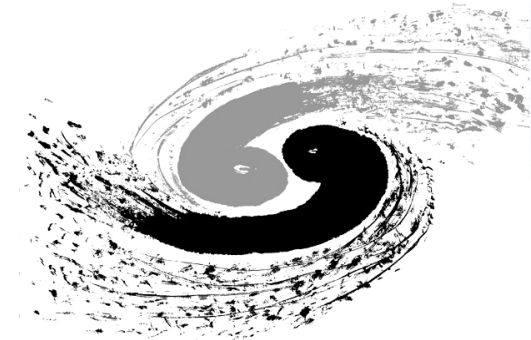




中国科学院高能物理研究所
INSTITUTE OF HIGH ENERGY PHYSICS, CAS



CCAST博士后面试报告

报告人： 张水涵
报告日期： 2025.01.11

目录

01

个人简介

02

以往工作成绩

03

未来研究计划



个人简介

张水涵

本科毕业于吉林大学计算机系

2020年进入中国科学院高能物理研究所 直博

GPA: 3.91/4.0

- **导师:** 朱科军研究员
- **参与实验:** JUNO, TAO, HEPS-BPIX4
- **获奖情况:**
 - 2022-2023学年年度三好学生
 - 2023-2024学年年度三好学生
 - 2023-2024学年中国科学院大学学业一等奖学金
 - 2024年12月第二十四届所长奖学金表彰奖

研究方向: 高性能数据获取与处理

- **工程维度:** 参与多项实验
- **系统维度:** 深度参与数据流和在线系统研发
- **技术维度:** 熟练掌握C++、Python编程, ROOT框架, 机器学习算法和数据库技术

个人简介

参与国内外会议：

获得OUTSTANDING POSTER和优秀报告

- **TIPP 2023** - Design and Implementation of the DAQ System for the HEPS-BPIX 6M Detector, The 6th Technology and Instrumentation in Particle Physics Conference, 口头报告
- **RT 2024** - A ROOT-based General Online Data Visualization System, The 24th IEEE Real Time Conference, Mini oral and Poster
- **CEPC workshop 2024** - Artificial Intelligence Applied Researches on Online Monitoring System, Poster
- **第二十届全国科学计算与信息化会议**, 台中微子实验数据获取系统的设计与开发, 口头报告
- **第二十一届全国核电子学与核探测技术学术年会**, HEPS-BPIX 6M探测器DAQ系统的设计与开发, 口头报告
- 其他会议：多次参与**JUNO**合作组会议, 并作TAO TDAQ系统工作进展口头报告

发表论文：

以第一作者发表论文2篇, 合作作者发表论文3篇

- **Zhang, S.**, A ROOT-based General Online Data Visualization System. *IEEE Transaction on Nuclear Science* (2024). (已发表, 一作)
- **Zhang, S.**, Chen, C., Ji, X. *et al.* Design and implementation of TAO DAQ system. *Radiation Detect Technol Methods* (2024). <https://doi.org/10.1007/s41605-024-00496-3> (已发表, 一作)
- Wu, Y., Yu, Z., **Zhang, S.** *et al.* The Online Software of JUNO Data Acquisition System. *IEEE Transaction on Nuclear Science* (2024). (已发表, 合作)
- Chen, C., Peng, Y., Zeng, T., **Zhang, S.** *et al.* Design and Development of JUNO DAQ Data Flow Software. *IEEE Transaction on Nuclear Science* (2024). (已发表, 合作)
- Yang, X., Ji, X., Zhu, K., **Zhang, S.** *et al.* Design and Implementation of DAQ System for HEPS-BPIX4. *IEEE Transaction on Nuclear Science* (2024). (已发表, 合作)

研究工作概要

➤ TAO台中微子实验数据获取系统研制

- **核心数据流方案研究实现**：设计优化子系统的数据读出和处理方案
- **难点数据流全链路集成与性能优化**：全规模数据量性能测试优化、联调测试等
- **创新**
 - **通用的远程监控系统研究**：大规模数据量下系统可扩展性
 - **在线实时异常检测机制研究**：异常数据实时反馈报警

➤ 参与其他实验

- JUNO-SPMT数据获取系统的研发
- HEPS-BPIX4分布式数据获取系统的研发

需求分析

架构设计

系统研发

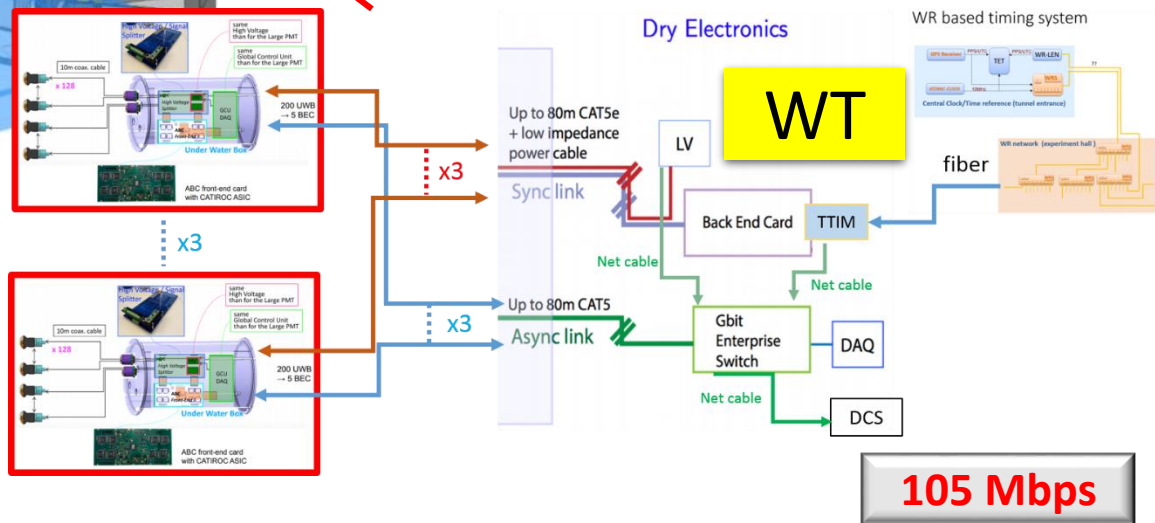
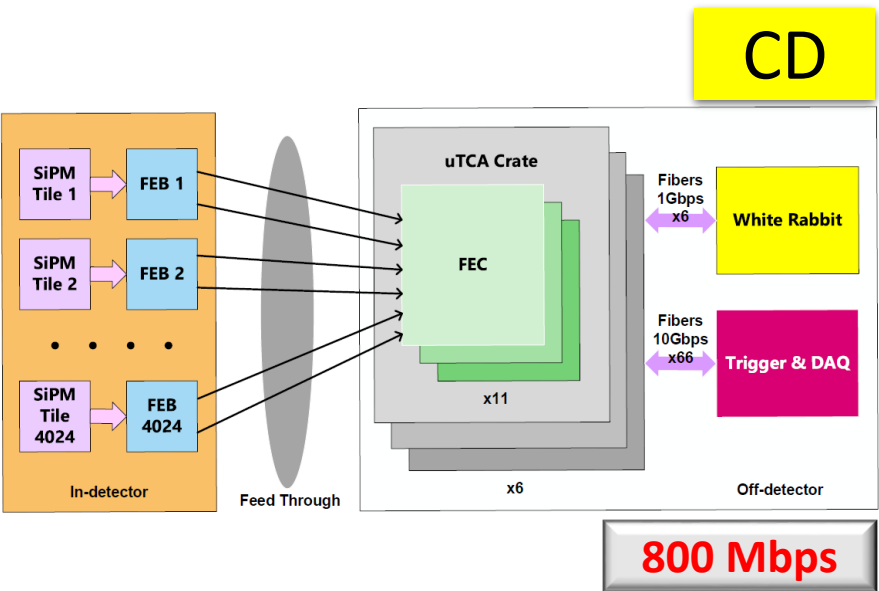
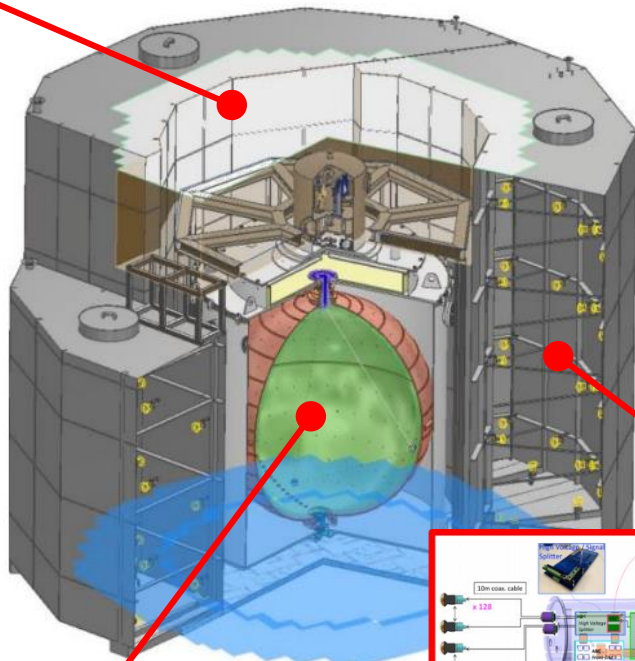
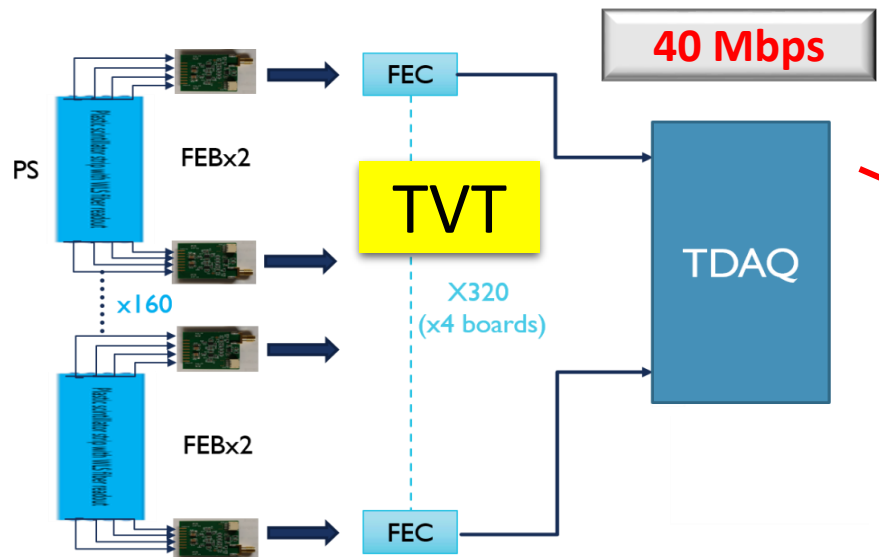
联调测试

实验运行

研究背景简介

TAO-WT和JUNO-SPMT共享相同的电子学结构

- 读来自三个独立探测器电子学或触发系统的原始数据
- 压缩数据存盘带宽到 **<100Mbps**

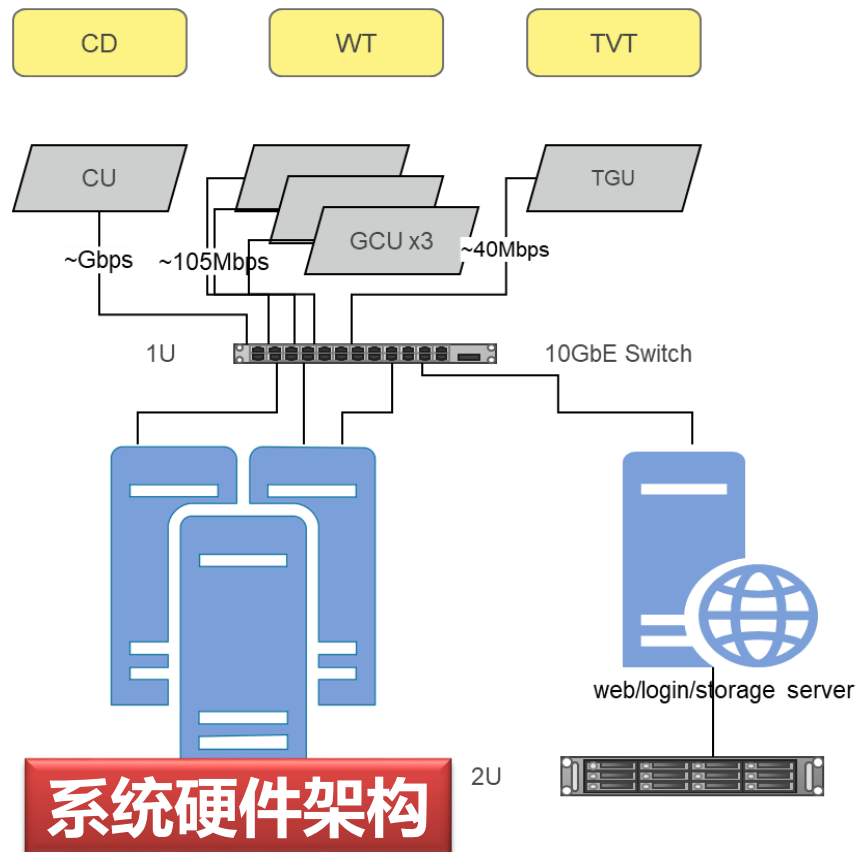
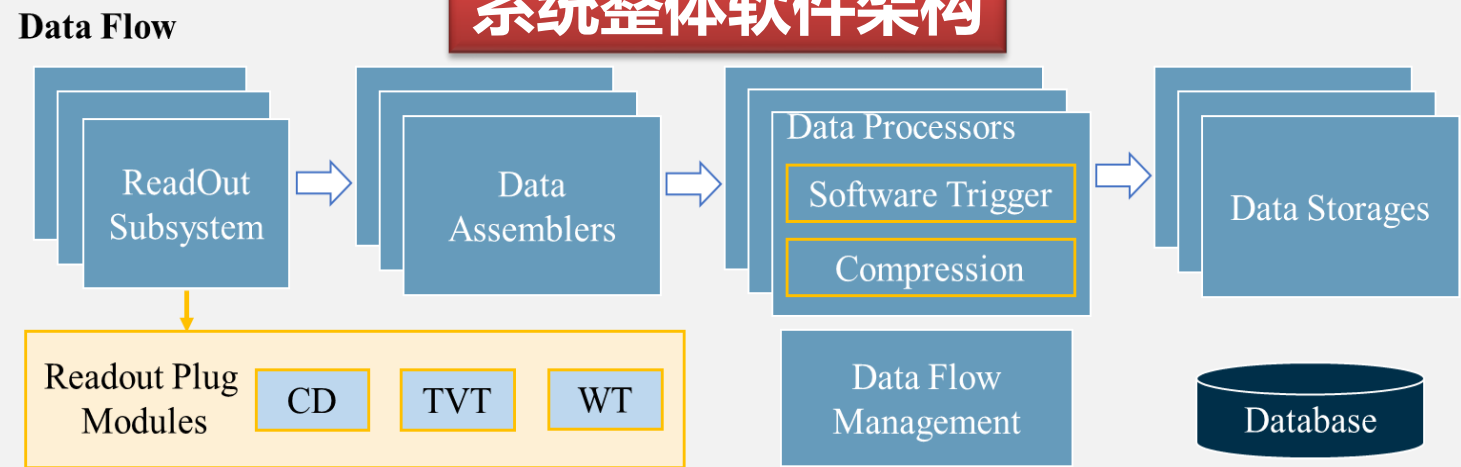
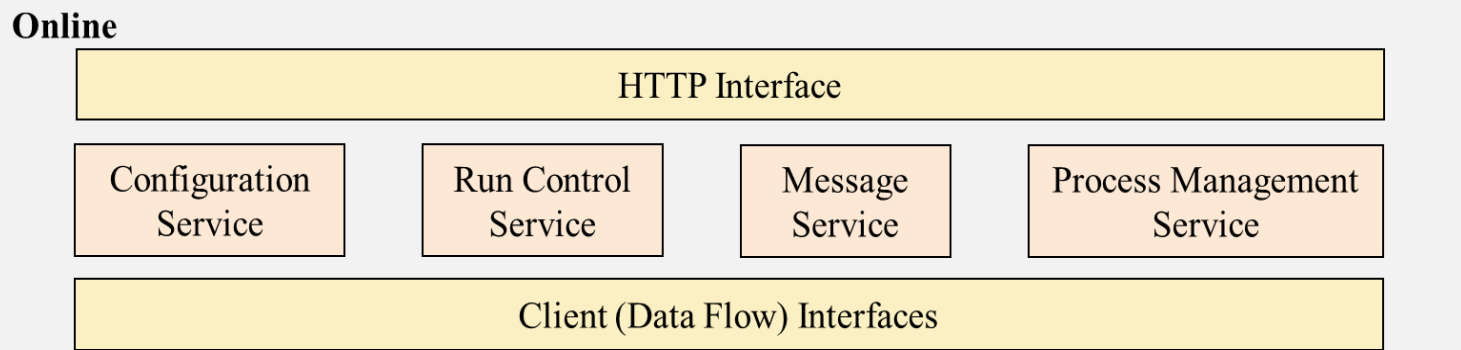


研究成果：数据获取系统架构设计与实现

- 研究内容：兼容JUNO DAQ已有架构，设计完整的适应TAO实验的定制化DAQ系统架构
- 优势：共享数据文件格式和传输系统；集成监测方案，共同运行维护；多接口设计方便实验定制化和扩展，提高系统灵活性

设计搭建基于插件式的模块化分布式系统软件架构

系统架构融合，无缝集成，完美适配于TAO实验

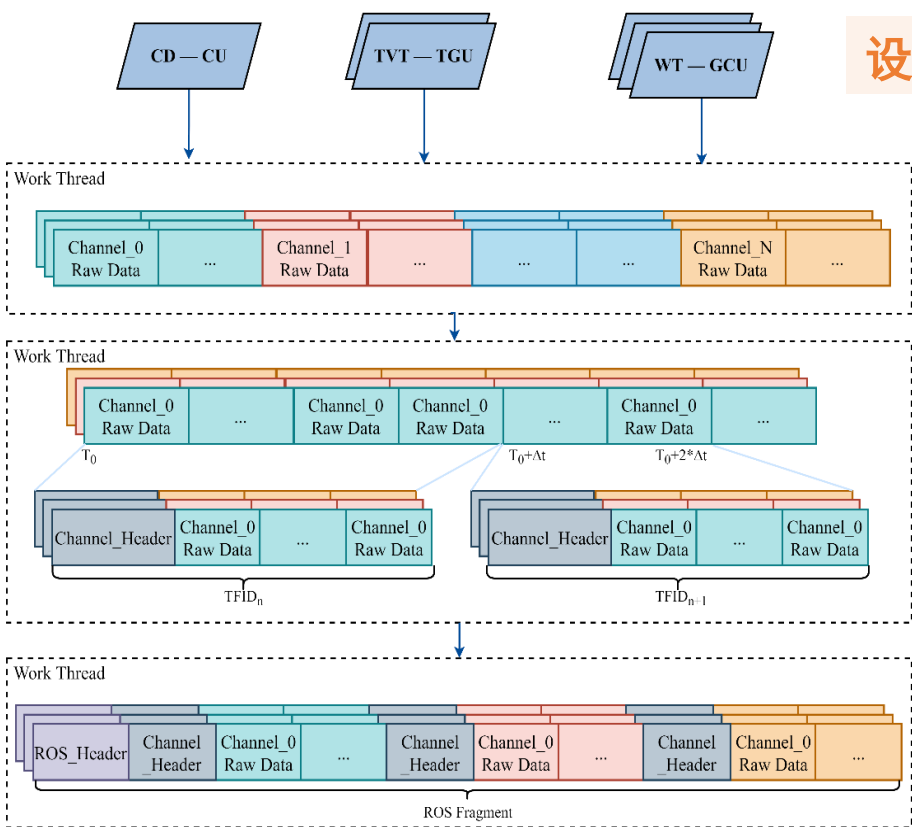


研究内容 → 定制化电子学和触发系统配置，实现TQ、波形等数据读出

已实现多平台部署，成功获取超过30TB的TQ和波形数据

如何实现上万维通道并行数据读出和异常处理机制**是一项研究难点**

设计三级业务异常处理机制



数据读出业务
Epoll高并发的数据读出
&支持TCP/IPbus协议

预处理业务
通道级的数据处理，检查数据格式和时间戳，异常数据包报警并跳过

一级组装业务
设计强制组装策略应对低数据量或无数据通道

Data Stream	Readout module	Interface	Readout Bandwidth
CD	1 CU	SiTCP+10 GbE	~800 Mbps
WT	128 PMTs/GCU	IPbus/TCP + GbE	~105 Mbps
TVT	2 TGUs	SiTCP + GbE	~40 Mbps

读出性能优化策略

- 原始数据不对齐：拆分原始数据包，只解码一次数据
- 预处理和缓存：缓存1s数据后向后传递
- 算法逻辑重整优化：减少内存拷贝次数 & 移位操作

数据读出接口

CPU资源：
~180个核 → ~125
数据检查：
5ms → 180us

数据流核心读出和电子学配置功能就绪，配合多个测试平台同步联调、验证、优化

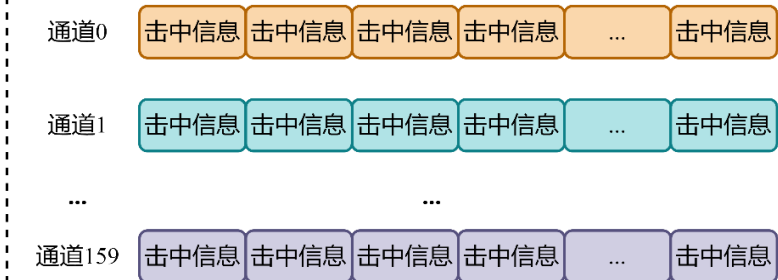
研究内容

专用的数据处理算法，实现高性能并行在线数据处理

不同类型数据处理算法的实现和算法正确性的验证是一项**关键技术难点**

不同颜色代表不同通道

层符合软件触发



基于已知映射关系

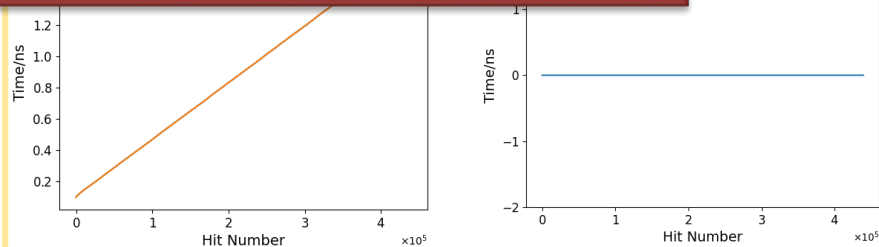


N=2的nHit软件触发算法

nHit软件触发



实现多类型的并理事例处理流



交叉验证触发算法的正确性

修复问题

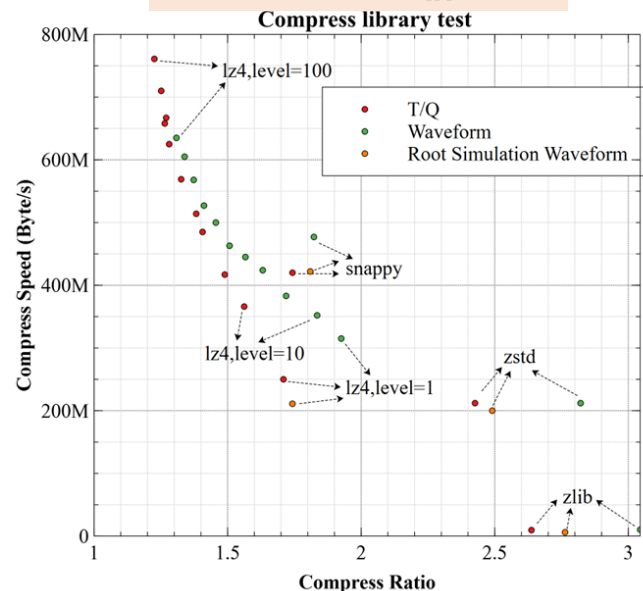
- 漏触发&误触发
- 触发窗中相同通道重复计算



解决与优化策略:

- 边界问题特殊分析
- 事例打包逻辑重整

数据压缩



数据流已集成多种数据处理算法，验证了数据流接口的灵活性和算法的正确性

研究成果：在线软件研究与实现

在线软件

研究目的 实现基于web的远程监控平台，确保**实时、准确、全面**的数据动态监测和控制管理

关键问题：运行控制、配置管理、实时数据监测、在线故障分析显示、发布报警等

实现多维度的数据监测方案

基于Exporter、Prometheus和Grafana的系统状态监测

基于ROOT的在线数据质量监测



Redis



Kafka

优化策略

- 分布式架构
- 增量更新
- 负载均衡

System Status Monitoring

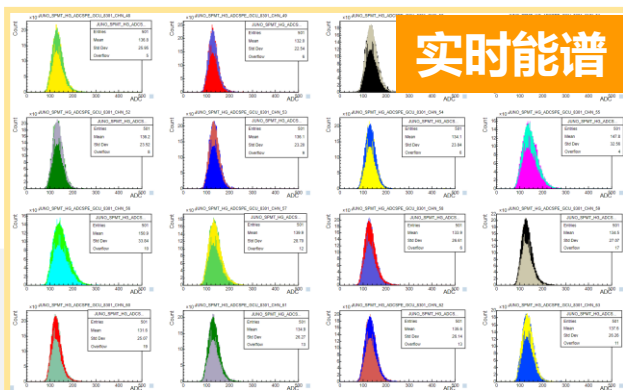


在线监测系统架构

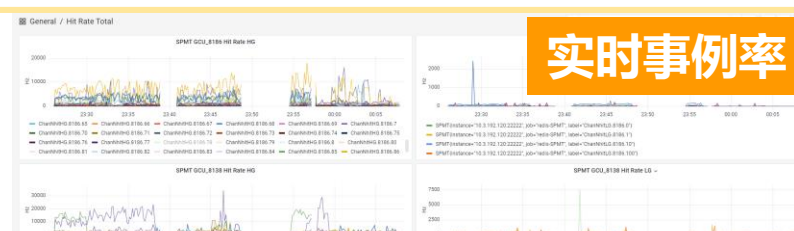
Online Data Monitoring



监控模块具有高度可复用性，成功集成到JUNO和TAO DAQ web GUI，为JUNO实验提供**核心技术**



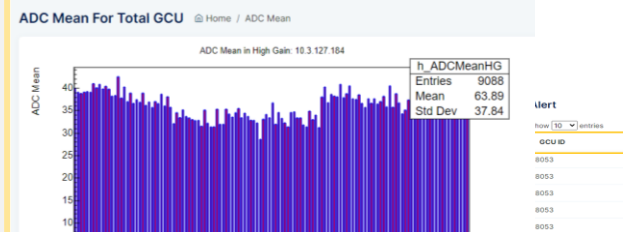
实时能谱



实时事例率



运行控制



GCUI ID	Error Type	Time
8053	GCUWAIT_FOR_TIMEOUT!	2023-06-19 02:57:53
8053	GCUWAIT_FOR_TIMEOUT!	2023-06-19 02:58:03
8053	GCUWAIT_FOR_TIMEOUT!	2023-06-19 02:58:14
8053	GCUWAIT_FOR_TIMEOUT!	2023-06-19 02:58:31
8053	GCUWAIT_FOR_TIMEOUT!	2023-06-19 02:58:42

支持按列排序

支持按关键字检索

在线软件已经成功与JUNO现有架构集成，整套监测系统已部署于4大测试平台，应用于10次SPMT关灯测试，完成最多**同时25600个通道成功监控**

实时报警

创新点：通用在线数据可视化系统 (ROBOT)

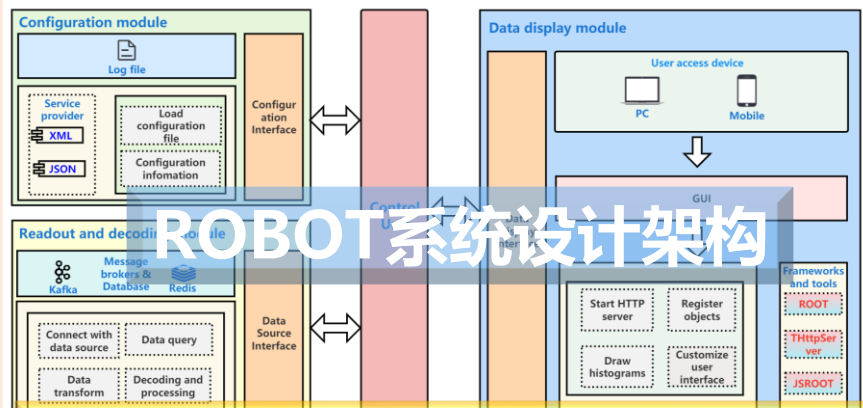
研究目的

解决不同实验**重复设计开发**在线监测系统的痛点

- 方便对接不同实验
- 支持不同数据源采集
- 基于ROOT框架使用友好

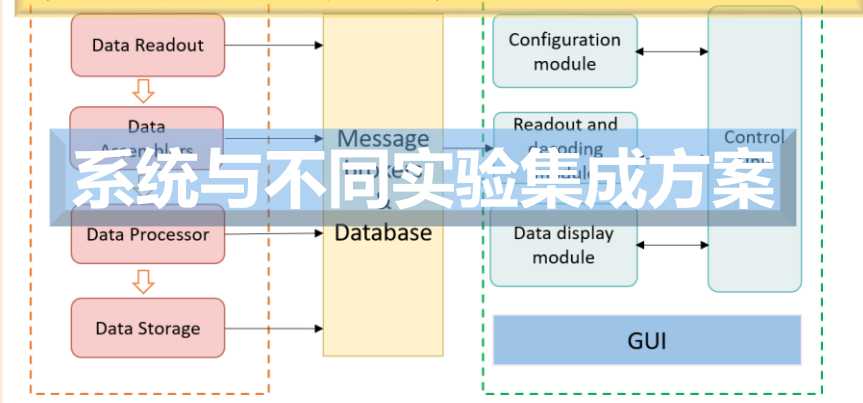
研究设计通用的数据可视化系统，实现高效便捷的联调及实验监测

ROBOT系统架构和集成方案



ROBOT系统设计架构

独立设计实现系统整体架构和功能



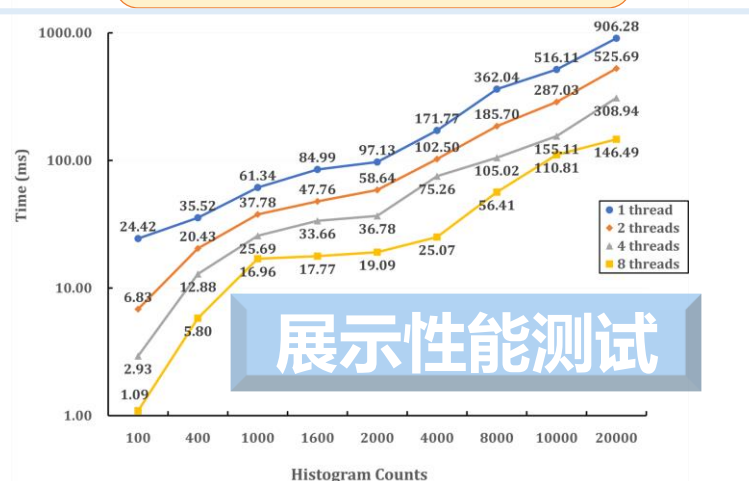
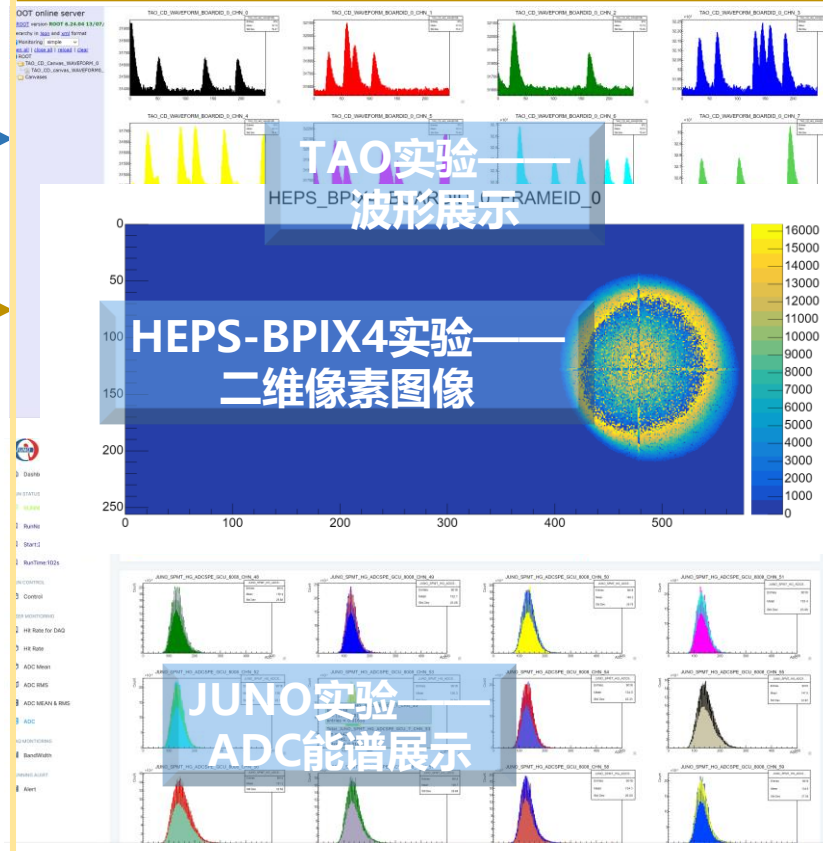
系统与不同实验集成方案

丰富的配置管理：
JSON和XML

多类型数据读出接口
基于Redis和Kafka

多种数据展示方案
TH1D和TH2D

已有具体实验应用



展示性能测试

创新点：通用在线数据可视化系统 (ROBOT)

扩展性研究 ➔ 面向**大规模系统**下上万维数据通道，系统响应的实时性和稳定性

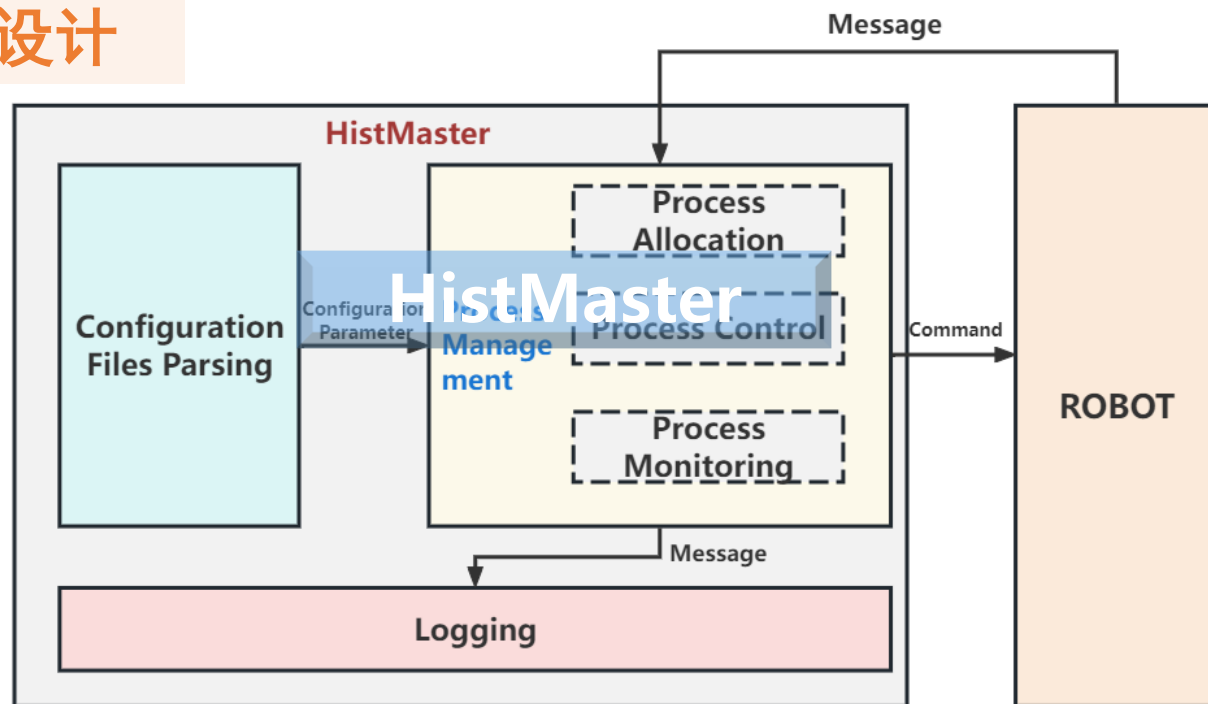
系统优化：异步处理，缓存机制，水平扩展

部署策略：级联配置 + “HistMaster”设计

Section	Parameter	Description
Common	DATASOURCE	Define data source types: Redis, Kafka, etc
	IP	Define IP of the data source
	PORT	Define port of the data source
	WEB_PORT	Define port displayed on the webpage

Detectors	SUB_DECTOR	Define the name of the sub-detector plugin
	BOARD_NUM	Define the number of boards for display
	CHANNEL_NUM	Define the number of channels for each board
	GROUP_NUM	Specify the number of processes for large-scale data volume
	DRAW_TYPE	Define the histogram types for display
	CANVAS_PICS	Define the layout for each canvas
...	...	

级联配置格式



级联配置——提供多系统配置管理方式

➤ 共享通用部分 & 定制化展示功能

系统在终端用户的有益反馈下正逐渐完善，将在实验运行中正式应用于JUNO和TAO实验

创新点：基于机器学习的实时直方图异常检测

研究目的

解决大规模系统下统计+人工检查异常数据点**耗时耗力，易漏检**的痛点

基于人工智能和机器学习实现**自动化实时异常检测**

数据集准备
人工标注~20w条样本
生成自动化脚本辅助

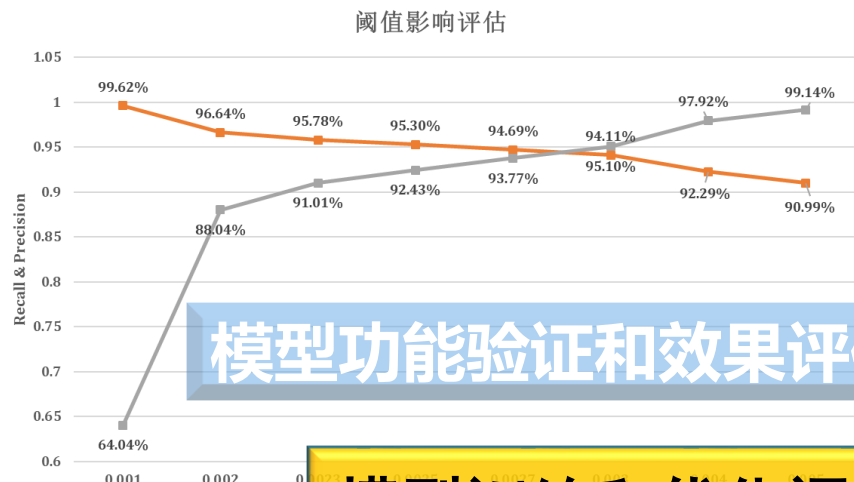
模型搭建训练
AE无监督学习模型

模型效果评估
多维度模型参数
及结构优化

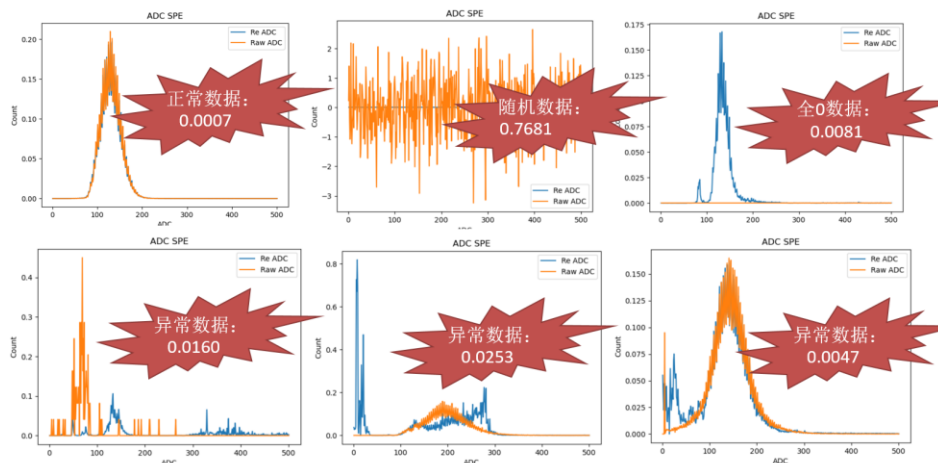
CPU环境	直方图数量	检测时间	精准率
AE模型	2560	0.2324s	0.64
统计方案	2560	2.3433s	0.33

实际部署效果：

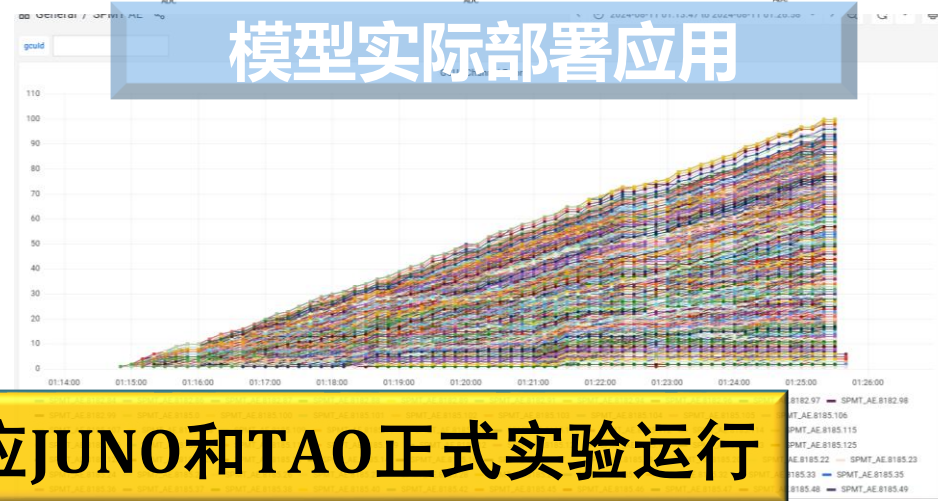
- 减少查看60%的通道
- 99%异常直方图被模型成功预测



模型训练和优化评估基本完成，匹配适应JUNO和TAO正式实验运行



模型实际部署应用

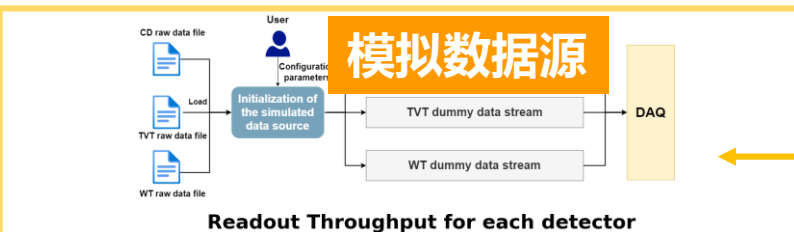


研究成果：全链路数据流性能测试与集成部署

研究内容：开发模拟数据源以及搭建测试平台完成系统集成部署

模拟全规模数据量测试&联调测试，验证和优化关键性能指标

实际联调工作多地同步开展，完成通信接口验证，提供了完整的测试平台，友好的测试界面和各种物理量的实时显示



模拟全规模数据量验证全链路数据流性能

模拟测试中全链数据流可稳定运行一个月以上



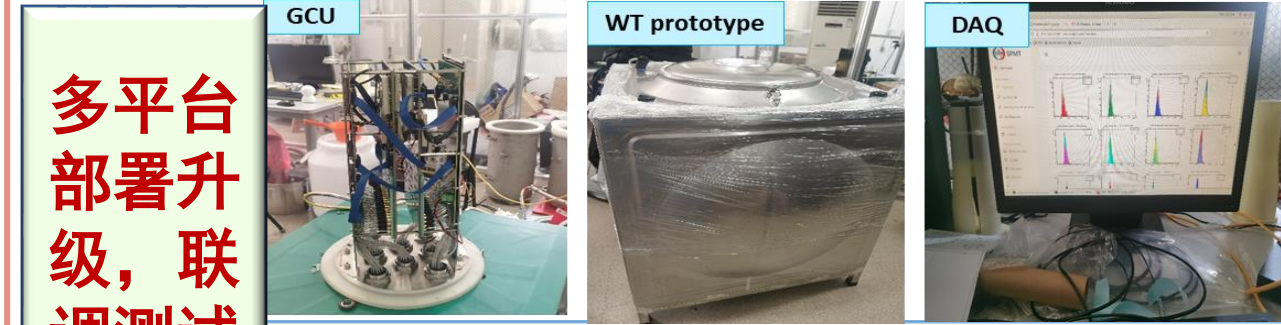
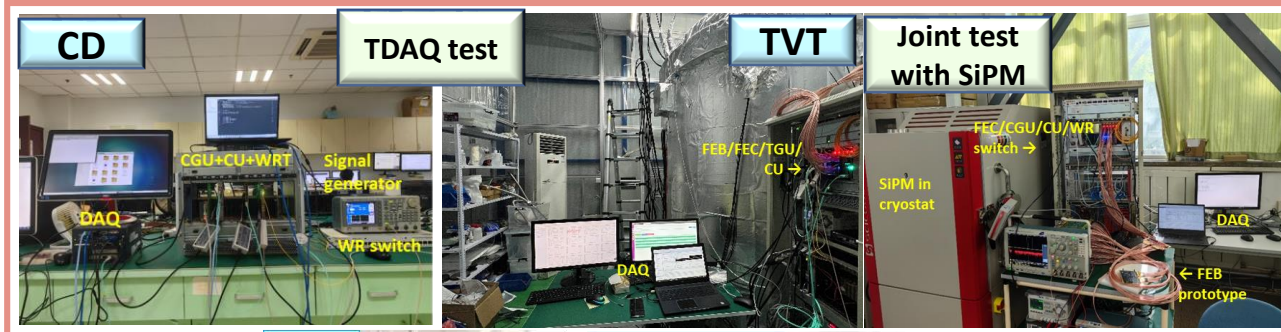
数据流读出性能测试

全链路系统多平台下联调测试验证通信接口和各级系统功能

Detector	Channel Numbers	Hit Rate	Data Window	Trigger Window	Nhit Value	T1	T2	CPU Cores
WT-SortAlg_Single	100	5.1kHz	500ns	300ns	5	2832μs-2922μs	2997μs-3057μs	1(0.3-0.4)
WT-SortAlg_Total	300	5.1kHz	500ns	300ns	5	3199μs-3302μs	4087μs-7433μs	1(0.5-0.8)
WT-FillTableAlg_2							1257μs-1565μs	1(~0.2)
TPS-SortAlg							990μs~1029μs	1(~0.1)
TPS-FillTableAlg_25ns	4	80kHz	300ns	100ns	2	600μs~608μs	981μs~1022μs	1(~0.1)

需较少的CPU资源即可完成数据处理

数据处理性能测试



多平台部署升级，联调测试期间取数超过30TB

地点	目的	时间	结论
科大	TDAQ系统联调	22.11至今	验证了TDAQ功能与逻辑
三号厅	完整读出链验证平台	23.7至今	验证读出链
车库	WT小模型验证系统	22.12至今	SPMT逻辑功能验证
工厂	TAO模型实验	23.10至今	展开100+路联调与运行

未来研究计划

1. JUNO+TAO实验的运行调试

2025年迎来实验关键节点

1

调试任务

- JUNO和TAO实验全系统级联调
- 确保系统测试期间功能、性能以及稳定性
- 模拟简单到复杂极端工况，全方位保障系统正常工作
- 测试问题整理分析、解决优化

2

正式上线运行

- 硬件维护
 - 服务器管理 & 网络设备管控 & 存储设备运维
- 软件管理
 - 系统故障处理 & 中间件与数据库维护

如何实现高效的监测软硬件系统和异常报警，保障数据质量是一项**关键研究课题**

搭建一套**智能化数字化平台**实现多系统间智能监测和联动控制

未来研究计划

2. 人工智能在数据获取系统中的应用研究

研究内容 ➤ 基于人工智能的智能化数字化平台的搭建

智能监测系统 + 智能助手

研究内容1

提供实验运行的统一入口，实现对实验全方位运行控制

研究内容2

研究开发一套基于异构计算的实时处理软件，实现对大规模数据量的实时采集和集群调度处理

研究内容3

搭建一套独立于现有系统的在线实时智能监控系统，智能感知实验的运行状态，实现多系统间的智能监测

BESIII/LHAASO

JUNO/TAO

CEPC

立足于JUNO+TAO，面向CEPC

未来研究计划

2. 人工智能在数据获取系统中的应用研究

研究目的

人工检测存在滞后性，易遗漏

数据获取系统智能监测方案

监测类型一

基于时间的参数曲线的实时异常检测

- ✓ 基本思想：整合**先进的统计和机器学习方法**，对参数曲线进行预测，提前识别潜在的异常趋势
- ✓ **异常评分系统**，计算每个时间点的数据与预测模型的偏差生成异常评分，判断异常程度

监测类型二

直方图数据的实时异常检测

- ✓ 基本思想：基于机器学习的直方图异常检测
- ✓ 直方图数据转换为特征向量训练机器学习模型
- ✓ **算法的选取至关重要**，监督学习和无监督学习
- ✓ **自适应学习的能力**，模型持续学习新数据，调整自己内部参数，提高检测精度

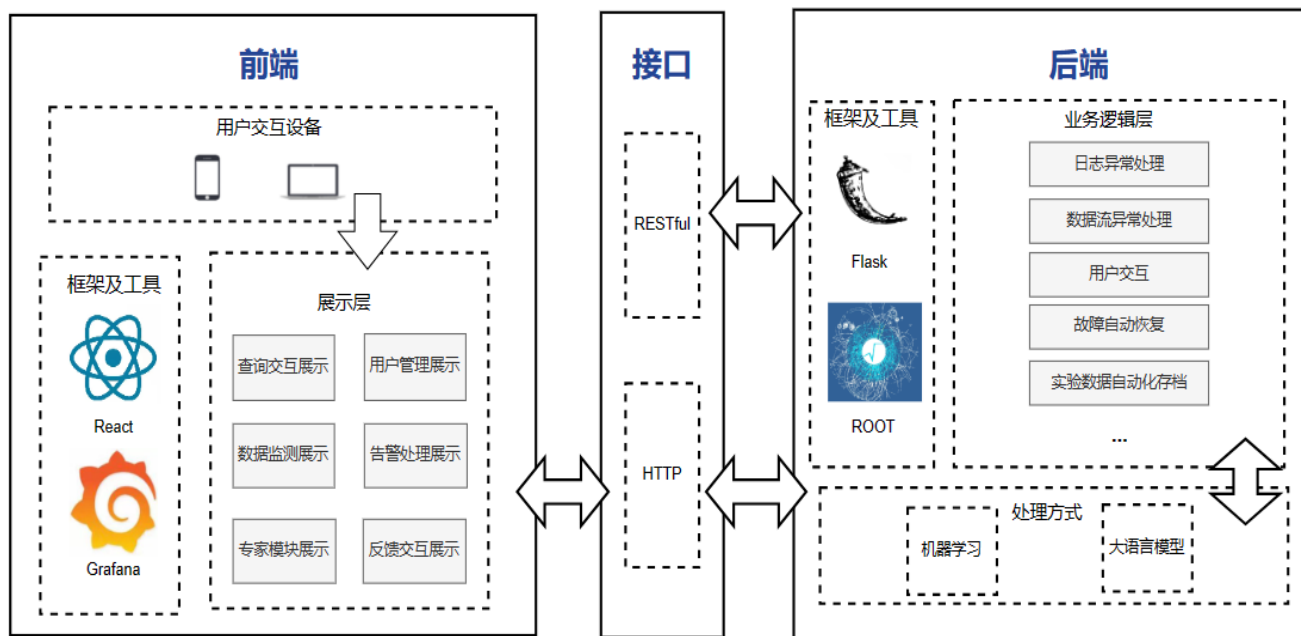
未来研究计划

2. 人工智能在数据获取系统中的应用研究

现有方案

传统的实验值班系统仅在前端提供基础的运行控制和实时监测功能

智能助手方案



后端的智能助手实现和前端的智能交互支持

- ✓ 强大的自然语言处理能力：实现对复杂问题的解释和问题解答
- ✓ 专门构建的知识库：查询操作步骤和根因分析
- ✓ 动态更新机制：定期更新实验数据和发现结果更新至知识库

大语言模型技术嵌入到值班系统，提供更直观、便捷的交互查询方式

未来研究计划

➤ 预期成果:

- 实验成果: JUNO与TAO实验成功完成调试, 关键性能指标符合设计指标要求, 进入正式稳定科学运行
- 学术产出: 发表3-5篇高影响力论文, 在国际会议上分享成果至少2次
- 技术沉淀: 开源一套基于AI的数据获取系统工具包

