

基于 SciCat 的合肥光源实验数据管理系统开发

林 广 张大地 张 震 贺 博 孙晓康 刘功发

(中国科学技术大学 国家同步辐射实验室 合肥 230029)

摘要 合肥光源(Hefei Light Source, HLS-II)是以真空紫外和软 X 射线为主的专用同步辐射光源,用户的实验效率与数据处理效率是其运行的核心指标,因此设计并开发了基于元数据目录框架 SciCat 的合肥光源实验数据管理系统,以有效提高用户的实验效率和数据处理效率。通过设计开发数据采集服务、元数据系统、文件存储系统与数据分析平台,以及接入现有的用户服务系统,实现涵盖用户认证、课题申请、专家审批、机时分配、数据采集、数据存储、数据分析、数据发布等实验及数据全生命周期的统一管理。该系统已在合肥光源的软 X 射线成像实验站进行了部署和功能测试,达到了设计要求。

关键词 合肥光源, SciCat, 实验数据, 全生命周期管理, 元数据目录

中图分类号 TL99

DOI: 10.11889/j.0253-3219.2022.hjs.45.020101

Development of experimental data management system for HLS-II based on SciCat

LIN Guang ZHANG Dadi ZHANG Zhen HE Bo SUN Xiaokang LIU Gongfa

(National Synchrotron Radiation Laboratory, University of Science and Technology of China, Hefei 230029, China)

Abstract [Background] Hefei Light Source (HLS-II) is a dedicated synchrotron radiation source with featured the characteristics spectrum in vacuum ultraviolet and soft X-rays regions. The user's experiment efficiency and data processing efficiency are the principal indicators of HLS-II operation. [Purpose] This study aims to develop an experimental data management system based on SciCat to improve the user's experimental and data processing efficiency. [Methods] As the core component of HLS-II experimental data management system, the metadata system was designed and developed within the framework of SciCat, an open source metadata catalog framework used to realize the unified management of the whole life cycle of experimental data of large scientific devices. Data acquisition service, metadata system, file storage system and data analysis platform were integrated into SciCat, and the hierarchical data format 5 (HDF5) was adopted as the standard experimental data format. Interface between existing user service systems and SciCat was implemented to achieve the management of the whole experiment and experimental data lifecycle, including user authentication, proposal application, expert approval, machine time schedule, data acquisition, data storage, data analysis and data publication. [Results & Conclusions] The system has been deployed and functionally tested at the soft X-ray microscopy experimental station of HLS-II, satisfying all design requirements.

Key words HLS-II, SciCat, Experimental data, Whole lifecycle management, Metadata catalogue

中央高校基本科研业务费专项资金(No.WK2310000102)、中国科学院合肥大科学中心重点研发项目(No.2019HSC-KPRD003)资助

第一作者: 林广, 男, 1996年出生, 2019年毕业于华南师范大学, 现为硕士研究生, 研究领域为加速器控制技术

通信作者: 孙晓康, E-mail: sunxk@ustc.edu.cn

收稿日期: 2021-09-10, 修回日期: 2021-11-29

Supported by the Fundamental Research Funds for the Central Universities (No.WK2310000102), Key R&D Project of Hefei Science Center, Chinese Academy of Sciences (No.2019HSC-KPRD003)

First author: LIN Guang, male, born in 1996, graduated from South China Normal University in 2019, master student, focusing on accelerator control technology

Corresponding author: SUN Xiaokang, E-mail: sunxk@ustc.edu.cn

Received date: 2021-09-10, revised date: 2021-11-29

大科学装置为多学科前沿课题研究提供了良好的基础研究平台,对于解决国家需求中的战略性、基础性和前瞻性科技问题具有重要意义^[1-3]。大科学装置的实验研究过程比较复杂:用户进行科学实验前,首先需要在用户服务系统提交实验课题申请书;待实验课题申请书获批后,再从实验站获得机时分配后才能开展实验;实验期间及完成后,一般通过人工方式从实验站拷贝数据,搭建环境分析数据,得出实验结果。这种由人工管理实验数据的方式导致实验与数据处理效率较低。通过建设大科学装置实验数据全生命周期统一管理平台,以有效提高用户的实验效率和数据处理效率,进而提高大科学装置的科学产出,是国内外大科学装置信息化建设的一个重要趋势^[4-5]。实验数据统一管理平台的建设方案可分为两类:通过相对独立开发的方式实现和基于合作开发的元数据目录框架实现。采用前一种方案的有德国 ANKA 光源^[6]、美国 NSLS-II 光源^[7]以及日本 J-PARC 中子源^[8]。后者例如英国 DIAMOND 光源^[9]、ISIS 中子源^[10]以及中国散裂中子源(China Spallation Neutron Source, CSNS)^[11-12]以元数据目录框架 ICAT^[13]为核心建成了实验数据统一管理平台;瑞典 MAX IV 光源^[14-15]、上海同步辐射光源(Shanghai Synchrotron Radiation Facility, SSRF)^[16]、高能同步辐射光源(High Energy Photon Source, HEPS)^[17]与硬 X 射线自由电子激光装置(Shanghai High repetition rate XFEL and Extreme light facility, SHINE)^[18]目前也正在规划设计基于元数据目录框架 SciCat^[19]的实验数据综合管理服务平台。

合肥光源(Hefei Light Source II, HLS-II)是以真空紫外和软 X 射线为主的专用同步辐射光源,目前拥有 11 条光束线及实验站^[20]。作为面向国内外用户的大科学装置,合肥光源现有的用户服务系统仅针对实验课题审批、机时分配等,而在海量实验数据的管理上,存在着数据资源分散、缺乏数据分析环境、无方便有效的交流共享途径等问题,难以满足国内外科研人员多样化访问实验数据的需要。针对以上问题和需求,我们基于开源元数据目录框架 SciCat 开发了合肥光源实验数据管理系统。该系统结合现有的用户服务系统,可实现从用户认证、课题申请、专家审批、机时分配、数据采集、数据存储、数据分析到论文发表时引用发布的数据等实验及数据全生命周期的统一管理。

1 系统架构

合肥光源实验数据管理系统的总体架构如图 1 所示,由数据采集服务、元数据系统、文件存储系统 and 数据分析平台组成。

根据合肥光源不同实验站的实际情况进行定制开发的数据采集服务(Data Acquisition Service, DAQ Service)部署在相应实验站的本地服务器上,负责获取实验原始数据和收集相关的元数据。元数据系统由元数据摄入服务(Metadata Ingestion Service)、消息队列集群、Web 后端、元数据目录数据库和 Web 前端构成,用于将元数据信息统一编排成元数据目录,并提供标准化数据访问接口和可视化操作界面。文件存储系统负责实现实验数据文件的

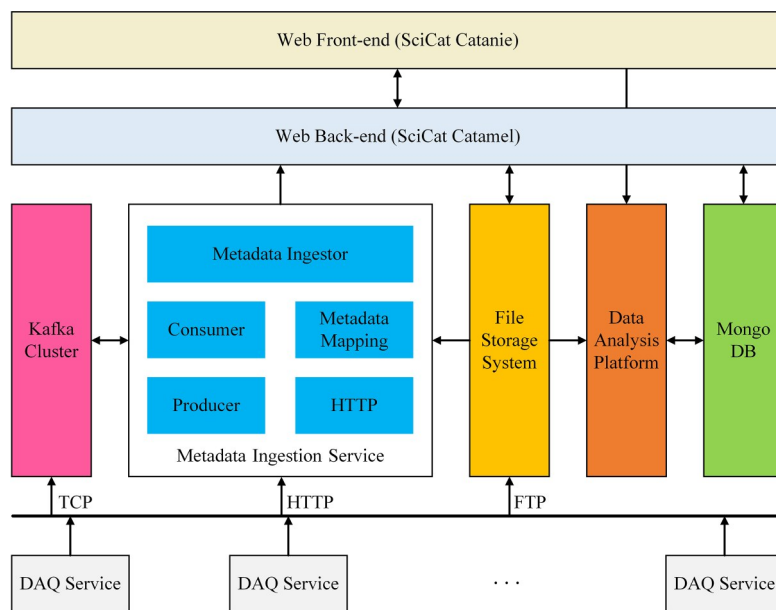


图1 基于 SciCat 的合肥光源实验数据管理系统架构图

Fig.1 Architecture of experimental data management system for HLS-II based on SciCat

传输校验、持久存储、远程下载等功能。基于 Web 的交互式离线数据分析平台由前端用户界面、后端服务软件与底层资源调度软件组成,为用户在实验结束后提供了灵活快捷、功能强大、资源丰富的离线实验数据分析环境。

2 元数据系统

元数据系统是合肥光源实验数据管理系统的核心组成部分,也是实现大科学装置实验数据全生命周期统一管理的基础。

2.1 元数据目录框架 SciCat

SciCat 是一个用于实现大科学装置实验数据全生命周期统一管理的开源元数据目录框架,由瑞典 MAX IV 光源、欧洲散裂中子源 (European Spallation Source, ESS) 与瑞士保罗谢尔研究所 (Paul Scherrer Institute, PSI) 合作研发。SciCat 的设计采用了流行的微服务架构(图2),具有耦合度低、扩展性高、可靠性强等特点。Web 后端 (Catamel API Server) 采用 Node.js 框架 LoopBack 开发。LoopBack 支持通过轻量级的数据交换格式 JSON (JavaScript Object Notation) 定义元数据模型,可以轻松扩展元数据的类型和数量,并且能够根据元数据模型自动生成具有表现层状态转换风格的应用程序接口 (RESTful API),提供给其他子系统或外部程序调用以对元数据进行增删改查操作。采用基于文

档的分布式数据库系统 MongoDB 作为元数据目录数据库。MongoDB 面向集合存储,不需要定义任何模式,因而可以存储比较复杂的数据类型,满足统一管理不同实验站的异构元数据的实际需求。Web 前端 (Catanie GUI Server) 基于 JavaScript 前端框架 Angular 进行开发,提供了基本的数据检索与展示功能,并在设计上预留了数据下载、数据分析、数据发布等有待进一步开发功能的页面。

在 SciCat 框架中,元数据摄入服务可以基于命令行界面或基于消息队列实现。元数据摄入服务负责汇集从大科学装置实验站和用户服务系统获取的元数据,再利用 Web 后端提供的 RESTful API,将元数据编排成目录存储到 MongoDB 数据库。Web 前端为用户提供了可视化界面进行数据检索、数据存档、数据下载、数据分析和数据发布等操作,并保存用户操作过程中产生的元数据信息。SciCat 框架全面记录了实验数据从开始的课题申请到最终的论文发表时引用发布的数据整个生命周期。

相对于元数据目录框架 ICAT,SciCat 的数据结构灵活,可扩展性强,因此我们基于 SciCat 框架实现元数据系统。需要指出的是,SciCat 仅开源了 Web 前端与 Web 后端的服务程序,对于元数据摄入、实验文件存储、数据发布、对接现有用户服务系统等服务模块的开发,以及元数据条目的扩展,则需要根据大科学装置实际需求自行设计实现。

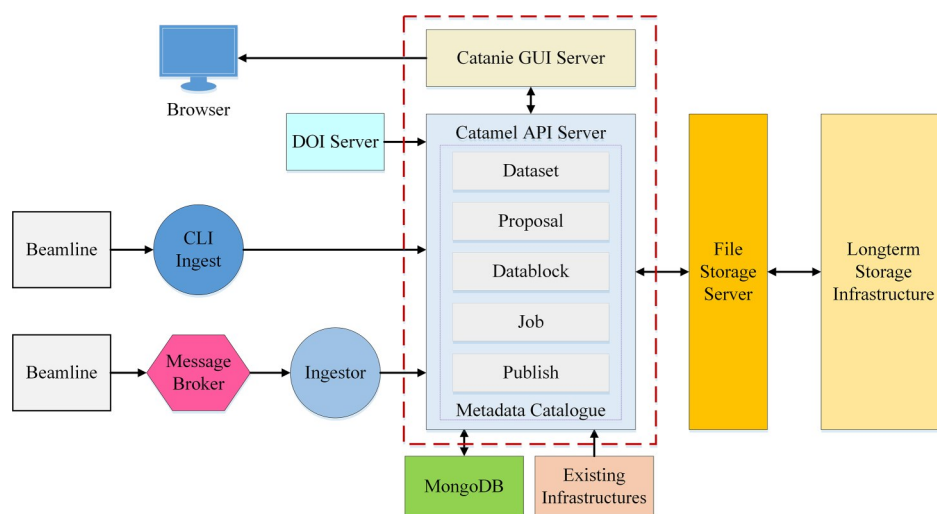


图2 元数据目录框架 SciCat 架构图
Fig.2 Architecture of SciCat metadata catalogue framework

2.2 元数据摄入

元数据目录是一个被动的系统,需要数据采集服务从不同的数据源提取元数据,再由元数据摄入服务将汇集的元数据编排成元数据目录写入数据

库,元数据摄入流程如图3所示。元数据摄入服务是元数据摄入流程中的关键部件,我们在 Node.js 环境下开发的元数据摄入服务由 HTTP 模块、消息队列生产者模块、消息队列消费者模块、元数据映射模

块和元数据摄入器模块组成。

数据采集服务分别从用户服务系统获取用户信息元数据、样品信息元数据以及实验课题信息元数据,从实验站本地服务器的文件系统或数据库提取实验结果信息元数据,从控制系统获取实验条件信息元数据。数据采集服务采用基于 TCP (Transmission Control Protocol) 的通讯协议将收集的元数据发送到消息队列 Kafka 集群。Kafka 是 Apache 软件基金会开发的一种高吞吐量的分布式发布订阅消息系统。引入消息队列 Kafka 作为中间件有以下作用:1)元数据摄入服务与数据采集服务通过遵守 Kafka 的接口约束来传输数据,从而降低元数据摄入服务与多个不同的数据采集服务之间的耦合性;2)元数据摄入服务需要间接读写数据库,数据处理速度相对单个数据采集服务的数据传输速度较慢,且数据采集服务传输数据是突发性的,Kafka 可以起到数据缓冲的作用;3)Kafka 能够将数据持久

化,直到数据已被完全处理,避免了元数据摄入过程出错,导致数据丢失。此外,考虑到某些实验站的数据采集服务可能无法集成 Kafka,我们在元数据摄入服务的开发中提供了基于 HTTP 协议的转发元数据 API。数据采集服务可以使用转发元数据 API 传输元数据至元数据摄入服务,再由元数据摄入服务的 Kafka 生产者模块转发到 Kafka 集群进行缓存。元数据摄入服务的 Kafka 消费者模块负责从 Kafka 集群读取元数据信息。元数据映射模块提取元数据信息中的实验站类型字段,依据描述实验站元数据类型与 SciCat 元数据类型对应关系的配置文件,将实验站的异构元数据统一标准化为 SciCat 规范的元数据。元数据摄入器模块从文件存储系统获取文件存储信息元数据后,通过 Web 后端提供的 RESTful API 将标准化的元数据进行元数据目录编排,即按顺序更新 MongoDB 数据库中的 User、Proposal、Sample、Dataset 等元数据模型。

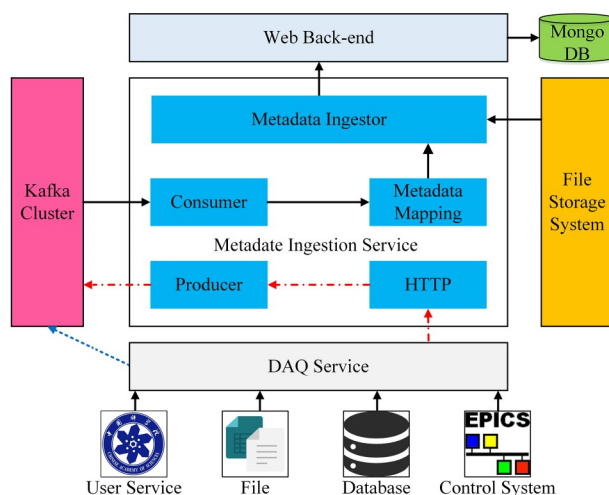


图3 元数据摄入流程
Fig.3 Process of metadata ingestion

2.3 现有用户服务系统接入

用户正式开始科学实验之前,需要在中国科学院重大科技基础设施共享服务平台 (Chinese Academy of Sciences, Large Research Infrastructures User Service Platform, CAS LSSF) 注册账号、提交实验课题申请,审核通过后在合肥光源用户实验管理系统 (HLS-II UEMS) 申请实验机时。合肥光源实验数据管理系统的开发,需要对接上述两个现有的用户服务系统,实现用户共享与单点登录。由于 SciCat 默认除本地账户外仅支持基于轻量目录访问协议 (Lightweight Directory Access Protocol, LDAP) 进行用户认证与授权访问,而目前中国科学院重大科技基础设施共享服务平台尚未提供基于 LDAP 的

认证服务,因此我们对 SciCat 的 Web 前后端进行二次开发,以实现用户共享与单点登录(图4)。

用户的个人账户信息与实验课题信息存储在中国科学院重大科技基础设施共享服务平台的数据库,合肥光源用户实验管理系统会定期将信息同步到本地数据库。用户在合肥光源实验站正式开展科学实验时,数据采集服务通过合肥光源用户实验管理系统后端 API 获取用户信息元数据与实验课题信息元数据。元数据摄入服务调用二次开发的 API 将用户信息与实验课题信息传输至合肥光源实验数据管理系统的 Web 后端。Web 后端根据用户信息元数据创建 User 模型实例,写入 MongoDB 数据库,实现了用户共享。同时,Web 后端依据实验课题编号划

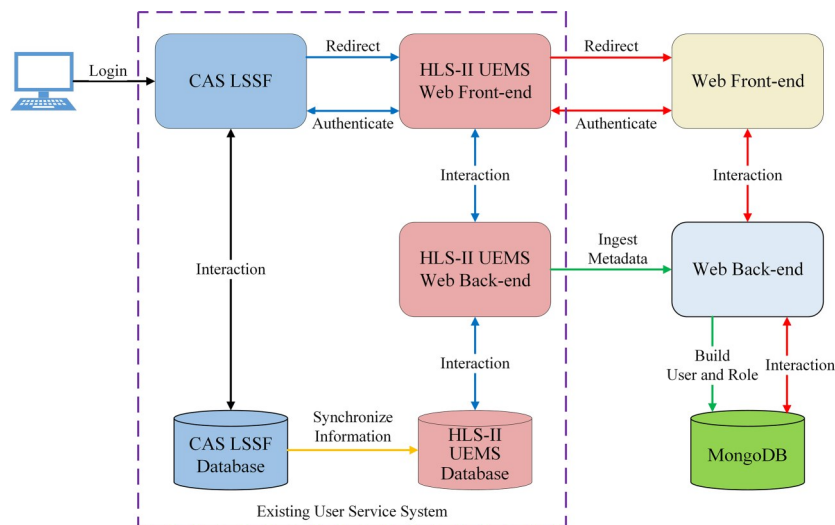


图4 接入现有用户服务系统示意图

Fig.4 Schematic diagram of interface between existing user service system and SciCat

分用户组,建立基于用户角色的数据访问授权机制。我们将用户登录 SciCat 的认证方式由基于 LDAP 认证修改为基于 MongoDB 数据库与现有用户服务系统的认证,使得用户可以在中国科学院重大科技基础设施共享服务平台成功登录后,通过网页重定向的方式登录到合肥光源实验数据管理系统,实现了不同系统之间的单点登录。

3 文件存储系统

文件存储系统负责承载海量的异构实验数据资源,需要统一数据保存格式,设计存储结构,以提供实验数据文件的长期存储、跨域远程文件下载、接入离线数据分析环境等服务。

3.1 文件格式标准化

合肥光源是一个涉及多学科交叉的大科学装置,其科学实验所产生的数据具有多源、异构、分散、海量的特点,需要将这些异构的实验数据文件格式标准化,以便统一存储管理。HDF5 (Hierarchical Data Format 5) 是一种设计用于存储和管理数据的数据文件格式。它支持层次式数据类型,可以存储复杂的异构数据;数据对象的数量和大小没有限制,为大数据提供了很大的灵活性;支持跨平台分享数据,拥有完善的生态系统。这些特点使 HDF5 在记录和存储实验数据时具有很大的优越性,因此,我们采用 HDF5 作为合肥光源实验数据管理系统的标准实验数据格式。数据采集服务负责将实验原始数据与元数据封装成标准的 HDF5 格式文件,使用 FTP 协议上传至文件存储系统指定目录。

3.2 存储目录结构设计

标准化后的实验原始数据、在线分析的处理数据以及离线分析的结果数据等实验数据资源都将保存至文件存储系统的中心磁盘。为了更高效地组织存储中的所有类型的数据文件,我们在合肥光源实际需求的基础上,参考高能同步辐射光源与保罗谢尔研究所的解决方案^[21-22],设计了文件存储系统的存储目录结构(图5)。通过实验站、用户组、日期与样品编号共同标记每次科学实验产生的数据,能够方便有效地管理海量实验数据资源。

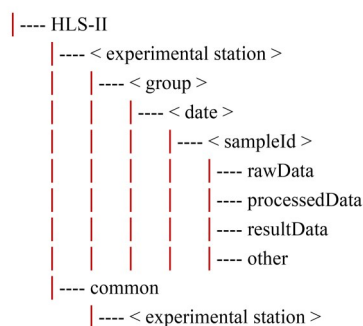


图5 存储目录结构示意图

Fig.5 Schematic of the storage directory structure

3.3 文件下载

为了满足国内外科研人员跨越时间及地理限制访问实验数据的需要,我们开发了文件下载功能,实现过程如图6所示。用户在 Web 前端数据下载功能页面选择要下载的文件,Web 前端调用后端提供的文件下载 API 提交待下载的文件信息;Web 后端对前端提交的文件下载信息进行封装后,发送至文件存储系统;文件存储系统开始打包待下载的文件,同

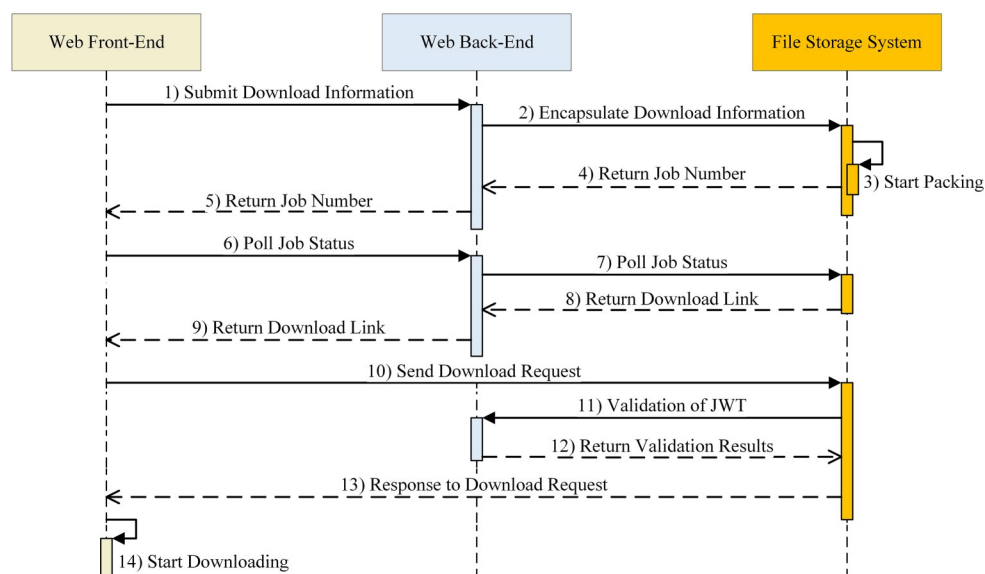


图6 文件下载序列图

Fig.6 Sequence diagram of file downloading

时返回文件打包作业编号;Web前端依据文件打包作业编号轮询作业状态,直至文件存储系统完成文件打包返回下载链接;Web前端在下载链接末尾附加携带有身份信息和过期时间的JWT(JSON Web Token)作为参数,直接向文件存储系统发起下载请求;文件存储系统向Web后端请求检验JWT的有效性,验证通过后响应Web前端的下载请求;Web前端开始下载文件到用户本地设备。JWT是一种开放标准(RFC 7519),用于在网络应用环境间以JSON对象的形式安全地传输信息。使用JWT对文件下载的有效性进行验证,能够提升数据访问的安全性。

4 数据分析平台

在实验完成后,用户需要根据实验原始数据,以及探测器刻度、数据采集时探测器的状态、样品特征信息、控制系统参数等元数据,搭建专用的数据处理环境,依托大规模计算资源进行离线分析,以得到精度更高、效果更好的结果数据。为了实现用户对实验数据的分析、识别、挖掘、分类等多方面的应用需求,我们开发了基于Web的交互式离线数据分析平台,由前端用户界面、后端服务软件、底层资源调度软件三部分组成(图7)。

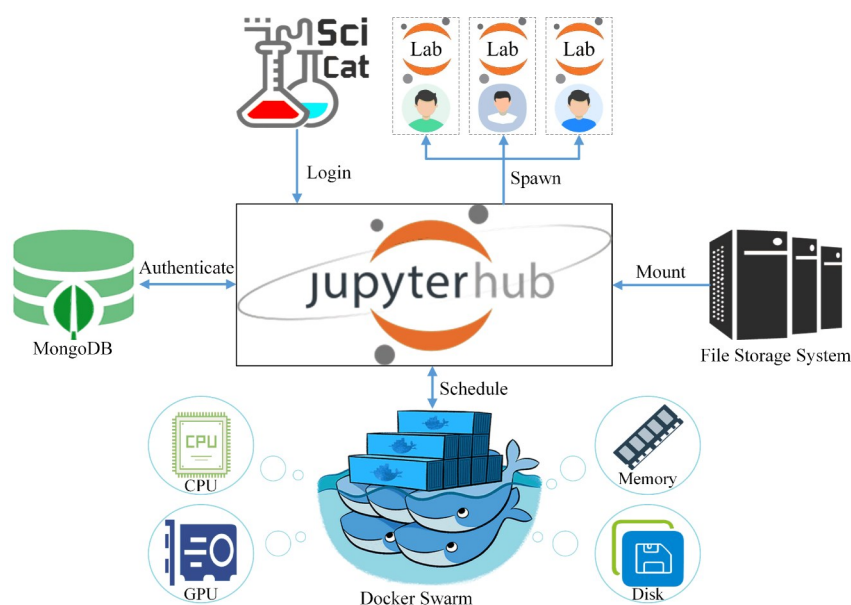


图7 交互式数据分析平台架构图

Fig.7 Architecture of interactive data analysis platform

前端用户界面基于 JupyterLab 实现。JupyterLab 是一个基于 Web 的交互式集成开发环境,支持多种流行编程语言、集成度高、扩展性强、具有灵活而友好的用户界面,可以定制化配置以支持数据科学、科学计算和机器学习中的各种工作流程。我们在 JupyterLab 的基础版本上进行二次开发,集成了 Python、Julia 与 R 语言编程环境, TensorFlow、PyTorch 等主流深度学习框架,以及容器化条件下的基于 CUDA (Compute Unified Device Architecture) 深度学习的 GPU 加速库 cuDNN。此外,我们将前端用户界面 JupyterLab 的访问入口集成到 Web 前端,方便用户使用。

后端服务软件基于 JupyterHub 实现。JupyterHub 用于为多用户使用 JupyterLab 环境提供服务,负责创建、管理、代理多个 JupyterLab 实例。JupyterHub 默认支持使用基于 PAM 或基于 OAuth 的认证器为用户登录提供身份认证服务。我们开发了基于 MongoDB 数据库的 JupyterHub 认证器,使得用户无需输入账号密码即可从 Web 前端通过网页重定向登录到 JupyterLab 的使用界面,实现了与元数据系统的统一用户认证。用户首次登录时, JupyterHub 根据配置文件自动创建 JupyterLab 实例,并将用户保存在文件存储系统的实验数据目录挂载到 JupyterLab 实例。

底层资源调度软件基于 Docker Swarm^[23]实现。Docker Swarm 是 Docker 官方提供的集群管理工具,负责将 Docker 主机池抽象为单个虚拟 Docker 主机,统一管理主机上的各种资源。我们将 JupyterLab 与 JupyterHub 制作成 Docker 镜像,通过 SwarmSpawner 建立关联,以容器的形式运行在数据分析平台集群

上。Docker Swarm 负责管理调度数据分析平台集群的 CPU、GPU、内存与磁盘等计算资源,根据后端服务软件的需求进行资源分配、资源约束以及资源回收处理。

5 系统测试

合肥光源的软 X 射线成像实验站(BL07W)是一个用于研究高分辨三维结构、元素吸收边和谱学成像的科学实验平台。我们以软 X 射线成像实验站作为系统测试平台,对开发的合肥光源实验数据管理原型系统进行了数据采集、传输、存储、下载、分析与发布等功能测试。

基于 Python 开发的数据采集服务部署在软 X 射线成像实验站本地服务器,通过控制系统局域网与文件存储系统、元数据系统进行数据传输;标准化的实验数据保存至文件存储系统;元数据摄入服务利用 Web 后端的 RESTful API,将数据集(dataset)与实验课题、实验样品、用户角色、数据文件、管理策略等元数据建立关联,并为数据集分配用于数据发布的唯一持久标识符(Persistent Identifier, PID)写入 MongoDB 数据库。通过认证的用户和管理员可以从 Web 界面访问元数据目录(图 8)。

用户在基于 Web 的交互式离线数据分析平台上,利用提供的数据分析环境与计算资源,对实验数据进行了电子计算机断层扫描(Computed Tomography, CT)图像三维重建处理,重建结果如图 9 所示。

测试表明,基于 SciCat 的合肥光源实验数据管理系统的各项功能符合设计预期,具有工程可用性。



图 8 合肥光源实验数据管理系统 Web 界面
Fig.8 Web interface of HLS-II experimental data management system

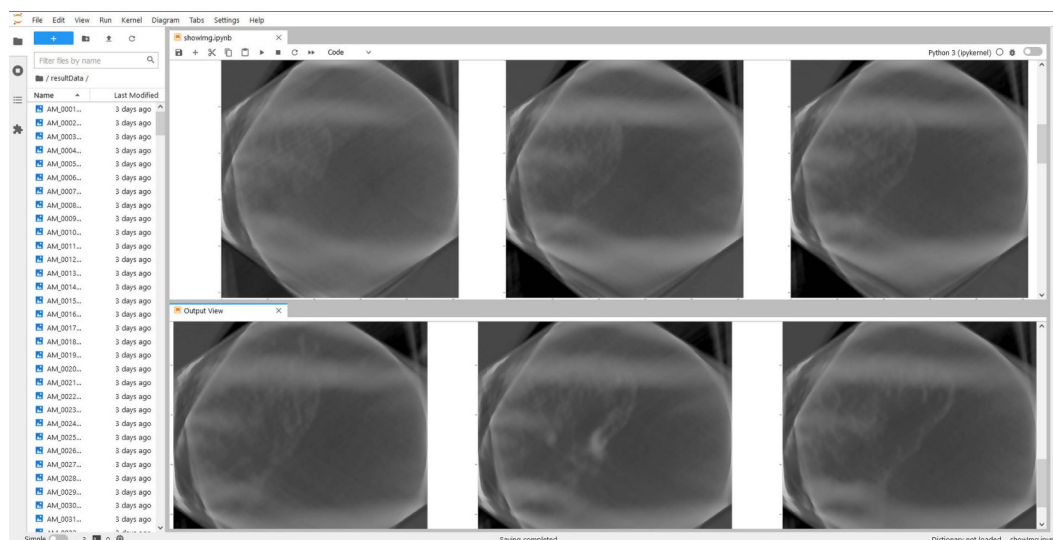


图9 CT图像三维重建的结果展示
Fig.9 Display of 3D reconstruction results of CT images

6 结语

本文描述了基于开源元数据目录框架SciCat的合肥光源实验数据管理系统的设计、开发与测试过程。通过设计开发数据采集服务、元数据系统、文件存储系统与数据分析平台,以及接入现有的用户服务系统,展示了涵盖用户认证、课题申请、专家审批、机时分配、数据采集、数据存储、数据分析、数据发布的实验及数据全生命周期的流程管理和远程数据服务功能,并在合肥光源的软X射线成像实验站进行部署和功能测试,达到了设计要求。目前系统的数据发布功能还处在内部测试阶段,下一步我们将与数据对象识别码(Digital Object Identifier, DOI)注册机构合作以实现实验数据的全球网络发布与共享,并针对数据下载服务性能、数据处理效率等软件性能开展测试和优化工作。

致谢 感谢合肥光源软X射线成像实验站的刘刚研究员、陶夏禹同学在系统测试方面提供的帮助。

作者贡献说明 林广负责文章起草和最终版本的修订;张大地负责文件存储系统开发与网络技术支持;张震负责数据采集系统开发;贺博负责用户服务系统对接;孙晓康负责资料收集整理及文章结构设计;刘功发负责研究提出及设计。

参考文献

- 何佩瑶, 赵越, 刘宇, 等. 相变材料 $\text{Ge}_2\text{Sb}_2\text{Te}_5$ 的X射线辐照致变效应研究[J]. 核技术, 2021, 44(3): 030102. DOI: 10.11889/j.0253-3219.2021.hjs.44.030102.
HE Peiyao, ZHAO Yue, LIU Yu, *et al.* Study on X-ray irradiation induced effect of phase-change material $\text{Ge}_2\text{Sb}_2\text{Te}_5$ [J]. Nuclear Techniques, 2021, 44(3): 030102. DOI: 10.11889/j.0253-3219.2021.hjs.44.030102.
- 袁静, 罗立强. 同步辐射微区X射线荧光和吸收谱技术在大气、土壤和动植物分析中的应用[J]. 核技术, 2014, 37(8): 080101. DOI: 10.11889/j.0253-3219.2014.hjs.37.080101.
YUAN Jing, LUO Liqiang. Synchrotron μ -XRF and XAFS in element distribution and speciation of air, soil and biological samples[J]. Nuclear Techniques, 2014, 37(8): 080101. DOI: 10.11889/j.0253-3219.2014.hjs.37.080101.
- 亢舜彤, 吴桐, 陈焯辉, 等. 鼠脑神经网络的同步辐射3D成像研究[J]. 核技术, 2020, 43(7): 070101. DOI: 10.11889/j.0253-3219.2020.hjs.43.070101.
KANG Shuntong, WU Tong, CHEN Zhuohui, *et al.* 3D imaging of rat brain neural network using synchrotron radiation[J]. Nuclear Techniques, 2020, 43(7): 070101. DOI: 10.11889/j.0253-3219.2020.hjs.43.070101.
- Wang C P, Steiner U, Sepe A. Synchrotron big data science[J]. Small, 2018, 14(46): 1802291. DOI: 10.1002/sml.201802291.
- Bethel E W. Towards a data-centric research and development roadmap for large-scale science user facilities[C]//2017 IEEE 13th International Conference on e-Science (e-Science). October 24-27, 2017, Auckland, New Zealand. IEEE, 2017: 462 - 464. DOI: 10.1109/eScience.2017.72.
- Vondrous A, Jejkál T, Ressmann D, *et al.* Beamline data management at the synchrotron anka[ED/OL]. 2021-10-01. <https://accelconf.web.cern.ch/PCaPAC2014/papers/>

- fpo034.pdf.
- 7 Arkilic A, Allan D, Chabot D, *et al.* Databroker: an interface for NSLS-II data management system[ED/OL]. 2021-05-10. <https://accelconf.web.cern.ch/ICALEPCS2015/papers/wed3o02.pdf>.
 - 8 Moriyama K, Nakatani T. A data management infrastructure for neutron scattering experiments in J-PARC/MLF[ED/OL]. 2021-10-11. <https://accelconf.web.cern.ch/ICALEPCS2015/papers/wepgf060.pdf>.
 - 9 Scientific data management at diamond light source[EB/OL]. 2021-08-12. <https://www.slideserve.com/betsy/scientific-data-management-at-diamond-light-source/>.
 - 10 ISIS - ICAT project[EB/OL]. 2021-08-12. <https://icatproject.org/collaboration/facilities/isis/>.
 - 11 张俊荣. CSNS/HEPS 科学数据处理软件的设计和开发[C]. 遵义: 第十九届全国核电子学与核探测技术学术年会, 2019.
ZHANG Junrong. Software architecture of data analysis & management for CSNS[C]. Zunyi: The 19th Annual National Conference on Nuclear Electronics and Nuclear Detection Technology, 2019.
 - 12 Tian H L, Zhang J R, Yan L L, *et al.* Distributed data processing and analysis environment for neutron scattering experiments at CSNS[J]. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2016, **834**: 24 - 29. DOI: 10.1016/j.nima.2016.07.043.
 - 13 ICAT project[EB/OL]. 2021-08-12. <https://icatproject.org/>.
 - 14 Sjöblom P, Amjad A, Bell P J, *et al.* Stable operation of the MAX-IV laboratory synchrotron facility[C]// Proceedings of the 17th International Conference on Accelerator and Large Experimental Physics Control Systems. New York, NY, USA, 2019: 530 - 535. DOI: 10.18429/JACoW-ICALEPCS2019-MOPHA133.
 - 15 Hardion V, Barsek A, Bolmsten F, *et al.* The MAX IV laboratory scientific data management[C]//Proceedings of the International Conference on Accelerator and Large Experimental Physics Control Systems. 2017.
 - 16 Wang C P, Yu F, Liu Y Y, *et al.* Deploying the big data science center at the Shanghai Synchrotron Radiation Facility: the first superfacility platform in China[J]. Machine Learning: Science and Technology, 2021, **2**(3): 035003. DOI: 10.1088/2632-2153/abe193.
 - 17 齐法制, 黄秋兰, 胡皓, 等. 高能同步辐射光源科学数据处理平台规划与设计[J]. 数据与计算发展前沿, 2020, **2**(2): 40 - 58.
QI Fazhi, HUANG Qiulan, HU Hao, *et al.* The design of science data platform for high energy photon source[J]. Frontiers of Data & Computing, 2020, **2**(2): 40 - 58.
 - 18 怀平. 上海自由电子激光(SHINE)及其科学数据系统规划[C]. 遵义: 第十九届全国核电子学与核探测技术学术年会, 2019.
HUAI Pin. Overview of SHINE and its data system[C]. Zunyi: The 19th Annual National Conference on Nuclear Electronics and Nuclear Detection Technology, 2019.
 - 19 SciCat project[EB/OL]. 2021-08-12. <https://scicatproject.github.io/>.
 - 20 Hu H, Qi F Z, Zhang H M, *et al.* The design of a data management system at HEPS[J]. Journal of Synchrotron Radiation, 2021, **28**(Pt 1): 169 - 175. DOI: 10.1107/S1600577520015167.
 - 21 Ingest manual PSI[EB/OL]. 2021-08-12. <https://scicatproject.github.io/documentation/Ingestor/ingestManual.html#sec-3-2-3/>.
 - 22 Swarm mode overview[EB/OL]. 2021-08-12. <https://docs.docker.com/engine/swarm/>.