

医疗大数据双活系统集成技术应用比较研究^{*}

丁涛^{1,2} 刘峰³ 杨赛¹

(¹ 汉中市中心医院 汉中 723000 ² 西安交通大学 西安 710049 ³ 宁夏医科大学总医院 银川 750004)

[摘要] **目的/意义** 探讨系统集成双活技术在医疗大数据安全保护中的应用。**方法/过程** 采用比较研究法,从建设复杂度、保护内容、支持模式、应用效果等方面分析 3 类常用双活技术的优缺点,针对不同安全保护需求和场景,提出技术应用解决方案。**结果/结论** 不同层面系统集成双活技术实现的数据保护效果有差异,核心业务系统宜采用数据库端和基于底层存储的“双写式”双活系统集成融合;一般业务系统宜采用“双写式”双活系统集成架构与虚拟机技术融合;特殊业务应用宜采用存储底层硬件“单写式”双活系统集成,能够以较高性价比实现医疗大数据的安全保护,提升数据安全。

[关键词] 大数据; 数据存储; 业务连续性保护; 数据保护; 双活及多活系统

[中图分类号] R-058 **[文献标识码]** A **[DOI]** 10.3969/j.issn.1673-6036.2023.07.015

A Comparative Study on Application of Integration Technology of Medical Big Data Dual Active System

DING Tao^{1,2}, LIU Feng³, YANG Sai¹

¹Hanzhong Central Hospital, Hanzhong 723000, China; ²Xi'an Jiaotong University, Xi'an 710049, China; ³General Hospital of Ningxia Medical University, Yinchuan 750004, China

[Abstract] **Purpose/Significance** To explore the application of system integration dual active technology in medical big data security protection. **Method/Process** The comparative study method is adopted to analyze the advantages and disadvantages of three commonly used dual active technologies from the aspects of construction complexity, protection content, support mode, application effect, etc., and it proposes technical application solutions for different security protection requirements and scenarios. **Result/Conclusion** The data protection effects achieved by the dual active technology of system integration at different levels are different. The core business system should adopt the integration of the database side and the dual write dual active system based on the underlying storage; general business systems should adopt the dual write dual active system integration architecture and virtual machine technology integration; special business applications should adopt the single write dual active system integration of the underlying storage hardware, which can achieve the security protection of medical big data with a high cost performance ratio and improve data security.

[Keywords] big data; data storage; business continuity protection; data protection; dual active and multi active systems

[修回日期] 2023-01-03

[作者简介] 丁涛, 高级工程师, 发表论文 18 篇。

[基金项目] 陕西省卫生健康科研基金项目(项目编号: 2021D018); 陕西省汉中市中心医院青年启动基金项目(项目编号: YK2108)。

1 引言

随着医院信息化建设的逐步深入,信息技术广泛应用于医疗机构日常诊疗、管理、服务等活动中,成为医院运营的重要基础设施^[1],其安全、可靠、稳定运行对各项工作顺利开展至关重要。2018 年国家卫生健康委员会发布《全国医院信息化建设标准与规范(试行)》^[2],明确指出医院信息化建设要涵盖各应用及数据库服务器、存储磁盘阵列、集群软件和应用容灾软件等组件,具备数据容灾能力,支持多种技术本地化高可用,对医疗大数据的安全性、容灾能力建设的需求更加迫切^[3]。部分医院为提高系统和数据抗风险能力,运用系统集成技术实现大数据安全保护,主要方式包括基于底层存储硬件的数据克隆、复制,基于应用或数据库的群集,基于操作系统的热备,以及其他不同层级双活技术、虚拟化技术等。魏智等^[4]通过对医院数据安全需求分析,提出一种基于数据库和存储的双活数据中心,最大限度实现业务系统的高可用和数据容灾;曹凯^[5]通过软件定义网络(software-defined networking, SDN)、超融合等新技术应用,构建跨数据中心的超融合虚拟化资源池,减少系统单点故障、性能瓶颈;上官斌^[6]研究显示,越来越多机构开始逐步推进系统双活建设,其中最重要、最困难的是如何实现大型数据库双活,包括仲裁一致性、性能同步、异常切换决策等。

建立数据中心双活容灾保护机制(以下简称双活)可以有效促进大数据相关业务的高可用,确保数据一致性和系统安全性,已被广泛应用于医疗大数据安全保护。但是,不同双活建设层级达到的数据保护预期效果差别较大,如何选择合适的双活技术,以较高性价比实现大数据安全保护,具有一定研究意义。本文通过对常用医疗大数据双活保护技术分析,从实施复杂度和风险、建设与运维成本、软硬件兼容性以及对不同结构数据的保护能力,资源消耗、性能、技术开放程度等方面,总结不同维度双活技术的优缺点,结合医院数据保护场景和功

能需要,讨论技术方案的难点与问题,为医院大数据中心的可靠运行、应急容灾和数据保护提供思路。

2 医疗大数据安全需求分析

医疗大数据包括不同时期、厂商、业务、信息系统应用所积累的各类患者病历、文书,检验、检查结果等电子化资料,以及医院管理、运营、公共卫生服务中形成的各种与医疗相关的数据信息^[7]。根据“智慧医院”建设程度,信息系统日益增加使数据累积和应用环境趋于复杂,机构医疗大数据呈现出海量、业务及时化、存储多样化、价值高端化的特点^[8],其安全需求主要包括:一是医疗大数据的价值决定其安全基础是物理数据安全性,避免数据丢失、损毁以及其他因物理数据安全造成的公共事件;二是从应用角度,关注系统和数据容量增长衍生的各种问题,对数据应用层服务能力要求更高;三是关注数据连续性和系统运行鲁棒性,能够持续为患者和医疗服务提供数据支撑能力;四是从技术角度更加关注各种不同物理硬件、应用软件的高性能,在高并发下为不同业务需求提供安全可靠的服务;五是关注复杂情况下,不同类型(包括各种结构化、非结构化、半结构化)数据管理,以及数据集逻辑与隐私保护^[9]。

3 医疗大数据双活系统集成技术与应用

3.1 双活系统集成技术应用意义

医院业务系统应用具有较强行业特点,医疗大数据保护一般不为每个系统或独立应用进行特定数据安全保护体系建设,常用做法是从医疗数据的采集、存储、访问、应用等环节建立完整安全防护体系^[10]。双活系统集成技术被广泛应用,特别是基于数据中心的超融合应用,具有低成本、低风险、快速交付等特点,有助于医院减少资源浪费,既达到保护效果又能在共享和应用过程中实现数据管理的安全可靠和规范有序。

3.2 双活系统集成技术概况

双活系统集成就是由两套或两组以上存储、主机同时为业务系统提供数据读写及存储服务，数据同时以双副本或更多副本形式分别存在两个不同物理空间。同时，所有数据均在线，当任意底层物理部件出现损坏时业务数据不会丢失，甚至业务平台不会停止，见图 1。

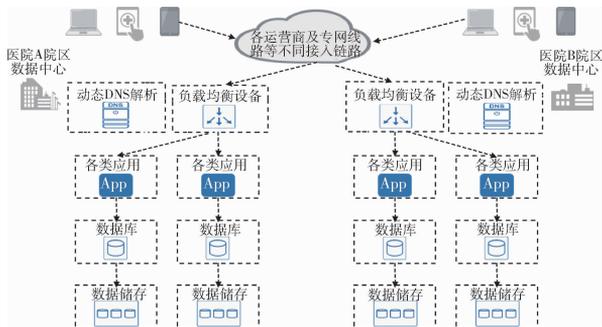


图 1 双活数据中心

双活系统集成技术并没有标准的定义或者行业规范，一般根据不同层级数据和业务保护需要，形成涵盖存储、数据库、文件、网络、业务等不同层面的系统集成部署技术，本文主要讨论业务数据系统集成双活技术。

3.3 医院常用双活系统集成技术

3.3.1 双活系统集成技术分类 按照数据副本动作发生的位置和处理方式，双活系统集成技术可分为 5 种：基于数据库的双活系统集成；基于主机（文件系统层）的双活系统集成；基于存储硬件的“单写式”双活系统集成；基于存储硬件的“双写式”双活系统集成；重构业务应用，通过应用实现双活系统集成。其中，第 5 种是根据业务特点、访问逻辑、访问模式等进行应用软件重构，再添加一系列软件组件，封装部分业务模块，甚至改变业务流程，进而实现双活。由于医院业务特殊性，除小部分技术实力强、业务模式简单的新兴互联网医疗平台可以实现外，其应用范围相对较小。因此本文以讨论前 4 种技术为主。

3.3.2 数据库端双活集成 数据库端双活集成技术一般基于数据库的分布式服务构建。在同一应用

环境下，除保持数据一致性外，将数据库服务分散至两台以上的服务器。其典型应用形式分为两类，一是以甲骨文公司为代表的 Oracle RAC 结合自动存储管理（automatic storage management, ASM）解决方案。ASM 支持条带化和磁盘镜像，且比传统的独立磁盘冗余阵列（redundant array of independent disks, RAID）更灵活，可以为每个数据文件指定冗余级别。当 ASM 为普通文件分配一个扩展数据块时，会分配另一个辅助副本，并将辅助副本存储在主副本以外的其他磁盘组块中，在发生磁盘故障时不会导致数据丢失。二是基于数据库日志的数据库文件副本同步复制技术。在数据库运行过程中，所有客户端请求都会写入数据库日志，再写入相应数据库文件里，并通过日志详细记录数据写入先后顺序，便于解决数据恢复时的一致性和完整性问题。部分数据库服务商研发出基于日志的副本复制技术，在写日志进程中将缓存里的日志依托通信传输给副本库实例，副本库根据接收的日志执行数据更新操作，确保副本库与主库的数据一致；该技术能为数据库提供双文件支持，以同步或异步形式更新，支持手动或自动故障转移，并可将副本配置成只读模式，进行其他数据操作，从而减轻主库压力。

3.3.3 主机端双活集成 在主机端层面，可以基于逻辑卷镜像方式提供双活系统集成服务，常见技术形式有 Unix、Linux 系统的逻辑卷管理（logical volume manager, LVM）模式、Windows 系统的动态卷镜像功能等。依托操作系统原生的输入/输出（input/output, I/O）组件，基于不同站点的各种物理存储卷，利用镜像方式“组合”成一个在线逻辑卷，再将逻辑卷提供给操作系统和业务系统存取数据。来自不同存储的两个独立物理存储卷，通过存储区域网络（storage area network, SAN）提供给计算服务器操作系统，这两个卷可以本地化部署，也可以通过 SAN 跨数据中心形成远程双活数据。通过对 LVM 物理分区单元写入进行实时双向复制，当本地卷数据及远端卷数据同时写完时，才能算一个完整的数据写入；当其中一个存储卷写

入超时或写入失败时，该存储卷会被标为故障或离线状态，等写入恢复后，重新手动同步，以确保镜像副本与原生副本数据一致。在创建逻辑卷时，LVM 已经将逻辑卷与两个或多个不同底层物理存储卷完成映射，针对任何数据文件的读写都需要由文件系统和 LVM 完成与底层物理数据块交互，在磁盘底层出现坏块或某个底层存储卷彻底损坏时，LVM 无须对底层数据块指针进行动态转移，以保证数据不丢失、业务不停止。

3.3.4 基于存储底层硬件的双活集成 基于存储底层硬件实现的双活系统集成主要有两种方式：一种是“单写式”；另一种是“双写式”。“单写

式”双活系统集成是指针对底层的数据卷，所有数据“写”操作均由一方完成，并且进行数据单向同步。基于存储底层硬件的“单写式”双活系统集成，见图 2。存储集群 A 为源端，存储集群 B 为从端，存储集群内部的控制器 A1 和 A2，B1 和 B2 分别构成高可用（high availability, HA）环境，在搭建集成环境时两个存储集群建立从属关系，源端任何写入操作都交由本地卷 SVM_A1 完成，而对端控制器接收到前端服务器的写入请求时，会将该请求转发至源端 SVM_A1，由源端 SVM_A1 完成写入操作，同时给目标端 SVM_A1 - mc 复制一个写入操作，直到两边 SVM 同时写入完成后，数据和副本才完成写入。

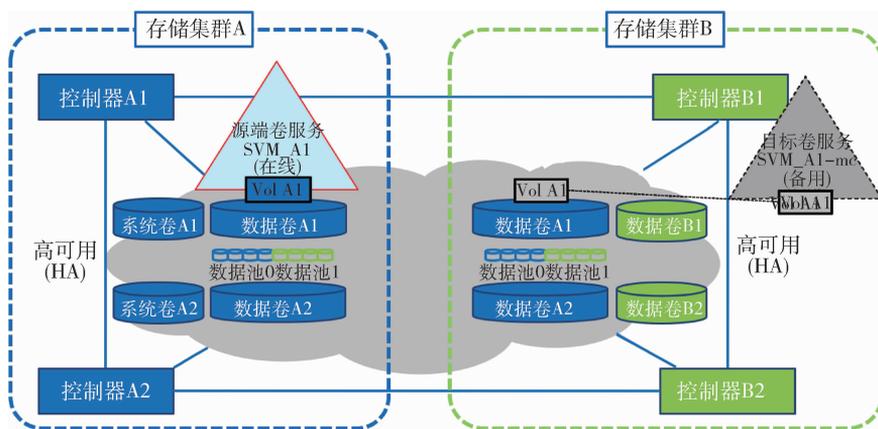


图 2 基于存储底层硬件的“单写式”双活系统集成

因此，针对数据卷 A1，无论是来自源端还是对端服务器写入请求，均会被转发给源端 SVM_A1 完成“写”操作，再按照从源端 SVM_A1 到目标端 SVM_A1 - mc 单向方式完成数据复制“写”操作。当源端出现故障或灾难时，业务系统会有一个属主切换过程，类似故障转移，在此期间业务会发生中断或性能急剧降低。

存储“双写式”双活系统集成技术与“单写式”最大的不同之处在于针对底层数据卷，对端和源端在接收到数据写入请求后，会自发完成“写”操作。这种架构的对端及源端数据卷均是在线、可读写状态，因此在底层任意存储硬件出现问题时不需要故障切换动作，确保业务连续性

和数据安全性；另外，源端和对端双方均可同时提供读写操作，其整体性能优于“单写式”架构。

4 技术比较与讨论

不同层面双活技术实现的数据保护效果存在差异，见表 1。医院各种业务系统众多且集成环境复杂，很难用一种技术实现所有层面的数据保护，因此，结合不同需求和应用场景以及业务响应要求，分级分类选择合适的双活数据保护技术，才能以较高性价比实现医疗大数据有效保护。

表 1 不同层面双活系统集成技术对比

对比项目	数据库端	主机端	存储端	
			“单写式”	“双写式”
实施复杂度及风险性	较小	小	适中	较小
建设及运维成本	适中	低	适中	略高
与现有环境兼容性	较差	好	一般	较好
非结构化数据保护	不支持	支持	支持	支持
结构化数据保护	支持	不支持	支持	支持
虚拟机保护	不支持	支持	支持	支持
双数据中心模式下, 应用对数据双读操作	支持	支持	支持	支持
双数据中心模式下, 应用对数据双写操作	部分支持	不支持	不支持	支持
故障切换及恢复时间	短	3~5 分钟	分钟级	无 IO 停止
主机端资源消耗	小	小	无	无
存储端资源消耗	无	无	一般	一般
性能影响	无	一般	较小	较小
厂商/技术锁定	一般	无锁定	较深锁定	较深锁定

4.1 核心业务系统及数据库保护

按照国家卫生健康委员会《关于进一步推进以电子病历为核心的医疗机构信息化建设的通知》要求^[11], 三级医院到 2020 年电子病历系统功能应用水平分级评价要达到 4 级以上。其中, 医院信息系统、电子病历系统等作为核心应用要实现业务及数据库保护。从业务角度看, 核心系统的容灾恢复点目标和恢复时间目标接近 0 才能达到高可用^[12], 满足安全需求; 从数据角度看, 在确保数据一致性的前提下, 应该尽可能实现数据多份冗余, 因此, 采用数据库端和基于底层存储的“双写式”双活系统集成融合, 可以最大限度地实现整体保护效果。或者结合不同主机的故障转移群集能力, 实现计算层面高可用, 将故障转移群集与存储底层硬件“双写式”双活结合, 进一步提升核心业务生产系统的稳定性以及高可用。

4.2 一般业务系统及数据库保护

一般业务系统的响应及时性要求不高, 恢复时间点目标要比核心业务系统大, 特别是部分医院建设资金投入有限, 对除核心业务以外的绝大多数系

统采用超融合、虚拟化等方式进行部署, 通过建立虚拟资源池节省硬件开销, 可以降低运维成本与难度。针对这部分业务及数据库, 可以将数据空间采用“双写式”双活系统集成架构进行构建, 而上层业务则依托虚拟机平台自身的快照技术, 或者其他多主机共享群集技术, 实现业务高可用与数据冗余灾备, 例如虚拟机快照技术、RoseHA、Windows Cluster、赛门铁克 VCS 等, 能够提高生产系统可靠性, 以较低成本实现较好效果。

4.3 特殊业务应用及数据保护

根据使用场景, 医院还有一些特殊信息系统应用, 特别是多影像、多文件存储及管理, 其业务模式、性能需求、数据特征以及数据量和其他系统差异较大。一是数据量巨大, 单日新增数据量超过 10 GB, 随着时间推移, 积累的海量数据形成巨大保存压力; 二是数据形式以影像、图片等文件为主, 部分系统产生的文件数据还具有小而多的特点; 三是数据调阅频度不固定, 但操作人员对数据的响应速度有很高要求。这部分特殊业务以及数据保护, 要综合考虑存储系统的性能、数据保护技术以及成本, 建议采用大容量网络存储方式构建数据仓库,

结合分层存储技术,保障数据调阅性能;在数据保护技术选择上,可以采用存储底层硬件“单写式”双活系统集成技术,确保数据安全冗余和成本低廉,形成性价比较高的综合解决方案。

5 结语

医疗大数据不仅保障医院日常业务运行,同时还服务于机构科研创新等方面,医疗大数据安全保护十分关键^[13]。通过对不同层面双活系统集成技术的研究与合理应用,可以为医院医疗信息数据及业务连续性保护提供一定支撑,提高医院网络信息安全水平,使信息化、大数据更好地服务于医疗健康事业。

参考文献

- 1 李岳峰,胡建平,庾兵兵,等.医院信息化功能与技术标准框架设计与探讨[J].中国卫生信息管理杂志,2022,19(3):390-394.
- 2 刘同柱,王德斌,徐兵,等.智慧医院建设标准体系研究及实施策略[J].中国卫生信息管理杂志,2020,17(6):715-720.
- 3 张琼瑶,黄基,李倩文,等.基于人工智能的大数据治

理平台实践与探索[J].中国数字医学,2021,16(10):31-36.

- 4 魏智,黄昊.医院双活数据中心的设计与应用[J].网络安全和信息化,2022(5):63-65.
- 5 曹凯.基于SDN的双活HCI超融合架构在医疗信息化建设中的应用[J].电子技术与软件工程,2022(10):34-38.
- 6 上官斌.关于大型数据库双活难点问题研究[J].电脑知识与技术,2022,18(8):23-25.
- 7 许培海,黄匡时.我国健康医疗大数据的现状、问题及对策[J].中国数字医学,2017,12(5):24-26.
- 8 王灵芝,郝明.医疗大数据的特征及应用中的伦理思考[J].医学与哲学(A),2017,38(4):32-35.
- 9 庄一峰.基于大数据背景的医院网络信息安全防范分析[J].网络安全技术与应用,2020(11):138-140.
- 10 王靖夫.大数据平台数据安全体系架构研究[J].网络安全技术与应用,2020(11):76-77.
- 11 韦鑫,李星霖.基于电子病历系统应用评价标准的合理用药系统建设实践[J].无线互联科技,2020,17(17):60-61.
- 12 原国家卫生和计划生育委员会.基于电子病历的医院信息平台技术规范[S].北京:中国标准出版社,2014.
- 13 张振,杨翠涓,徐静,等.健康医疗大数据应用发展现状与数据治理[J].医学信息学杂志,2022,43(7):2-8.

(上接第 77 页)

参考文献

- 1 杨晓光,马力,李娜,等.医疗联合体内基于云计算系统的脑卒中分级诊疗模式研究[J].中国全科医学,2018,21(4):462-464,470.
- 2 王森,于广军,刘海峰,等.基于云平台的区域医联体信息系统研究与实现[J].中国医院,2017,21(8):19-23.
- 3 王毅,张芳,宋凡,等.医联体医疗信息平台研究与应用[J].医学信息学杂志,2022,43(3):62-65.
- 4 郑少宇,滕飞,马征,等.支持临床决策的医学知识图谱的构建与应用[J].重庆医学,2021,50(1):163-166.
- 5 张文芳,陈桢,刘旭东,等.支持细粒度属性直接撤销的CP-ABE方案[J].软件学报,2019,30(9):

2760-2771.

- 6 赵霞,李中华.“十四五”期间医院信息化建设发展的若干思考[J].中国医院,2021,25(1):64-66.
- 7 袁静,张敏,李暄,等.基于容器云的医联体应用模式创新研究[J].医学与社会,2019,32(11):27-31.
- 8 汤佳,申俊龙,王静成,等.基于公共价值理论视角下医联体发展路径研究[J].中国医院,2020,24(5):29-31.
- 9 陈有兰,李伟,陈渝.我国医联体现状研究及发展策略[J].中国医院,2020,24(8):1-3.
- 10 苗豫东,吴建,张亮,等.分级诊疗制度变迁回溯及“十四五”期间的关键政策建议[J].中国卫生政策研究,2021,14(3):1-6.