

基于高可用集群的EPICS IOC冗余系统设计实现

徐国顺^{1,2}, 张祖超^{2,3*}, 田腾飞³, 黄子涵^{1,2}, 张杰^{1,3}

¹安徽理工大学计算机科学与工程学院, 安徽 淮南

²合肥综合性国家科学中心能源研究院(安徽省能源实验室), 安徽 合肥

³中国科学院合肥物质科学研究院等离子体物理研究所, 安徽 合肥

收稿日期: 2025年9月29日; 录用日期: 2025年10月26日; 发布日期: 2025年11月3日

摘要

在大型科研装置与工业控制系统中, EPICS架构因其分布式、高实时性的特性被广泛应用, 其中IOC作为关键组件, 其运行稳定性直接影响控制系统的安全性与数据完整性。针对IOC可能出现的宕机、网络中断等故障风险, 本文设计并实现了一套基于高可用集群的EPICS IOC冗余系统。该系统以Pacemaker与Corosync构建集群调度框架, 结合ETCD的分布式锁机制与DRBD的块级数据同步能力, 实现主备节点的自动切换与数据一致性保障。同时, 引入Autosave与Archiver Appliance工具对IOC状态及历史数据进行实时保存与归档。经多组故障模拟实验验证, 系统具备良好的容错能力与稳定性, IOC故障切换时长控制在30秒以内, PV数据持续性良好, 验证了系统在高可靠性控制应用中的可行性与稳定性。

关键词

EPICS, IOC冗余, 高可用集群, Pacemaker, 数据同步, 控制系统

Design and Implementation of an EPICS IOC Redundancy System Based on High-Availability Clusters

Guoshun Xu^{1,2}, Zuchao Zhang^{2,3*}, Tengfei Tian³, Zihan Huang^{1,2}, Jie Zhang^{1,3}

¹School of Computer Science and Engineering, Anhui University of Science & Technology, Huainan Anhui

²Institute of Energy, Hefei Comprehensive National Science Center (Anhui Energy Laboratory), Hefei Anhui

³Institute of Plasma Physics, Chinese Academy of Sciences, Hefei Anhui

Received: September 29, 2025; accepted: October 26, 2025; published: November 3, 2025

*通讯作者。

文章引用: 徐国顺, 张祖超, 田腾飞, 黄子涵, 张杰. 基于高可用集群的 EPICS IOC 冗余系统设计实现[J]. 计算机科学与应用, 2025, 15(11): 10-18. DOI: 10.12677/csa.2025.1511278

Abstract

EPICS (Experimental Physics and Industrial Control System) has been widely adopted in large-scale scientific facilities and industrial control systems due to its distributed architecture and real-time performance. As a core component, the IOC (Input/Output Controller) plays a critical role, and its operational stability directly impacts the system's safety and data integrity. To address potential failures such as controller crashes or network interruptions, this paper proposes and implements a high-availability EPICS IOC redundancy system. The system integrates Pacemaker and Corosync to form a cluster scheduling framework, uses ETCD for distributed locking, and leverages DRBD for block-level data synchronization, enabling automatic failover between the primary and backup nodes while ensuring data consistency. Additionally, Autosave and Archiver Appliance tools are employed to persist IOC states and archive historical data in real time. Experimental validation under multiple fault scenarios demonstrates the system's robust failover capability and stability. The switchover time between nodes can be controlled within 30 seconds, with no significant interruption in PV data, making the system suitable for high-reliability control applications.

Keywords

EPICS, IOC Redundancy, High Availability Cluster, Pacemaker, Data Synchronization, Control System

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

EPICS (Experimental Physics and Industrial Control System)作为分布式控制系统的行业标准,在粒子加速器、同步辐射光源及大型核聚变装置等领域发挥着核心作用。在其架构中,IOC (Input/Output Controller)作为连接底层硬件与上层应用的关键中间层,通过 Channel Access (CA)协议向上层暴露 PV (Process Variable),实现设备数据的采集与控制指令的下发。然而,在像全超导托卡马克(EAST)这样对运行时间要求极高的国家级大型实验装置中,IOC 往往需要长期高负载运行。一旦发生故障,不仅会导致控制数据丢失、设备失控,更可能直接中断宝贵的实验窗口。因此,在 EPICS 控制系统中建立一套高可靠性、状态完整的 IOC 冗余机制,以保障实验的连续性和系统的安全性,是至关重要的。

现有的 EPICS IOC 冗余技术方案主要包括传统硬件冗余方案和基于网络协议的冗余方案。其中,传统硬件冗余方案(如双机热备 Active/Standby 架构)依赖专用硬件实现故障切换与状态同步,具有高可靠性和配置简便的优点,但硬件成本较高、系统扩展性差,难以满足大型科研装置中多节点控制系统的灵活部署需求。基于网络协议的冗余方案(如利用 PROFINET 实现的冗余控制系统)能够在 I/O 层面实现数据冗余和链路保护,具有较好的实时性与兼容性,但在 IOC 层的状态同步和故障切换机制上仍不完善,一旦控制进程发生故障,系统无法快速恢复,存在控制中断与数据丢失风险。

针对上述不足,本文提出了一种基于高可用集群的 EPICS IOC 冗余系统设计。该方案以 Pacemaker 和 Corosync 作为集群调度与通信框架,结合 ETCD 的分布式锁机制和 DRBD 的块级数据同步技术,实现主备节点间的自动切换与数据一致性保障。同时,引入 EPICS 平台中的关键工具 Autosave 与 Archiver Appliance,分别负责 PV (Process Variable)实时值的保存与历史数据的归档存储,从而在控制层与数据层双重保障系统的稳定运行[1]。

2. 系统架构

冗余技术是一种通过引入额外设备或资源，以提升系统可靠性与运行安全性的常用手段[1]。在本设计中，IOC 冗余指的是：当主节点发生故障或 IOC 服务中断时，备用服务器能够自动接管控制任务并启动对应的 IOC 实例；一旦主节点恢复正常，备用节点则自动终止自身的 IOC 服务，恢复由主节点对外提供的控制功能。

本系统基于两台配置一致的 DELL PowerEdge R740 服务器与一台联想服务器构建，运行环境为 CentOS 8.5 操作系统，集成了 Pacemaker 2.1、Corosync 3.1、Autosave R5 和 Archiver Appliance 2.0 等关键软件组件，围绕 EPICS IOC 冗余方案展开系统性的设计与实现[1]。图 1 展示了该系统的整体架构：节点 1 与节点 2 分别承担主、备角色，均部署有功能相同的 IOC、Autosave 和 Archiver Appliance 服务模块，协同对外提供设备控制与数据归档能力；节点 3 部署 ETCD 服务器，用作分布式键值存储组件，负责协调集群中资源的运行状态与访问控制。

为实现主备节点之间的状态一致性，系统采用 DRBD (Distributed Replicated Block Device)进行块级数据实时同步，确保关键控制数据在切换过程中的完整性与一致性。systemd 作为现代 Linux 系统中的服务管理框架，负责服务单元的启动流程、生命周期管理及依赖控制。在本系统中，IOC、Autosave 与 Archiver Appliance 等关键服务均被注册为 systemd 管理单元，并与 Pacemaker 集群调度机制集成，实现服务的统一启动、运行状态监控与故障自动恢复[2]。

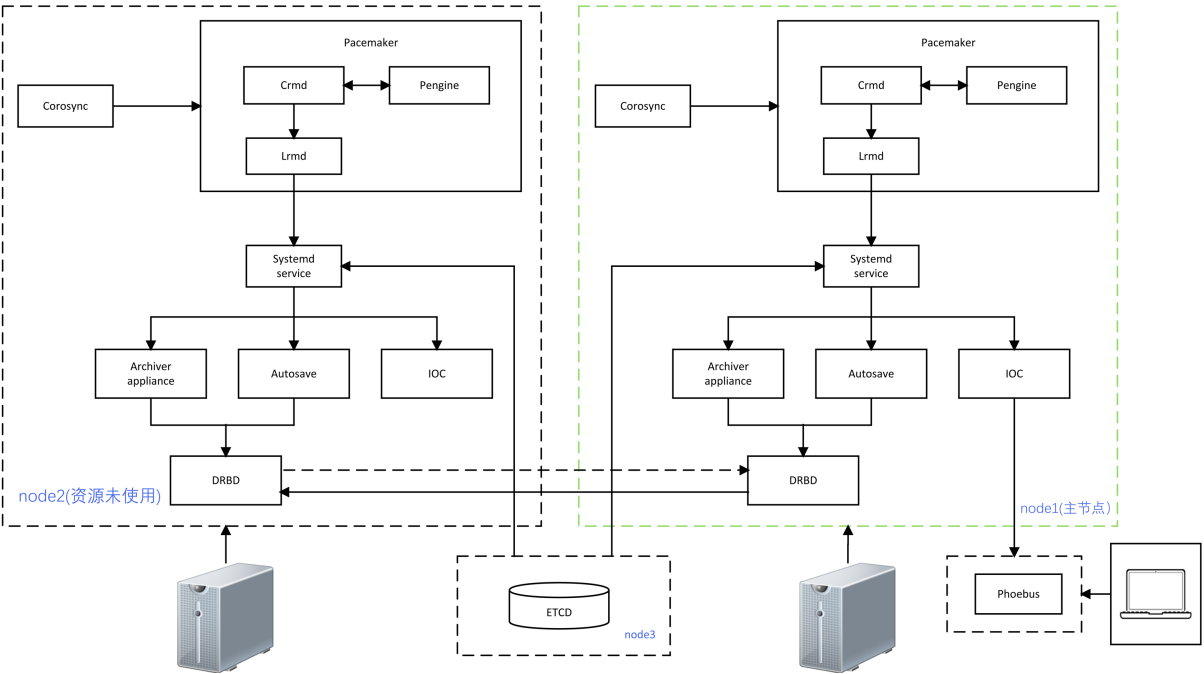


Figure 1. System overall architecture diagram
图 1. 系统整体架构图

3. 软件结构

为实现系统功能的模块化与高内聚低耦合设计，本控制系统从功能分工与服务边界出发，整体划分为三个协同运行的核心层级，分别承担集群调度、状态控制以及本地服务管理的职责[3]。图 2 展示了该系统的软件架构。

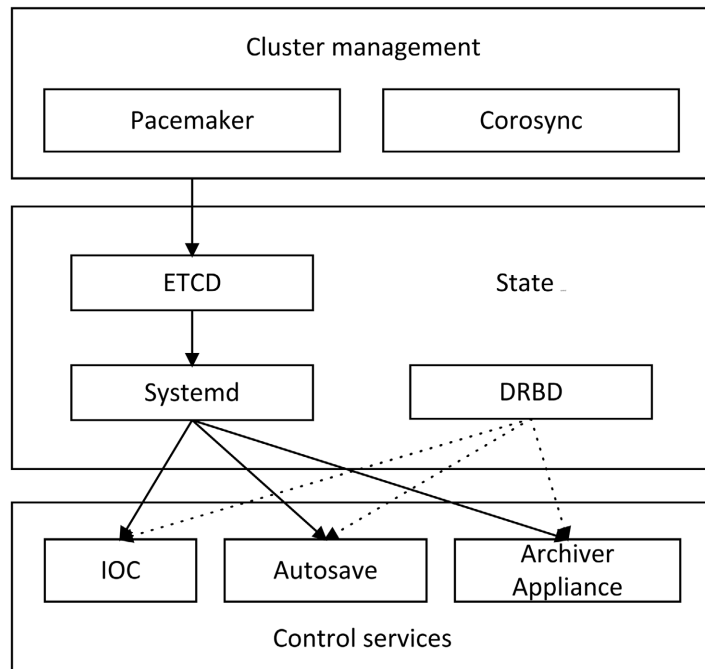


Figure 2. System software architecture diagram
图 2. 系统软件架构图

- 1) Cluster management: 通过 Pacemaker 和 Corosync 实现高可用感知与资源控制。
- 2) State control: ETCD 引入基于租约(Lease)的机制, 作为集群中资源控制权判断的分布式“锁服务”; DRBD 实现主备节点之间的块级数据实时同步, 保障关键 IOC 状态文件与归档数据在切换过程中的一致性与完整性[4] [5]。
- 3) Control services: systemd 负责统一接管 IOC、Autosave 和 Archiver Appliance 等服务的启动流程与运行状态反馈; 控制服务(IOC、Archiver Appliance、Autosave)被注册为 systemd 管理的资源单元, 通过配套脚本实现服务的自动化启动、状态监控以及故障联动响应, 确保运行流程高效、可靠[6]。

3.1. Pacemaker 与 Corosync 架构

Pacemaker 与 Corosync 最初源自传统高可用集群软件 Heartbeat, 后逐步演化为两个独立的模块。Pacemaker 是集群的资源调度核心, 负责主节点选举、服务状态监测以及资源接管策略的执行; 而 Corosync 则作为底层通信总线, 提供节点间的心跳检测、投票仲裁与状态同步功能, 构建起集群内部可靠通信的基础框架[5] [6]。图 3 展示了 Pacemaker 与 Corosync 的核心组件架构, 在集群运行过程中, 策略引擎(PEngine)会解析集群信息库(Cib)中记录的资源配置与节点状态, 生成资源调度图, 从而确定资源由哪个节点承担运行职责, 即主节点的选定。随后调度图由集群资源管理器(Crmd)分发至各节点, 用于协调资源状态维护与调度执行。当某节点被指定为主控节点后, 其本地资源代理(Lrmd)将接收资源启动指令, 并调用系统注册的 systemd 服务单元。

3.2. ETCD

Pacemaker 与 Corosync 构建了完整的主节点判定、服务启动、健康检测及故障切换恢复机制。在此基础上, 系统进一步引入了 ETCD 组件, 以增强资源控制的可靠性与一致性。ETCD 是由 CoreOS 开发的开源分布式键值存储系统, 基于 Raft 共识算法实现强一致性[7] [8], 并通过租约(Lease)机制为分布式锁

控制提供了可靠支撑。

在本系统中，每个受控资源在 ETCD 中对应一个唯一键。该机制特别适用于两节点集群结构，有效避免因网络异常引发的“双主”（脑裂）问题[1]。具体而言，节点在被 Pacemaker 调度后不会立即启动服务，而是首先向 ETCD 发起租约申请：若成功写入对应资源键，视为已获取主控权限，方可进入服务激活流程；若租约申请失败，即使获得调度资格，也不会启动服务，从机制上防止“双主”冲突的发生。

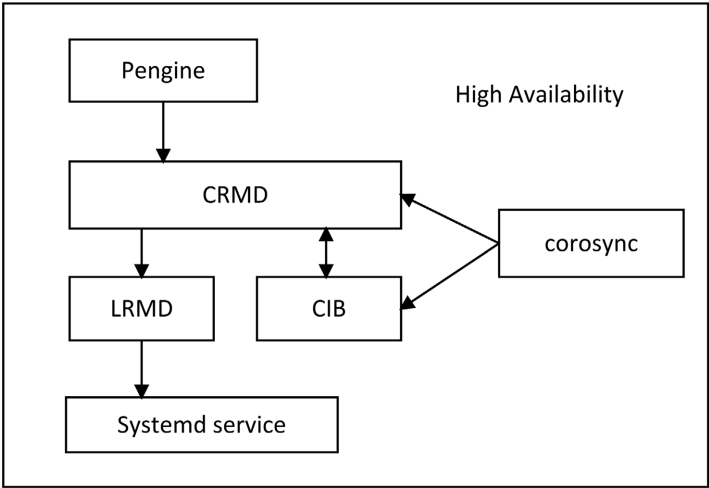


Figure 3. HA cluster architecture diagram
图 3. HA 集群架构图

3.3. 数据同步

系统的主备节点上分别部署了 Autosave 和 Archiver Appliance 软件，其中主节点和备节点上的 Autosave 和 Archiver Appliance 软件不会同时运行，仅主节点上的软件处于活动状态。Autosave 在 IOC 启动或重启时，从最近保存的文件中恢复各个 PV 的值，确保设备控制逻辑的连续性[9]。Archiver Appliance 将历史数据根据时间跨度划分至三个独立目录：ST (Short-Term)用于短期缓存，MT (Medium-Term)负责中期数据存储，LT (Long-Term)用于长期归档，支持数据的分层访问与集中管理[10]。

为确保主备 IOC 数据的一致性，部署时，在每台服务器上预留专用分区作为 DRBD 设备，并配置 DRBD 主备节点角色。集群在选举主备节点的同时选举主备 DRBD。Autosave 与 Archiver Appliance 的数据目录统一部署在 DRBD 设备中。运行期间只有主节点能将 Autosave 与 Archiver Appliance 的数据写入 DRBD 设备，DRBD 工具自动将主节点上 DRBD 设备数据同步给备节点的 DRBD 设备。

3.4. 系统运行流程与状态模型

本系统将系统运行流程抽象为一个有限状态机(Finite State Machine, FSM)。图 4 详细展示了这一 FSM 的完整状态转换图。核心状态包括：

- ① Standby：节点已加入集群但尚未取得主控资格，处于待命状态。
- ② Primary (idle)：被 PEngine 选为主节点，但尚未获得 ETCD 租约，资源未激活。
- ③ Primary (active)：节点被 Pacemaker 的 PEngine 模块选为主节点，但尚未成功获取 ETCD 租约，资源尚未激活。
- ④ Failed：因服务启动异常、ETCD 不可达、租约失效等问题，节点进入故障状态。
- ⑤ Terminated：节点因断网、宕机等严重故障彻底失去服务运行资格。

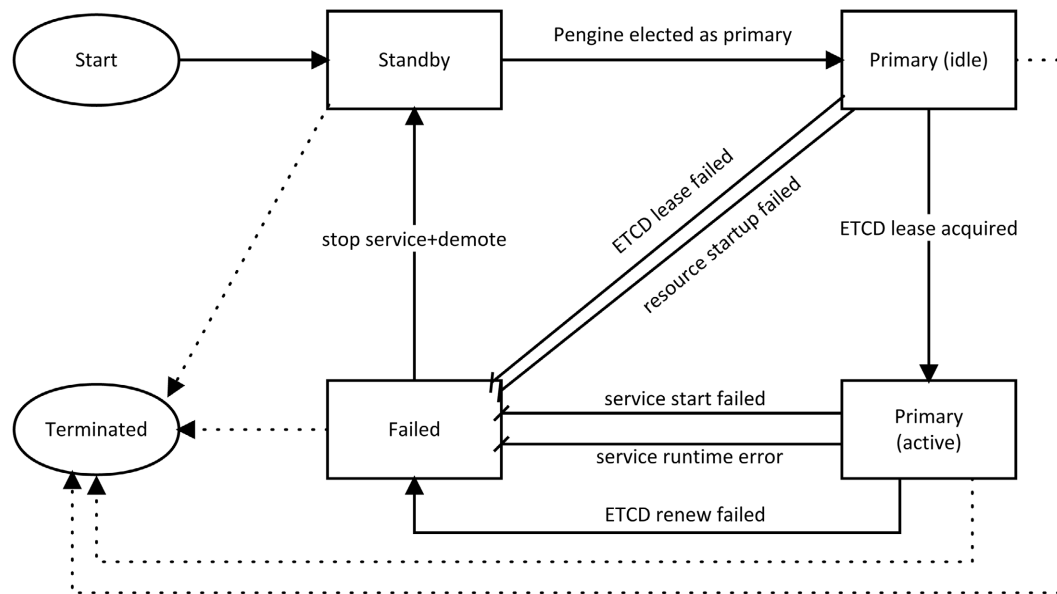


Figure 4. System operational status diagram

图 4. 系统运行状态图

系统运行逻辑可分为三个阶段：主控节点选举、主控资源激活、故障检测与切换。以下是典型的一次主控激活过程：

系统初始启动后，各节点处于 Standby 状态，等待 Pacemaker 的选举决策。PEngine 解析集群状态后，选定某节点为主控节点，该节点随即进入 Primary (idle)状态，表示获得调度资格但尚未真正启动资源。

在 Primary (idle)状态下，节点通过 systemd 启动服务单元，进入资源激活流程。控制脚本首先尝试从 ETCD 获取租约，若租约申请成功，状态转入 Primary (active)，表示该节点具备主控权限并正式启动 IOC 及相关服务。系统运行过程中，节点定期向 ETCD 续约，若发生租约续约失败、服务运行异常或底层故障(如 DRBD 挂载失败)，状态将转入 Failed。Pacemaker 检测到服务异常后，自动发起降级操作(stop service + demote)，回收资源控制权，并重新进入选主流程，调度备用节点接管，完成资源迁移与恢复。若节点遭遇严重系统故障，将直接进入 Terminated 状态，退出资源参与。

4. 测试与结果

基于第二章的软件架构，本文构建了 EPICS IOC 冗余控制系统的测试环境，旨在验证系统在主控节点失效场景下的容错能力、主备切换效率以及 PV 数据一致性保障能力。测试模拟主节点故障，通过网络断开方式触发切换机制，并通过工具持续观察 PV 状态变化和归档数据同步情况。

4.1. 控制权接管验证

测试开始前，使用 EPICS 工具 camonitor 对目标 PV 变量 reduncance:sync:test1 进行持续监听。图 5 展示了监听过程中关键日志变化。初始阶段，camonitor 成功连接至主节点 xgs 提供的 PV，并持续接收数据。于 11:04:55 模拟主节点故障(断开网络连接)，客户端立即检测到连接中断，并进入重连状态。最终在 11:05:21 成功连接至备节点 epicstest，其已接管原 IOC 服务并继续提供相同 PV。PV 值维持在 6.0，与断网前一致，表明服务状态完全恢复，主备切换顺利完成。

本轮测试从故障触发至连接恢复总耗时约 25 秒，PV 数据无明显中断，验证了系统具备良好的切换稳定性和容错能力。


```
[LAUNCH] 启动监控子进程...
[INFO] 成功连接 PV 'reduance:sync:test1', 当前 IOC: xgs:5064
=====
[2025-07-25 11:04:42] Value: 3.0 | IOC Host: xgs:5064 | TS: 1753412682.683014
[2025-07-25 11:04:43] Value: 4.0 | IOC Host: xgs:5064 | TS: 1753412683.68308
[2025-07-25 11:04:44] Value: 5.0 | IOC Host: xgs:5064 | TS: 1753412684.683175
[2025-07-25 11:04:45] Value: 6.0 | IOC Host: xgs:5064 | TS: 1753412685.68306
[2025-07-25 11:04:46] Value: 7.0 | IOC Host: xgs:5064 | TS: 1753412686.683346
[2025-07-25 11:04:47] Value: 8.0 | IOC Host: xgs:5064 | TS: 1753412687.68336
[2025-07-25 11:04:48] Value: 9.0 | IOC Host: xgs:5064 | TS: 1753412688.682865
[2025-07-25 11:04:49] Value: 0.0 | IOC Host: xgs:5064 | TS: 1753412689.682802
[2025-07-25 11:04:50] Value: 1.0 | IOC Host: xgs:5064 | TS: 1753412690.683571
[2025-07-25 11:04:51] Value: 2.0 | IOC Host: xgs:5064 | TS: 1753412691.684196
[2025-07-25 11:04:52] Value: 3.0 | IOC Host: xgs:5064 | TS: 1753412692.683089
[2025-07-25 11:04:53] Value: 4.0 | IOC Host: xgs:5064 | TS: 1753412693.683158
[2025-07-25 11:04:54] Value: 5.0 | IOC Host: xgs:5064 | TS: 1753412694.684664
[2025-07-25 11:04:55] Value: 6.0 | IOC Host: xgs:5064 | TS: 1753412695.68305
CA.Client.Exception.....
Warning: "Virtual circuit disconnect"
Context: "xgs:5064"
Source File: ../cac.cpp line 1237
Current Time: Fri Jul 25 2025 11:04:56.107615855
.....
[WARN] 与 PV 'reduance:sync:test1' 的连接断开, 正在重试...
[INFO] 子进程已退出, 3 秒后重启...
[LAUNCH] 启动监控子进程...
[ERROR] 无法连接到 PV: reduance:sync:test1
[INFO] 子进程已退出, 3 秒后重启...
[LAUNCH] 启动监控子进程...
[ERROR] 无法连接到 PV: reduance:sync:test1
[INFO] 子进程已退出, 3 秒后重启...
[LAUNCH] 启动监控子进程...
[INFO] 成功连接 PV 'reduance:sync:test1', 当前 IOC: epicstest:5064
=====
[2025-07-25 11:05:21] Value: 6.0 | IOC Host: epicstest:5064 | TS: 1753412775.299316
[2025-07-25 11:05:22] Value: 7.0 | IOC Host: epicstest:5064 | TS: 1753412776.299204
[2025-07-25 11:05:23] Value: 8.0 | IOC Host: epicstest:5064 | TS: 1753412777.299254
[2025-07-25 11:05:24] Value: 9.0 | IOC Host: epicstest:5064 | TS: 1753412778.299214
[2025-07-25 11:05:25] Value: 0.0 | IOC Host: epicstest:5064 | TS: 1753412779.29921
[2025-07-25 11:05:26] Value: 1.0 | IOC Host: epicstest:5064 | TS: 1753412780.299207
[2025-07-25 11:05:27] Value: 2.0 | IOC Host: epicstest:5064 | TS: 1753412781.299241
```

Figure 5. IOC switchover test interface
图 5. IOC 切换测试界面

4.2. 主备状态可视化展示

系统通过 Phoebus 工具在人机界面中实时呈现主备角色状态与服务活动情况。图 6 显示了主备 IOC 的标识信号。其中，主节点与备节点分别配置状态标识 PV：“主 IOC”与“备 IOC”，便于快速区分角色。此外设置的心跳信号 PV 可反映服务活动状态：主节点信号呈周期性跳变，表示服务活跃；而备节点心跳信号保持恒定，体现待命状态。通过界面反馈，操作者可直观判断系统当前运行状态及角色归属。

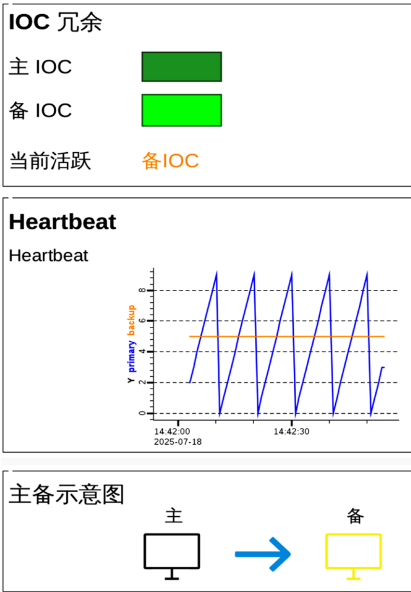


Figure 6. Phoebus interface
图 6. Phoebus 界面

4.3. 归档数据同步测试

为进一步验证 Archiver Appliance 的数据同步能力，测试中使用其客户端工具查看 PV test 的历史归档曲线，图 7 展示了切换后在备节点上的历史数据查询结果。系统配置中，主备节点各自持有名为 test 的 PV，其中主节点信号呈正弦函数变化，备节点信号为{-1, 1}的阶跃波。测试结果表明，备节点 Archiver Appliance 成功检索到了主节点生成的 PV 历史曲线，说明归档数据已顺利完成同步。

图 7 中存在一个时间段为 21:13:17 至 21:13:45 的数据断点，反映了归档服务在主备切换过程中的恢复耗时约 30 分钟。尽管存在短暂中断，整体数据一致性和历史记录完整性仍保持在可接受范围内，验证了系统的归档可靠性。

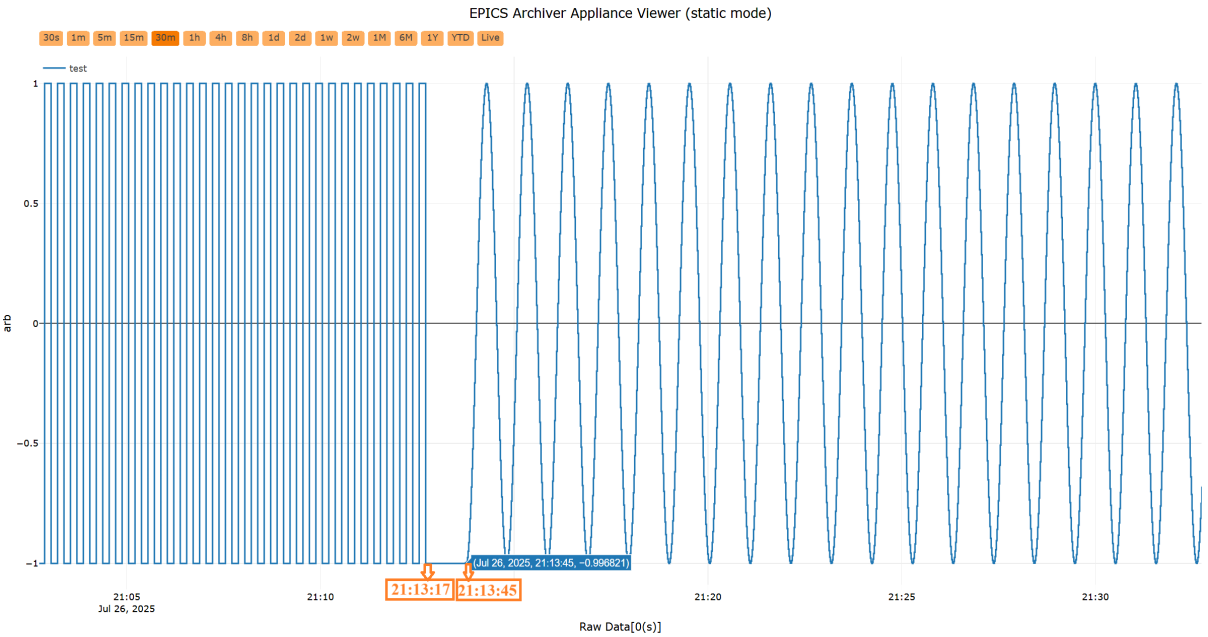


Figure 7. Archiver Appliance testing interface
图 7. Archiver Appliance 测试界面

5. 结论

本文围绕 EPICS 控制系统中对 IOC 稳定性与高可用性的核心需求，设计并实现了一种基于 Linux 高可用集群架构的 IOC 冗余系统。该系统集成 Pacemaker 与 Corosync 集群调度框架、ETCD 分布式锁机制、DRBD 块级数据同步服务，以及 Autosave 和 Archiver Appliance 数据持久化工具，实现了控制资源的自动接管与控制状态、历史数据的一致性保障。

通过多轮故障模拟测试，验证了系统在控制权切换、PV 值保持和数据归档同步等方面的稳定性与有效性。结果表明，在网络中断或进程异常等典型故障场景下，系统能够在 30 s 内完成无人工干预的主备切换，PV 数据连续性良好，归档信息无明显中断，具备满足高可靠性控制系统实时性与连续性要求的能力。

基金项目

合肥综合性国家科学中心能源研究院(安徽省能源实验室)项目(Grant No.24KZS304)；国家重点研发计划国家磁约束核聚变能发展研究专项(Grant No.2022YFE03010002 和 Grant No.2024YFE03020003)。

参考文献

- [1] 尹聪聪, 韩利峰, 李勇平, 等. EPICS IOC 冗余技术研究[J]. 自动化仪表, 2013, 34(12): 73-75+79.
- [2] 余强, 周芷伟, 袁恺, 等. CRAFT 1 kW@4.5 K 氦制冷机控制系统开发与应用[J]. 低温与超导, 2025, 53(6): 1-8.
- [3] 尹聪聪. 钍基熔盐堆 EPICS 控制系统可靠性技术应用研究[D]: [博士学位论文]. 上海: 中国科学院研究生院(上海应用物理研究所), 2015.
- [4] 黄子灏. PROFINET 在加速器控制系统中的应用研究[D]: [博士学位论文]. 合肥: 中国科学技术大学, 2019.
- [5] 李宇鲲, 杜垚垚, 叶强, 等. 基于高可用集群的服务化 EPICS 与数据处理方式[J]. 强激光与粒子束, 2025, 37(1): 81-87.
- [6] Li, X., Li, Y., Shao, R., *et al.* (2026) Multigroup Cross-Section Library Generation for Deterministic Electron-Transport Calculation. *Annals of Nuclear Energy*, **227**, 111937-111937. <https://doi.org/10.1016/j.anucene.2025.111937>
- [7] 吴汉楠, 岳敏, 马涛, 等. 荧光靶历史图像数据存储系统设计与实现[J]. 强激光与粒子束, 2024, 36(7): 73-79.
- [8] Bielewicz, M., Mazerewicz, P., Szewiński, J., *et al.* (2025) The New Precise Positioning System of the Heavy Hadron Calorimeter FPSD in the NA61/SHINE Experiment Based on the Siemens 1200 Controller Connected with the EPICS Software. *Electronics*, **14**, 3961-3961.
- [9] 薛康佳, 张玉亮, 王林, 等. CSNS 加速器服务器监控系统设计[J]. 核电子学与探测技术, 2023, 43(2): 404-408.
- [10] 雷蕾, 韩利峰, 徐海霞, 等. EPICS 环境下的软件规范管理[J]. 核技术, 2015, 38(6): 74-79.