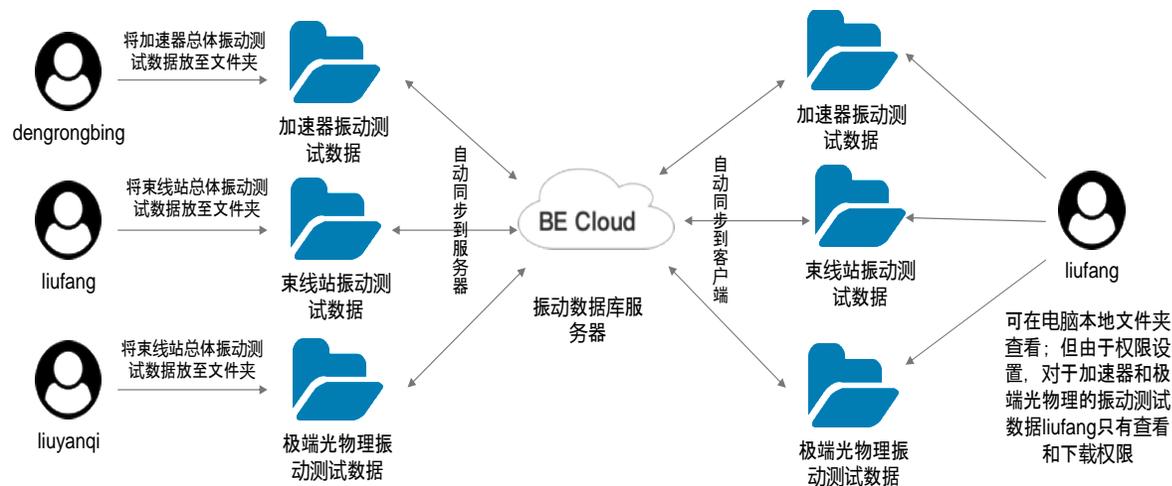
截止2020.3.19, SHINE DocDB使用情况

项目	数量
文件夹	620
文档	1013
作者	35
主题	83
会议	100
服务器空间占用	3.8G

Databases R&D

振动数据库试运行

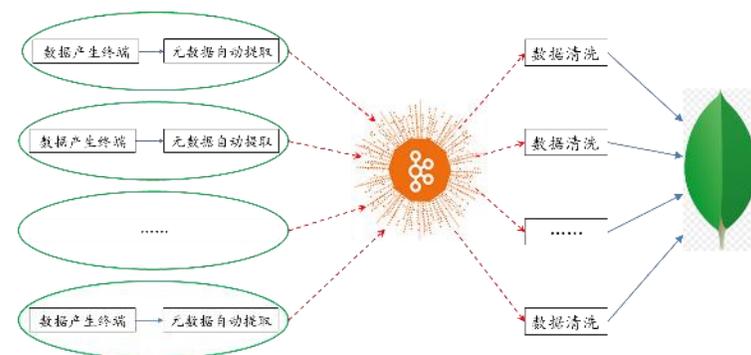
- 7月22日起进入**试运行**阶段
- 目前共计15个用户，其中振动相关用户4人
- 当前数据存储总量17GB，文件共2028个



SHINE振动数据库

科学数据库元数据提取测试

- 基于SciCat开发的科学元数据库
- 已完成科学元数据库的MongoDB写入、Kafka调试
- 基于**低温探测器实验数据**进行测试



低温TES-X 探测器元数据提取测试

Design Platform of SHINE BE

完成了工程设计数据平台的方案设计、资源整合、管理规则、上线服务

东线站工程设计数据平台用户手册

本手册使用对象为本项目高性能集群的普通用户，本手册旨在将集群的整体情况，包括软硬件资源、网络环境、资源管理等集群现状进行描述。

1. 集群配置

集群项目包含：

- (1) 管理：gpu01（登录、管理、计算共用）
- (2) CPU计算节点（4个）：cu01、cu02、cu03、cu04
- (3) GPU计算节点（4个）：gpu01、gpu02、gpu03、gpu04
- (4) 1套存储（10台AS13000）：inspur01-10
- (5) Slurm作业系统调度管理节点：gpu01

管理网络以太网环境：

计算网络采用 100Gb EDR 高速 InfiniBand 网络环境：

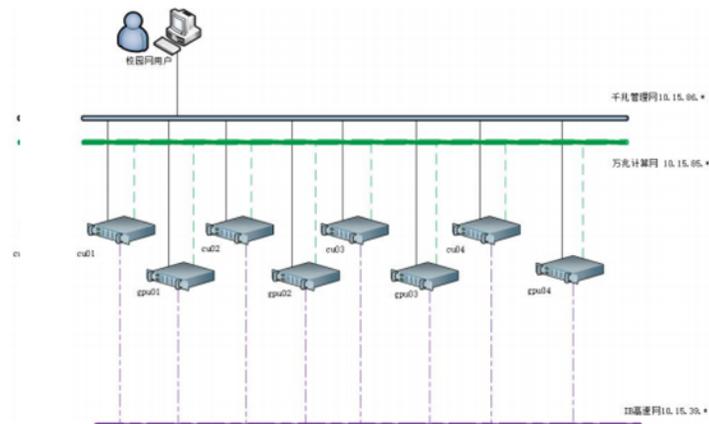
每个 GPU 节点包含 2 个 tesla v100 图形计算卡。

IB 网段命名规则是以太网段名前加 ib 前缀，如 ibcu01, ibcu02, ibcu03, ..., ibgpu04 等。

万兆网段命名规则是以太网段名后加-10g 后缀，如 cu01-10g, cu02-10g, ..., gpu04-10g 等。

1.1 集群网络环境

集群网络拓扑图：



涉及 涉及到的所有计算服务器网络配置信息如下表所示：

集群	主机名	万兆网络 掩码：24 位	IB 网络 掩码：24 位	千兆管理网 网关： 10.15.86.1 掩码：24
CPU 节点	cu01	10.15.85.28	10.15.39.20	10.15.86.19
	cu02	10.15.85.29	10.15.39.21	10.15.86.20
	cu03	10.15.85.30	10.15.39.22	10.15.86.21
	cu04	10.15.85.31	10.15.39.23	10.15.86.22
GPU 节点	gpu01	10.15.85.32	10.15.39.24	10.15.86.23
	gpu02	10.15.85.33	10.15.39.25	10.15.86.24
	gpu03	10.15.85.34	10.15.39.26	10.15.86.25
	gpu04	10.15.85.35	10.15.39.27	10.15.86.26

1.2 集群共享目录

1. /c
2. /data-ib

集群共享目录，用于安装公共应用软件

集群共享目录，用于存放用户家目录数据



1.3 集群硬件资源配置情况

集群	主机名	CPU 配置	内存配置	GPU 配置
CPU 节点	cu01	Intel xeon 6132 * 2 进程：28	192G	未提供
	cu02	Intel xeon 6132 * 2 进程：28	192G	未提供
	cu03	Intel xeon 6132 * 2 进程：28	192G	未提供
	cu04	Intel xeon 6132 * 2 进程：28	192G	未提供
GPU 节点	gpu01	Intel xeon 6132 * 2 进程：28	192G	NV_32GB_TESLA-V100*2
	gpu02	Intel xeon 6132 * 2 进程：28	192G	NV_32GB_TESLA-V100*2
	gpu03	Intel xeon 6132 * 2 进程：28	192G	NV_32GB_TESLA-V100*2
	gpu04	Intel xeon 6132 * 2 进程：28	192G	NV_32GB_TESLA-V100*2

1.4 集群软件资源配置情况

类型	软件	版本	备注
操作系统	CentOS	7.6	
编译器	icc	intel2018	/opt/intel
	gcc		
并行环境	intel-mpi	intel2018	/opt/intel
	mpich	3.2.1	/opt
	mvapich2	2.3	/opt
	openmpi	3.1.0	/opt
CUDA 环境	cuda	10.0	/usr/local
应用软件	Matlab	2019a	/data-ib/apps/Matlab-2019a

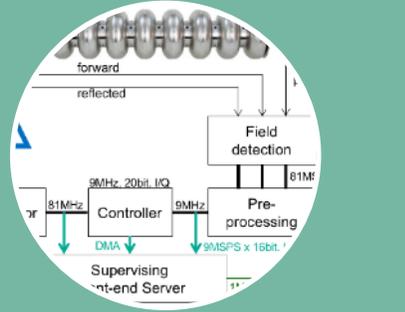
SBP集群：用户服务为主



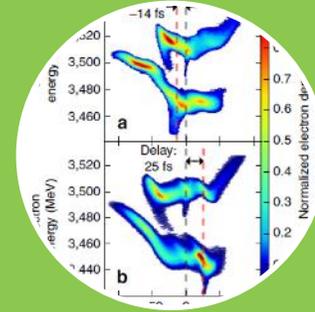
Machine Learning in SHINE



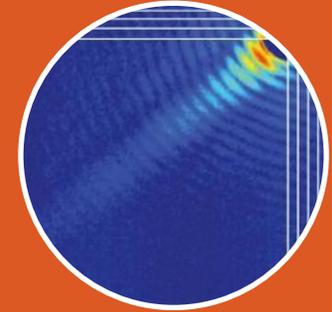
Automatic
Optimization



Anomaly
Detection



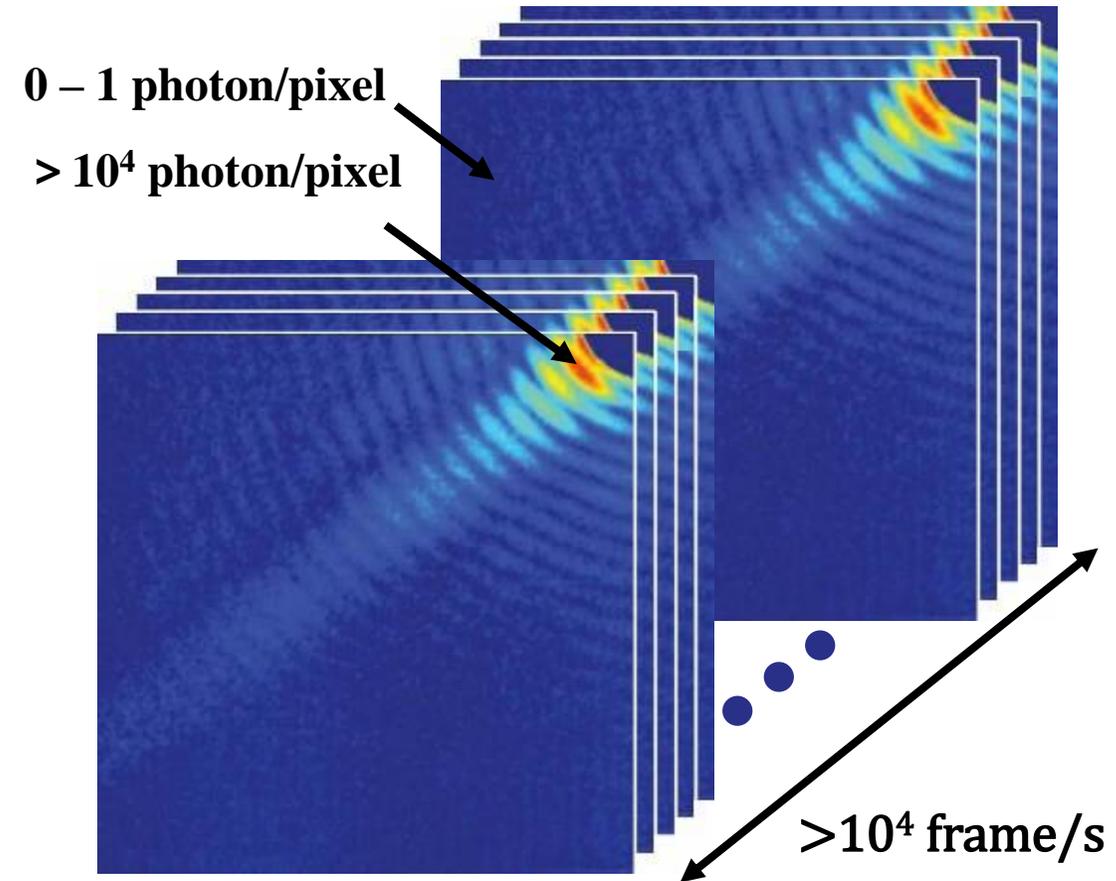
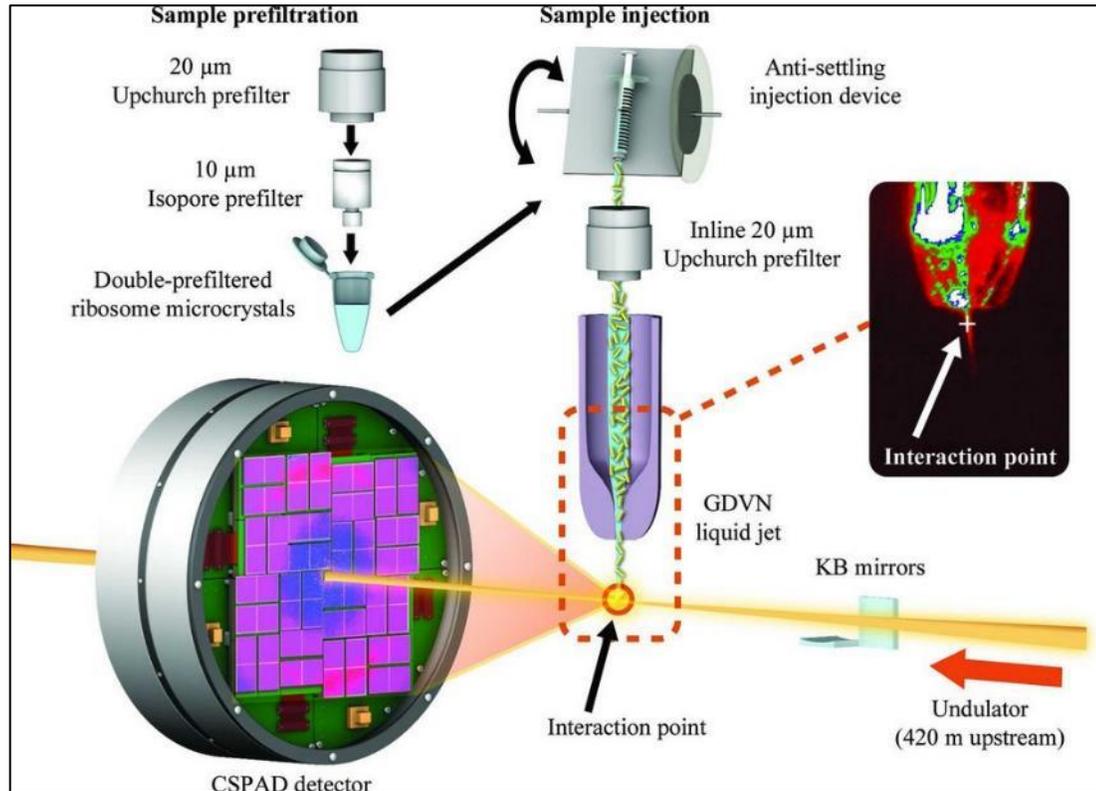
Diagnostic
Prediction



Data
Reduction



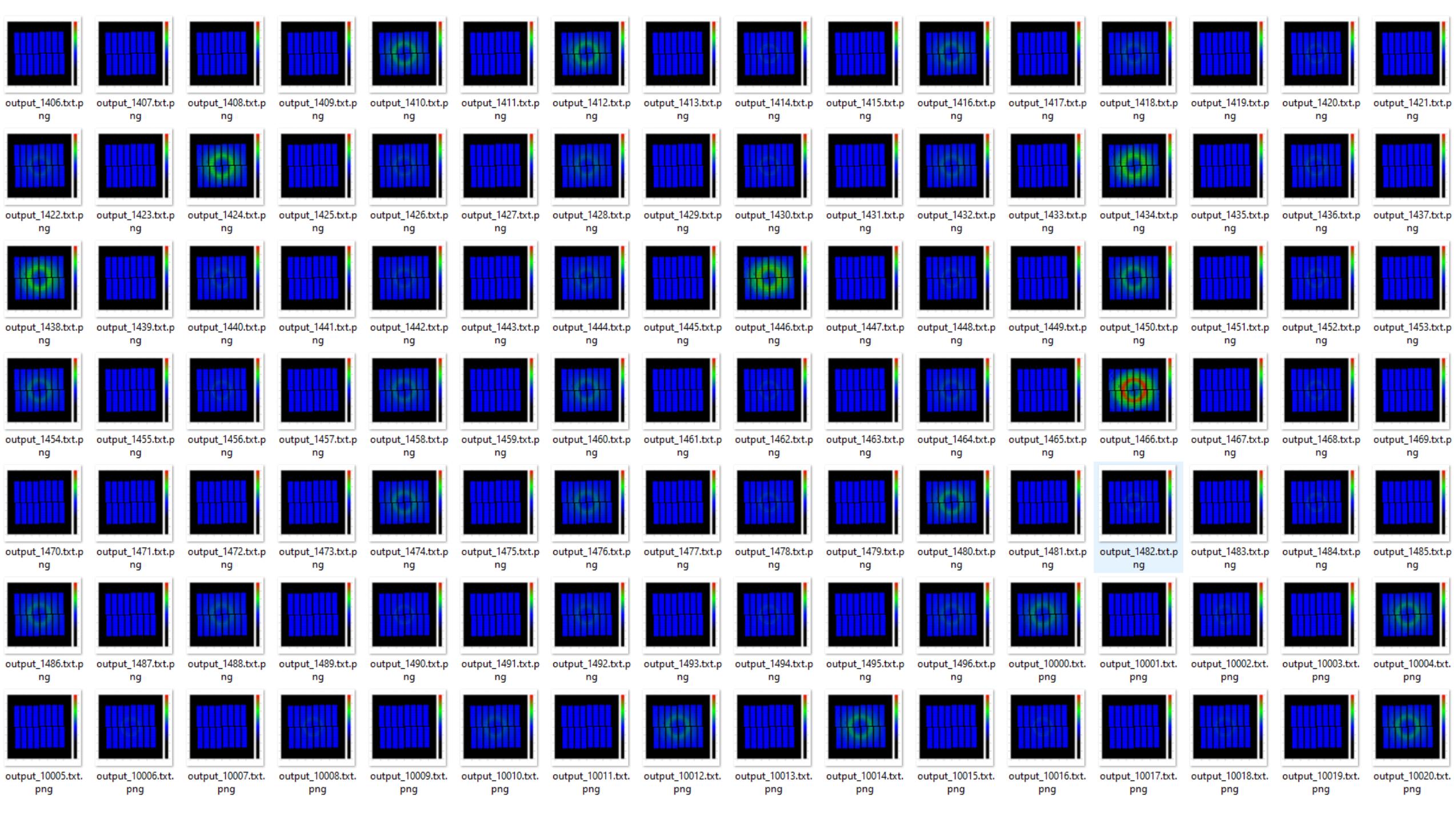
On-The-fly Data Reduction in XFEL



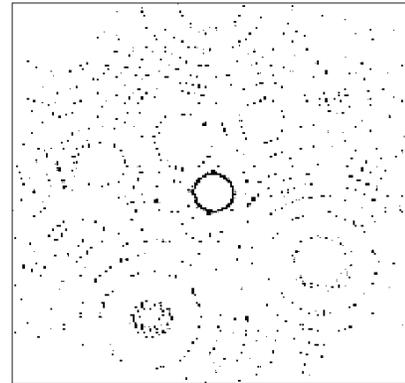
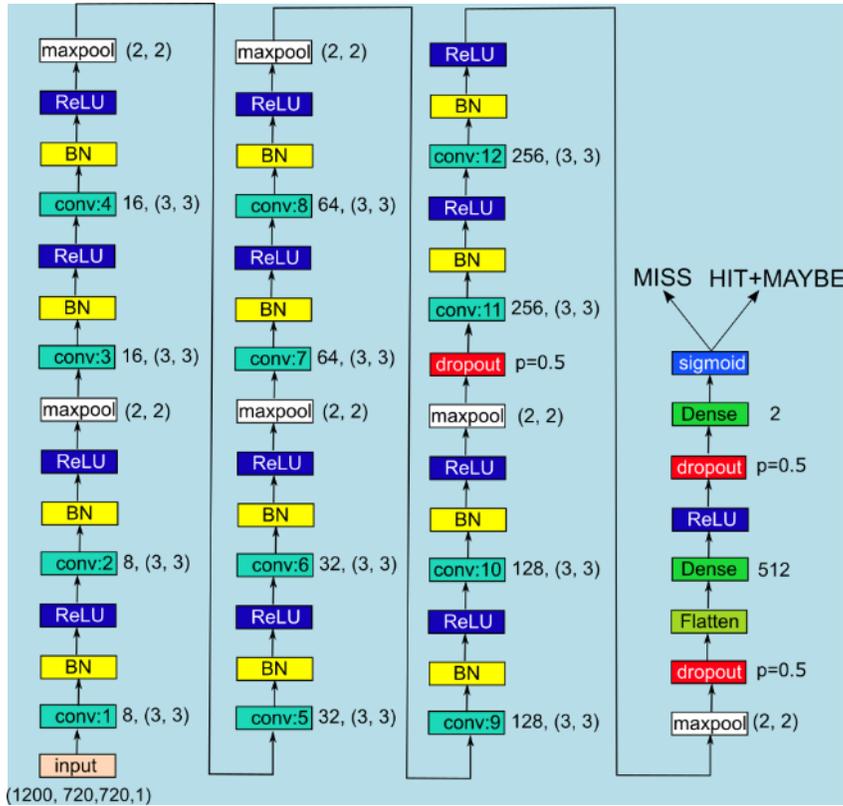
CPU/GPU based system not sufficient for data reduction

key: online reduction based on AI

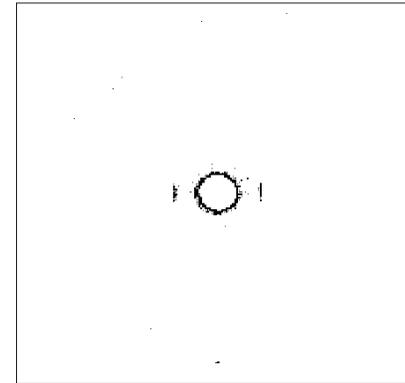
	2025	2030
Data Rate	100 GB/s	1000 GB/s
Raw Data (year)	30 PB	1 EB
Reduced Data (year)	3 PB	100 PB



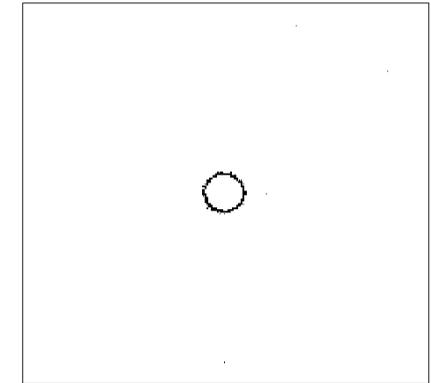
Data Selection based on ML



label = HIT



label = MAYBE



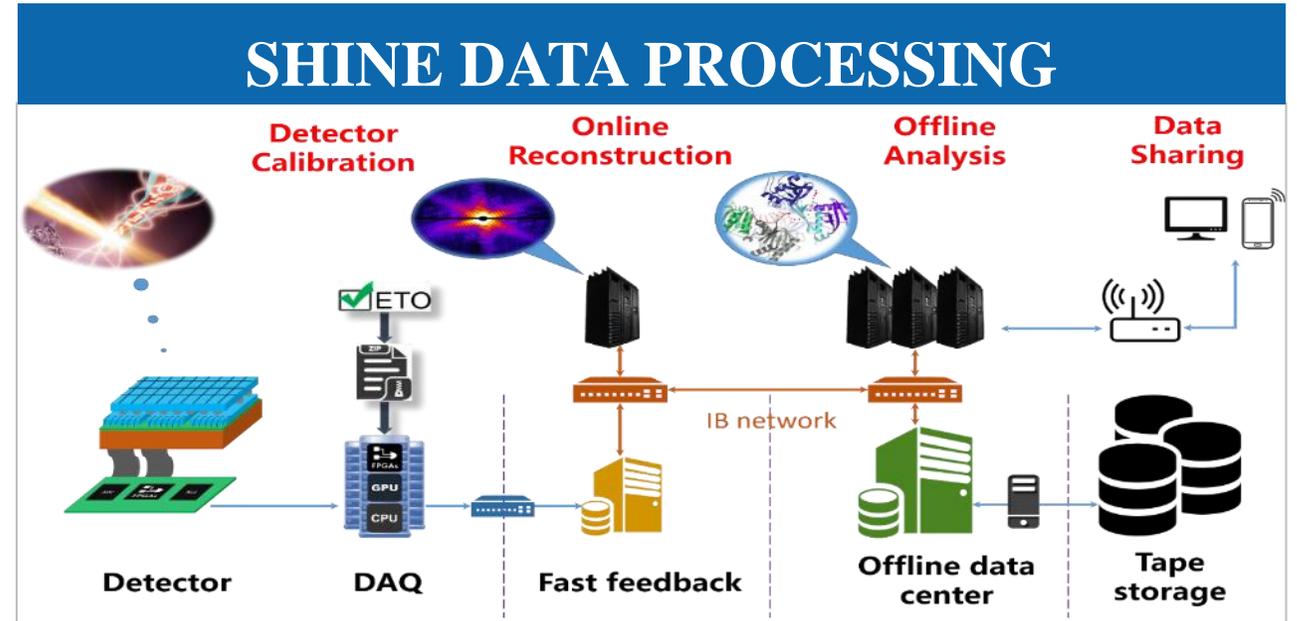
label = MISS

dataset	HIT + MAYBE 的权重		prediction		
			MISS	HIT + MAYBE	
LN83	3.5	actual	MISS	99.10%	0.90%
			HIT + MAYBE	1.54%	98.46%
LN84	1.1	actual	MISS	92.75%	7.25%
			HIT + MAYBE	1.04%	98.96%
LO19	1.9	actual	MISS	97.48%	2.52%
			HIT + MAYBE	4.32%	95.68%

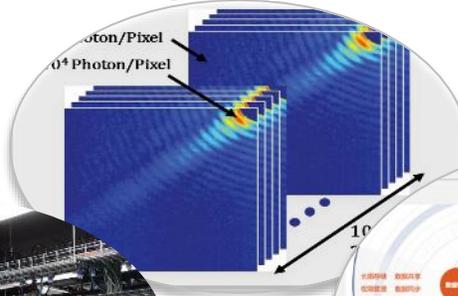
SHINE Big Data Strategy

□ Deliver Software and Tools to Enable Data-Driven XFEL Science

- Low-latency network & fast storage
- Leverage existing big data and data processing technologies
- AI Application
- Collaborations with top institutions, mega scientific facilities, and companies



High-Speed DAQ



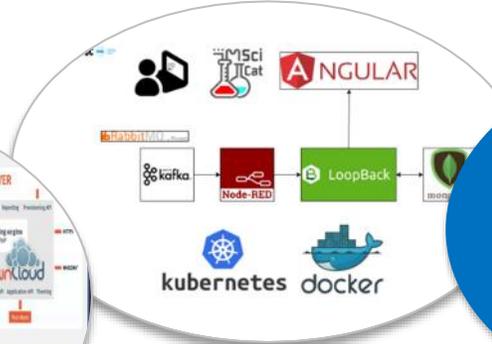
Data Center



Database



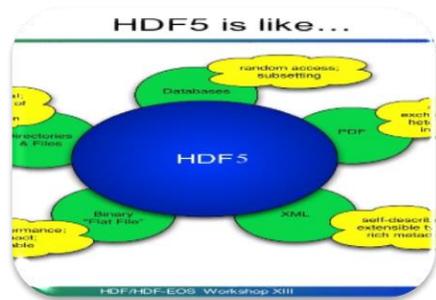
Meta Data



Big Data Management
Kafka
MongoDB

Cloud Computing
Machine Learning

Technologies for Big Data @ Advanced Light Sources



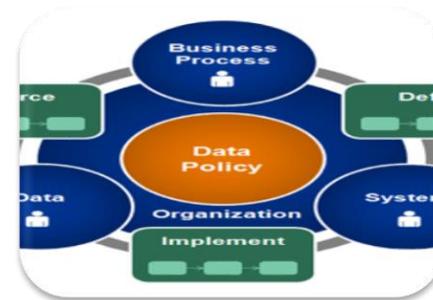
Data Format



Database



Meta Data



Data Policy



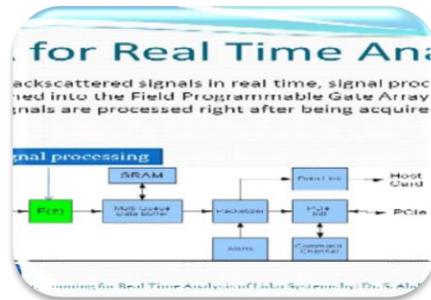
Software Framework



High-Speed Network



Data Reduction



Real-time Analysis



Virtual Reality



AI

Collaboration With Supercomputing Centers and Light Sources

National Super Computing Center



Data Achieving



Super Computing

Other Light Source@China



Data Sharing

Photon Science Data Center@ ZhangJiang Lab



Data Sharing



SLAC NATIONAL ACCELERATOR LABORATORY



Summary



- ❑ SHINE will enable the probing of structural and functional properties of materials, including the physical and chemical behaviors in condensed matters and biomaterials.
- ❑ SHINE offers scientist a rare opportunity to probe, potentially control, the transfer of charge and energy at atomic scale with fs resolution.
- ❑ SHINE poses grand challenges big data that will be handled by state-of-art data science and dedicated data center.

SHINE General Experiment Technology Dept. (GET)



Head of GET (General Experiment Technology, 实验辅助分总体) Department: Ping HUIAI (怀平)

Detector System

Data Acquisition &
Analysis System

Data Management
System

Laser System





Thank you for your attention!

