医院双活数据中心设计和应用

李 彬 苏悦洪 麦子铭 任忠敏 何彩升

（中山大学附属肿瘤医院信息科 广州 510060）

〔摘要〕 以中山大学附属肿瘤医院为例，在分析其医院信息系统现状的基础上，为实现医院主数据中心和容灾中心实时在线容灾目标，提出基于 EMC VPLEX 构建存储双活的数据中心，介绍数据中心的容灾方案并进行系统测试，为医院信息化建设提供借鉴。

〔关键词〕 双活数据中心；VPLEX；连续数据保护

〔中图分类号〕R－056 〔文献标识码〕A 〔DOI〕10.3969/j.issn.1673－6036.2017.02.007

Design and Application of Dual－active Data Centers of Hospitals *LI Bin, SU Yue－hong, MAI Zi－ming, REN Zhong－min, HE Cai－Sheng, Information Department, Cancer Hospital Affiliated to Sun－sen University, Guangzhou 510060, China*

〔**Abstract**〕 In order to achieve the real－time online disaster recovery objectives of the main data center and disaster recovery center, the paper takes Cancer Hospital Affiliated to Sun Yat－Sen University as an example to analyze the status of Hospital Information System (HIS), puts forward the construction of dual－active storage data center based on EMC VPLEX, introduces the disaster recovery schemes of the data center, conducts system test, and provides reference for information construction of the hospital.

〔**Keywords**〕 Dual－active data center ; VPLEX; Continuous data protection

1 引言

随着虚拟化、云计算等信息技术的高速发展，医疗卫生信息化建设日臻完善。然而如果信息系统在运行过程中中断将会导致医疗服务的中断，诊疗信息数据丢失其损失无法估量，因此信息系统安全稳定运行日益成为医院运维的重中之重。传统的医院核心业务均采用双机存储热备，存储系统主备切换方式提供稳定服务，在医院前期的信息化建设中起到了一定的作用，但也存在如需切换停机时间导致业务中断等弊端。基于以上情况，实现生产中心和容灾中心两台存储之间的数据同步及内存镜像，使得虚拟化平台上所有业务系统可以在两个双活中心之间进行无缝切换，即任何一台存储出现硬件故障或有虚拟平台的物理服务器出现故障，都不会造成业务中断，保障业务的持续运行及高可用性[1]。本文以中山大学附属肿瘤医院为例进行介绍。

〔修回日期〕 2016－12－06

〔作者简介〕 李彬，助理工程师。

2 医院医疗信息系统现状

2.1 存储架构

随着医院对信息化建设的重视和投入的加大，医院业务量不断增加，对于信息系统的稳定性需求也不断增加。目前，中山大学附属肿瘤医院内部使用的医疗系统包括医院信息系统、电子病历系统、检验系统、影像存储与传输系统、心电超声系统、病理系统、病案数字化系统、移动护理系统、合理用药系统、体检系统、医院物流系统等。同时，该

院也正陆续建设基于大数据、数据挖掘、云平台等的复杂大系统。采用基于 DS5300 和 DS4800 两重存储，运用 IBM 存储镜像技术做了存储层面的镜像复制，见图 1。采用传统的基于两套存储层面的镜像双活架构，可使存储文件在不同存储内保留两份数据，确保在一台存储出现故障时可以通过手动方式切换到另外一台存储，从而避免数据丢失。

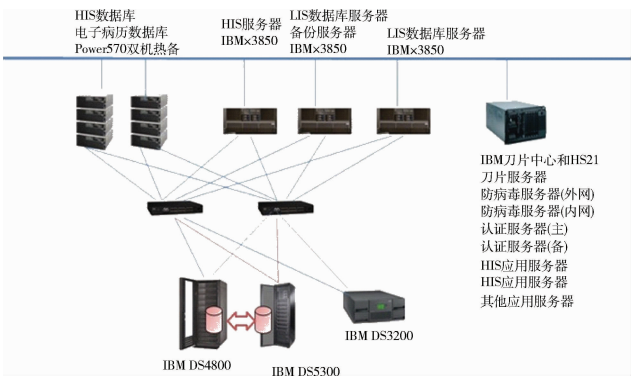

图 1 中山大学附属肿瘤医院存储架构

2.2 存在的问题

此镜像双活的架构出现以下几个问题：（1）因业务系统大部分基于 Windows 平台，所以在出现故障时，操作系统层面需要手动挂起存储盘，导致在故障切换出现数据恢复点时，数据库其实相当于从数据库挂机之后再恢复原来状态，所有 RPO 不等于零，RTO 在大约分钟到小时级别。（2）Linux 类型的操作系统，由于底层存储进行更换，为了识别到新的存储磁盘，需重启服务器，如果 Linux 系统通过如 heartbeat 方式设计双机热备，还必须采用双机前后主备服务器先后进行重启、双机配置需要重新写入等一系列恢复步骤。由此造成 RPO 数据不等于零，同时 RTO 恢复的时间也远超过通过冷备方式单机故障存储切换时间，出现越是高可用越难以恢复的情况。（3）采用磁带机恢复数据时，由于备份的数据库都是昨天的，所以会导致数据恢复时只能恢复已经备份的数据，同时通过最近的日志进行今天恢复，所有的 RTO 时间等于数据库恢复时间加上日志恢复时间。而且对于 RPO 的丢失有可能最大到 1 天左右，对于影像、检验这样医技临床系统来说，数据丢失会造成严重的医疗事故，对医院发展的影响是巨大的。所以如何在最短的时间成本上，最完整地恢复系统，保障医院业务正常运行，降低数据丢失风险，是设计数据中心双活解决方案的目标。中山大学附属肿瘤医院为了实现两个数据中心双活容灾，采用 EMC VPLEX METRO 方案加 RecoverPoint 连续数据保护方案，可以实现医院原有 IBMDS5300、DS4800 异构存储的虚拟化整合，而且通过 VPLEX 大内存提高旧存储的 IO 读写速度，提高使用效率。同时方案利用 VPLEX 的虚拟化设备，通过两套光纤交换机，实现两个数据中心之间的“2 +2”双活架构，来达到 RTO 为零、RPO 为零的目的。

3 医院数据中心容灾方案

3.1 方案设计要求

3.1.1 *系统兼容性* 方案充分考虑双活方案的体系兼容性，需配置集 FC – SAN、IP – SAN、NAS 于一体的高性能统一存储，具有分层存储功能；并且通过存储虚拟化异构整合目前医院在运行的多套存储系统。

3.1.2 *可靠性* 方案设计中考虑系统的双活架构、同步方式、容灾发生应急方案等方面，系统具有高可靠的保障，考虑到物理容灾以及逻辑容灾的保障，使得数据丢失容忍量（RPO）为 0，业务恢复时间（RTO）尽可能低，同时对于关键设备以及链路均采用冗余设计。

3.1.3 *灵活性、扩展性* 双活架构设计需要有灵活的扩展性，在现双数据中心的架构下，考虑可以达到两地 3 中心的技术架构，对于关键的双活设备采用同步或异步方式来满足不断增长的业务需求。

3.1.4 *可管理性* 作为中山大学附属肿瘤医院的双活容灾基础架构，必须建议完善的运行管理和维护手段来管理双活设备、存储设备、数据保护设备。

3.1.5 *实用性* 系统的设计要求必须从实际出发，依照医院现有的信息化设备，满足医院双活容灾的需求，选择具有最佳性能价格比的系统硬件和软件，减少系统运行费用，同时充分考虑未来的业务需求。

3.2 方案实施

3.2.1 VPLEX 双活方案 两个数据中心均通过 VPLEX 虚拟化网关管理 5600 主存储和其他异构存储，形成存储虚拟化的整合，同时基于 VPLEX 虚拟化设备，实现了两数据中心 METRO 架构以及“2+2”双活架构，见图 2，达到 RTO 为零、RPO 为零的建设目标[3]。

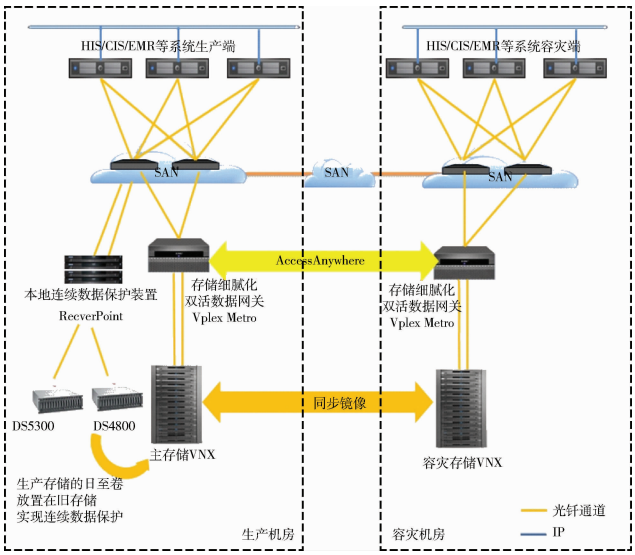

图 2 基于 VPLEX 存储虚拟化网关的双活数据中心方案存储构架

在 VPLEX 上的远程存储虚拟功能虚拟出虚拟卷给物理主机使用，底层两台 VNX5600 存储均为 Active，在两台存储底层划分相同大小卷进行镜像，同时在 VPLEX 系统中针对每一台物理主机构建跨多个 lun 的一致性组，保证在时间轴上数据一致性。对于物理主机来说，使用方式与传统的一样，主机只需要识别到 VPLEX 提供的虚拟卷即可，不需要修改原有使用模式。每当有一个 IO 写入 VPLEX 镜像卷时，均同时写入到两个数据中心的 VNX5600 存储中，保证两个数据中心存储数据的一致性。而当任意一套存储系统或者数据中心发生访问故障时，对于物理主机的业务连续性完全没有影响，另外一套正常存储设备可以继续支撑业务应用；且当故障存储恢复可以使用时，VPLEX 会自动将数据进行同步，可以自动化实现故障时数据中心正常运行，故障结束之后自动恢复，从而自动实现了远程双数据中心的双活。通过 VPLEX 实现了底层的存储双活，对于任何系统和数据库来说，始终使用的是 VPLEX 上的虚拟磁盘，至于底层的虚拟磁盘可以根据业务需求，合理地使用新旧存储。医院信息系统、集成平台等核心重要系统为了确保减少基于存储的单点故障，所有的双活磁盘均放在两套 EMC VNX5600 上，以保证系统读写性能和稳定性，同时对于一些旧的非核心系统如病案统计、合理用药等，可以采用基于 EMC VNX5600 和 IBM DS5300 两套新旧存储虚拟出来虚拟磁盘来搭建数据库，虽然性能会基于最差的存储的性能，但可以合理利用旧有的存储。EMC VNX5600 采用分层存储技术，通过将所有 SSD、SAS、SATA 磁盘组进行虚拟化，形成一个大 POOL，热点数据存放在 SSD 磁盘，常用数据存放在 SAS 磁盘，不活跃数据存放在 SATA 磁盘，从而提高热点数据 I/O 读写能力。自动分层存储技术给业务带来了更高性能上的提升，并且能节省成本投入[3]。

3.2.2 数据库双活升级改造 VPLEX 存储双活只是避免了存储底层的如磁盘故障、逻辑错误、磁盘数据丢失等单点故障，对于核心数据库来说，数据库使用服务器也是故障点之一。传统的数据库如 MY SQL、SQL SERVER 等均采用主备方式提供对外服务，部分服务器处于空运行，而且在切换过程中可能出现切换宕机风险。医院将原有的 DB2 数据库迁移到 Oracel 数据库，将原有的主机热备方式升级为“2+1”模式同时对外提供服务，见图 3，RAC 架构大大改善了以往“双机单活”双机热备模式下的硬件资源利用[4]。

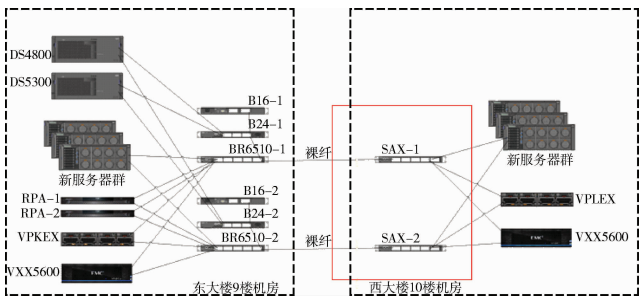

图 3 服务器双活数据中心架构

3.2.3 VMware 虚拟应用 数据库可以在两个数据中心机房同时对外提供服务，上层的应用服务器

采用 VMware 服务器虚拟化技术，通过在东大楼和西大楼机房同时部署一体化医生工作站虚拟化 ESXi 主机（东大楼 2 台，西大楼 1 台），实现一体化医生工作站应用的双活部署，当其中一台服务器有问题时，快速迁移到另外一台服务器保障业务快速恢复，通过虚拟化高可靠集群（HA）的构建有效保障了系统运行的连续性[5]。

3.2.4 双活数据中心网络规划 数据库双活对东西座两个节点间通讯链路的带宽和不间断性能需求较高，建议东座服务器核心交换机与西座核心交换机之间部署单独的存储备份专用链路，包含两条物理链路聚合实现高带宽和冗余性，见图 4。东座两台服务器核心交换机上为存储备份专用端口配置 VPC，完成专用链路的链路聚合配置。端口配置 trunk，只放通存储备份使用的 vlan180、vlan181、vlan800、vlan801。西座核心交换机目前为 VSS 结构，可以使用聚合端口配置两条存储备份专用链路。

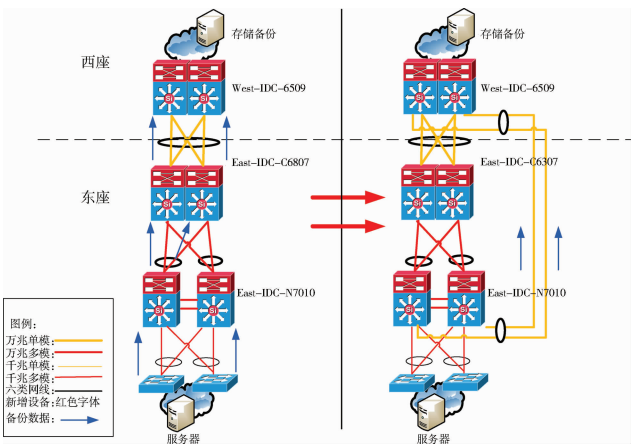

图 4 双活数据中心网络拓扑

3.2.5 连续数据保护方案 在解决了物理灾难带来的业务中断及数据丢失的同时，为避免数据文件误删除、数据库 bug、数据库升级失误等逻辑故障导致的数据丢失，医院构建连续数据保护策略，结合 VPLEX 引擎作为拆分装置，利用旧有的 IBMDS5300 和 DS4800，将生产存储的日志卷以 I/O 级别复制到旧存储，见图 5，故障发生时，只需通过选择需要恢复点，在测试环境恢复测试，在原存储上进行数据恢复，3 步简单操作，即将数据恢复至故障发生前的任意时间点，保障了 RPO = 0，且大大缩短了故障恢复时间，避免了因发生逻辑性数据错误时带来的损失，实现连续数据保护。

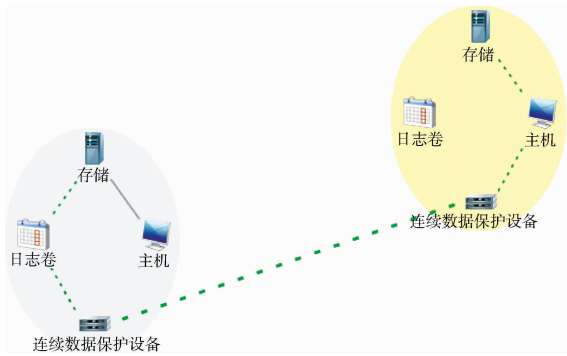

图 5 CDP 连续数据保护模型

旧有的存储系统（DS5300/DS4800）接入到 VPLEX 管理，配置 RecoverPoint - CDP，将 VNX5600 生产存储上的 CDP 日志卷放置于镜像后的旧存储系统集群上，存储规划为现有存储的 1/3 作为日志卷。经过测试连续数据保护的时间正常的业务运行大约是 3 个月到半年，如果出现比如大量的数据库写入和修改，保护的时间会被压缩到 1 周时间。连续数据保护进行数据库恢复时，需要先将日志卷在其他服务器挂载，确认需要恢复的数据库内容，然后将原有生产环境的服务器卸载数据库磁盘，如 Windows 平台进行磁盘脱机、Linux 平台进行 unmount 操作。进行日志卷恢复后，再重新挂载磁盘，重新启动数据库，因为数据库处于正常的运行状态，可以有效避免比如数据库逻辑错误，或者一些错误操作使得数据库宕机，使得 RPO = 零，RTO 依据日志卷恢复速度。

4 双活系统测试

4.1 VPLEX 双活环境故障测试

在 VPLEX METRO 的故障测试中，一体化核心系统采用 Oracle RAC3 节点在两个数据中心，其中一个数据中心是两个节点，另一个数据中心是一个节点。通过 SAN 网络以及核心交换机专线光纤走心跳线，数据存储统一采用 VPLEX 虚拟卷，数据存放在两个数据中心中。在 VPLEX 的 16 项故障测试

中，见图 6，测试结果符合预期，业务情况正常，满足双活需求，见表 1。

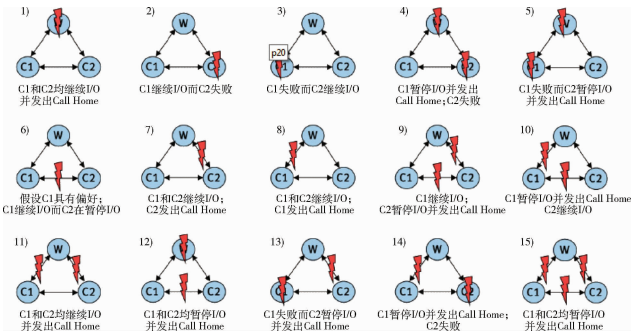

图 6 VPLEX METRO 故障状态

表 1 VPLEX 双活测试结果

序号	故障情景	预期结果	测试结果
1	Witness 故障	业务正常	业务正常
2	Cluster2 故障	业务正常	业务正常
3	Cluster1 故障	业务正常	业务正常
4	Witness 故障，C2 故障	业务中断	业务中断
5	Witness 故障，C1 故障	业务中断	业务中断
6	VPLEX WAN 口连接中断	业务正常	业务正常
7	Witness 和 C2 Manage Server 连接中断	业务正常	业务正常
8	Witness 和 C1 Manage Server 连接中断	业务正常	业务正常
9	Witness 和 C2 Manage Server 连接中断，WAN 口连接中断	业务正常	业务正常
10	Witness 和 C1 Manage Server 连接中断，WAN 口连接中断	业务正常	业务正常
11	Witness 和 C1，C2 Manage Server 连接中断	业务正常	业务正常
12	Witness 故障同时 WAN 口连接中断	业务中断	业务中断
13	Witness 和 C2 Manage Server 连接中断，C1 故障	业务中断	业务中断
14	Witness 和 C1Manage Server 连接中断，C2 故障	业务中断	业务中断
15	Witness 和 C1，C2 Manage Server 连接中断，WAN 口连接中断	业务中断	业务中断

4.2 连续数据保护测试

在 RPA 测试方案中，在随意一个时间点进行误删除操作，通过 RPA 回滚到删除前一个时间点进行数据恢复，再经过 RPA 覆盖数据库盘进行数据库恢复，测试结果找回了原来被删除数据库文件，而且数据库恢复之后可以正常启动。

5 结语

随着医院信息化的发展，医院各部门对信息系统有高度的依赖性，更加要求信息系统的高可用性。中山大学附属肿瘤医院通过基于 VPLEX 虚拟化设备进行双活数据中心建设，在满足存储双活的前提下，进行应用和数据库层次双活，为以后两地 3 中心打下了基础。容灾必须通过实战和演练，增强容灾意识，维护着信息系统稳定。

参考文献

1 汪兆来. 基于存储虚拟化技术的双活数据中心医院信息系统容灾平台研究与设计［J］. 中国医学装备，2015，(9)：65－68.

2 杨永福，黄黎明，尤超，等. 医院双活数据中心容灾模式建设的探索与实践［J］. 中国数字医学，2015，(8)：79－81.

3 董伟，李郑刚. 自动分层存储技术在现代存储中的应用分析与前景展望［J］. 云南大学学报：自然科学版，2013，35（S2）：54－58.

4 王文翠，李志强，秦芳，等. Oracle 10G RAC 在医院信息系统中的应用［J］. 中国数字医学，2015，10（12）：107－109.

5 巴江波，陈江，淡攀东，等. VMware 虚拟软件在医院数据中心建设中的应用研究［J］. 中国医疗设备，2015，30（2）：78－80.