doi: 10.3969/j.issn.1007-2861.2012.03.016

银行业数据中心系统灾备技术

王仲怡, 赵正德

(上海大学 计算机工程与科学学院,上海 200072)

摘要: 近年来随着数据中心体系结构的建设,灾难备份技术越来越受到各业界的重视.介绍数据中心系统灾备的意义和目标,给出灾备和恢复的指标体系以及灾备的实现技术.针对银行业数据中心的单点故障、逻辑故障以及自然灾难的情况,提出并设计数据及应用的备份和恢复方案,使业务能够可靠、安全和持续的运行.

关键词: 数据中心;灾难备份;数据恢复

中图分类号: TP 392 **文献标志码**: A **文章编号**: 1007-2861(2012)03-0299-06

Disaster Recovery of Data Center Systems in Banking Industry

WANG Zhong-yi, ZHAO Zheng-de

(School of Computer Engineering and Science, Shanghai University, Shanghai 200072, China)

Abstract: With the rapid development of data center construction in recent years, the importance of disaster recovery is increasingly recognized in all industries. This paper discusses the significance of data recovery for open systems in the banking industry, and introduces an index system of data and system recovery. A strategy is proposed to recover a system on the condition of one-point error, logic error and natural disaster in order to maintain security, sustainability and reliability of the system operations.

Key words: data center; disaster recovery; data backup

随着互联网、云计算、物联网等应用技术的不断发展,在金融界尤其是银行业这类信息技术应用程度较高的行业,其业务发展对信息系统的可用性和业务持续性提出了很高的要求.随着业务数据大集中工作的不断深入,集中到一起的不仅仅是数据,还有风险.灾难发生是无法预料且很难避免的,只有采取积极的灾备措施,建设完善的灾难备份系统,事先制定必要的灾难恢复方案和计划并加以演练,才能确保信息系统在灾难发生时能够及时恢复正常运转,将灾难造成的损失降到最低[1].

银行业信息系统是我国八大类重要信息系统之一,其安全运行直接关系到我国社会经济的稳定.当前工行、农行、中行、建行、招行和中国银联基本已完成了数据大集中的工作,但是配套的灾难备份中心和应急体系建设处于严重滞后状态.从中国农业银行2005年8月停机和中国银联2006年4月20日停机都可以看到情况的严重性[2].如何确保在全国大集中环境下的数据及信息系统安全已成为了当前银行业信息科技工作的重要课题和任务[3].因此,对于银行信息系统,做好灾难备份,加强信息安全事件的应急处理工作,增强在灾难打击中恢复业务的能力是信息安全中极为重要的一环,也是信息安全保

收稿日期:2011-02-16
基金项目:上海市教委重点学科建设资助项目(J50103)
通信作者:赵正德(1956~),男,教授,研究方向为计算机协同技术、操作系统与信息安全技术. E-mail:zhdzhao@163.com

障体系的基础建设.

1 灾备/恢复的指标和建设模式

1.1 灾备/恢复的指标[1]

容灾指标——恢复时间目标(recovery time object,RTO)、恢复点目标(recovery point object,RPO)和网络恢复目标(network recovery object,NRO)三要素可以作为指标性的主要评判依据[4].

RTO是容灾恢复的时间指标.广义的RTO是指从灾难发生造成业务中断,到通过各种方法恢复业务,使业务能够得以继续所需要的时间.通常越短的RTO意味着越高的容灾能力.狭义的RTO是指从决定进行容灾切换到业务可以继续运行的时间.一般用狭义的RTO指标评价IT层面的容灾能力.

宕机之后数据开始恢复的时间点称为恢复点.恢复点指标RPO,就是当业务恢复后,可以与灾难发生前那个时间点的状态进行相同的工作.通常RPO对应着灾难造成的数据丢失.如果RPO为0,相当于没有任何数据丢失,业务恢复后,可以进行与灾难发生前完全相同的工作,不需要任何额外的处理.

NRO即为网络恢复时间.由于现在的业务越来越依赖于网络,除非特别的业务类型(如账务清算),网络如果没有恢复正常,即便恢复了处理主机和数据,也无法提供对外服务.因此,NRO通常要小于RTO,大于RTO的NRO是没有意义的.

1.2 数据中心灾备建设模式

数据中心灾备模式以数据容灾为核心,以业务连续性为重点,以实现安全生产和运营为目标[5].灾备中心建设模式主要包括同城灾备中心、异地灾备中心和两地三中心这3种建设模式[6].

同城灾备中心是指灾备中心和生产中心建立在同一城市中,两个中心的距离一般都在20 km以上.同城灾备中心保留有生产环境的同步数据,它与生产中心的距离越远,对区域性灾难的抵御效果就越好,但是对生产系统的性能或是灾难恢复目标的影响也就越大.在现实的技术和投资条件下,同城灾备中心具有高等级、快速响应和高效率重续运行的特点,对80%左右的灾难事件有抵御能力.

异地灾备中心一般建立在与生产中心处于不同地域的其他城市中.没有严格规范要求异地灾备中心与生产中心的距离,但一般都在200 km以上.由于两个中心间的距离很远,如果异地灾备中心与生产环境的数据进行同步复制,可能会严重影响应用系统的性能,因此,异地灾备中心与生产中心间一般进行异步数据复制.以异地灾备中心的数据进行恢复,会丢失一定量的数据,如果仅仅是区域性的灾难就切换到异地灾备中心,业务恢复的时间也可能会较长.

两地三中心一般是指一个生产中心、一个同城灾备中心、一个异地灾备中心.通常生产中心的数据同步复制到同城灾备中心,同时还异步复制到异地灾备中心,即实现同城灾备中心的零数据丢失.在这种灾备中心模式下,80%的区域性灾难都可以通过最近的同城灾备中心来抵御,能够快速高效地在同城灾备中心实现业务的持续性.如果同城灾备中心进行的是同步数据复制,就能够实现数据的零丢失;如果采用活动备援站点的方式建设同城灾备中心,由于该中心以应用集群的方式工作,在发生区域性灾难时,切换到同城灾备中心,不仅能实现零数据丢失,还能实现应用的实时无缝切换.在出现小概率的大范围灾难时,如自然灾害地震时,同城灾备中心与生产中心同时不可用,可以切换到异地灾备中心,虽然会丢失少量的数据,但通过灾难恢复计划,实施经过日常灾难演练的步骤,在业务容许的时间内,可在异地灾备中心恢复业务系统.

与同城灾备中心模式和异地灾备中心模式相比,两地三中心的灾备模式结合了两种单中心模式的优点,提高了数据的冗余.灾备系统能够在遭遇区域性和较大范围的自然灾害时较快地响应,保证业务的连续性,提高RPO和RTO.该模式已被很多行业尤其是银行业广泛采纳.

2 存储虚拟化和高可用性技术

2.1 存储虚拟化技术

存储域网络(storage area network,SAN)能满足企业数据存储的需求,有助于应对管理存储的挑战.尤其在银行业的数据中心,面对海量数据时,SAN虚拟存储灵活和高效的优点尤为突出.本研究利用来自厂商HDS和EMC的存储系统创建虚拟存储池,使数据中心可以利用未被使用的盘卷,提高业务的灵活性.SAN虚拟化解决方案通过提供存储卷的单一视图,简化存储的管理.带内(in-band)虚拟化引擎通过主机和存储系统的数据通道,将逻辑卷分配给主机;带外(out-of-band)虚拟化引擎通过其他的网络和主机系统通讯.SAN虚拟化技术支持开放系统环境中的高性能性和持续可用性.

2.2　高可用性技术

高可用性是指通过尽量缩短因日常维护操作(计划内)和突发的系统崩溃(非计划内)所导致的停机时间,以提高系统和应用的可用性. 本研究在设计异地复制和实时远程切换时,还考虑了实现本地数据的快速恢复,通过本地高可用系统和本地数据中心建立本地容灾,容忍硬件毁坏等灾难造成的单点失效,在绝大多数故障情况下不需要操作或干预即能恢复信息系统的业务运行. 镜像技术是高可用性技术的一种,对灾难备份的实现有着重要意义. 本研究以 AIX 环境下的 LVM(logical volume manager)为例,通过逻辑卷管理配置,对逻辑卷做镜像. 每一个逻辑卷数据最多可有 3 份副本,即 1 份源,2 份源副本. 这 3 份副本可以分布在不同的物理分区中,即可以对应到不同的物理介质或物理存储设备. 从操作系统来看,每一个镜像的逻辑分区会对应着 2 ~3 个镜像的物理分区,这些分区的镜像关系通过 LVM 维护. 根据预定的策略,LVM 会管理如何对镜像进行读写操作. 随着 SAN 扩展技术的发展,可以将光纤通道 SAN 拓展到任何存在 IP 网络或具有一定带宽的地方.

另外,基于磁盘系统的灾难备份技术可采用同步数据复制模式和异步数据复制模式. 目前大多数高性能的企业级存储都具有远程复制功能,当灾难发生时,可通过启用灾备中心存储上的数据来实现信息系统的恢复. 比较典型的企业级存储复制技术产品有 IBM 的 Peer-to-Peer Remote Copy(PPRC)、EMC 公司的 Symmetrix Remote Data Facility(SRDF)以及 HDS 公司的 TrueCopy 等.

3　银行业数据中心系统灾备方案设计

在中国人民银行 2008 年 2 月 4 日发布的金融行业标准 JR/T 0044—2008《银行业信息系统灾难恢复管理规范》[7]中,明确了灾难需求等级的确定,通过业务的影响力和时间的敏感性划分为 1 ~3 级. 根据不同的灾难需求等级来确定最低恢复的要求,并以 RPO 和 RTO 来体现. 灾难恢复需求等级越高,对 RPO 和 RTO 的要求也越高,即反映了 RTO/RPO 与灾难恢复能力等级的关系. 灾难恢复能力的等级可分为 1 ~6 级,其中 6 级为最高,从而提出了不同的 RTO 与 RPO 的要求. 以 5 级和 6 级为例,5 级要求 RTO 在数分钟至 2 d 内,RPO 在 0 ~30 min 内;6 级要求 RTO 在数分钟内,RPO 为 0,即没有生产数据丢失.

根据银行业务的特点,综合系统使用频度、重要性及中断影响因素,大致给出一旦灾难发生后,各系统恢复的优先级. 优先级高的系统被预定义为关键业务系统,如核心交易系统,会优先考虑恢复,从而减少机构的损失,减轻灾难所造成的不良影响[8]. 因此,本研究要求关键业务系统的灾难恢复能力至少达到 5 级标准,相应的系统灾备指标如下:RTO < 灾难发生后 6 h,RPO < 灾难发生前 15 min 的数据,NRO < RTO 灾难发生后 6 h.

3.1　灾备系统架构的设计

图 1 所示为灾备系统架构设计方案 A. 在标准两地三中心方案的基础上,生产中心部署两台存储,一台存储为生产系统提供服务,并以基于存储的复制方式同步复制到另一台存储上. 两台存储存放在不同的生产机房以降低单机房内偶发的风险、不设立同城灾备中心,将数据通过基于存储的复制技术,从生产中心异步复制到异地灾备中心.

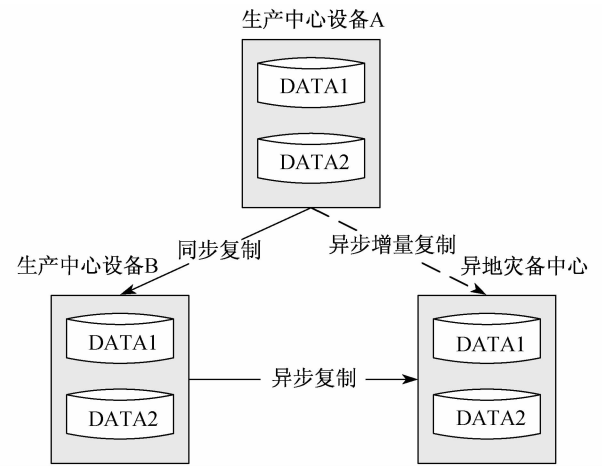

图 1　方案 A 的灾备系统架构

Fig. 1　Disaster recovery structure of Plan A

图 2 所示为灾备系统架构设计方案 B. 在标准两地三中心方案的基础上,生产中心部署两套存储,通过 LVM 卷镜像访问存储,在同城灾备中心同步复制生产中心的数据,在异地灾备中心通过异步复制技术异步增量复制生产中心的数据. 本方案可以实现生产中心本地高可用性. 生产中心、同城灾备中心和异地灾备中心以 2∶1∶1的比例配备存储设备.

综合比较 A 与 B 两种方案,结果如表 1 所示. 方案 A 和方案 B 都有较多较成熟的案例支持. 考虑到我国的大型股份制银行在 IT 运维方面的投资较大,且对存储的本地高可用性要求较高,日常运维都有较多分工明确的专业技术人员支持,因此,本研究选择方案 B 作为实施方案.

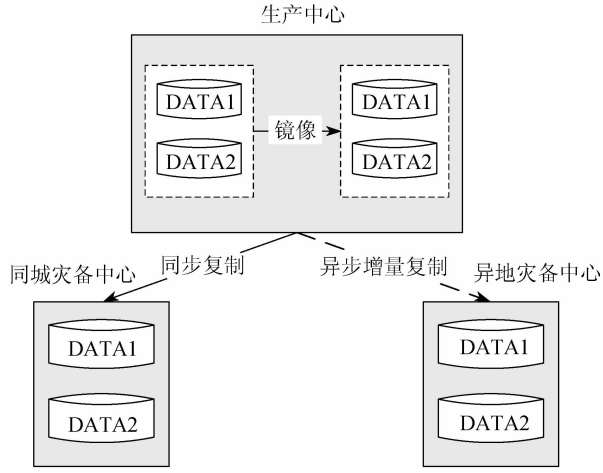

图2 方案B的灾备系统架构

Fig. 2 Disaster recovery structure of Plan B

表1 A与B两种方案的比较

Table 1 Compared Plan A with Plan B

指标	方案A	方案B
RTO	一般	好
本地高可用性	好	无
建设复杂度	较复杂	较简单
日常运维度	较复杂	较简单
投资	一般	低

3.2 银行业灾备方案设计

银行业数据信息中心系统每天有海量业务需要处理,并且要求系统提供7×24 h全年不间断的业务处理[9],这是对系统的稳定性、连续性提出的重大考验. 为提高系统的抗风险和抗冲击的能力,本研究提出了3种高效、灵活和可靠的灾备方式.

3.2.1 基于中间件和数据库集群的灾备方案

为提高系统中间件和数据库的高可用性,降低RTO和RPO,本研究提出了如图3所示的方案[10]:① 数据库和中间的Lpar独立搭建;② 数据库和中间件分别采用集群方式.

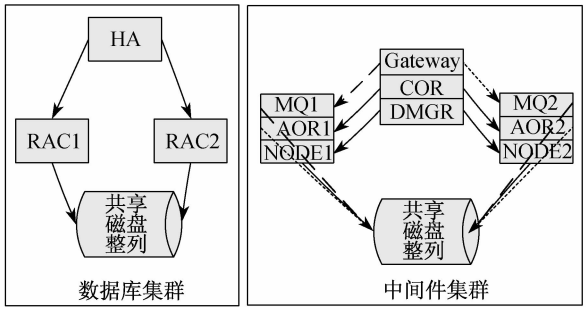

图3 基于中间件和数据库集群的灾备方案

Fig. 3 Disaster recovery strategy for middleware and database

本方案适用于在集群中单边Lpar宕机且共享磁盘整列能正常运作的情况. 在两台Lpar正常运行时,集群起到的是系统负载平衡的作用. 当集群中某一个Lpar宕机后,另一台Lpar将会独立工作,对于应用或系统前端几乎没有任何影响,可以说是透明的.

实验1 MQ集群方案.

```
$ echo " dis channel( * )" | runmqsc GW
1 : dis channel( * )
   CHANNEL(SYSTEM.DEF.CLUSRCVR)    CHLTYPE(CLUSRCVR)
   CHANNEL(SYSTEM.DEF.CLUSSDR)     CHLTYPE(CLUSSDR)
   CHANNEL(SYSTEM.DEF.SVRCONN)     CHLTYPE(SVRCONN)
   CHANNEL(TO.GW)                  CHLTYPE(CLUSRCVR)
   CHANNEL(TO.Z1)                  CHLTYPE(CLUSSDR)
   CHANNEL(TO.Z2)                  CHLTYPE(CLUSSDR)
```

运行runmqsc显示MQ的集群通道情况. 外部通道通过GW的接收通道和发送通道收发消息,而集群内部根据记录在GW中的完整信息库来完成消息的分发. 当Z1和Z2都能正常工作时,Z1和Z2均衡地处理队列中的消息,而当其中一个队列管理器发生故障时,另一台队列管理器会立即承担起故障机的任务. 因此,可以说这种方案的RTO和RPO约为0.

3.2.2 基于nfs的中间件、数据库及应用程序的灾备方案

在某些情况下,由于系统管理员或者系统操作员的疏忽或者缺乏经验,容易破坏系统中间件、数据库或应用程序. 为避免这种情况的发生,一方面需要加强对系统用户权限的管理,另一方面需要对中间件及应用作相应的备份,使系统能够在最短的时间内恢复正常.

图4为本研究设计的一种基于nfs的中间件、数据库及应用程序的灾备方案. 通过网络文件系统nfs将需要备份的Lpar与nfs Server相关联,在各台Lpar上创建脚本,通过cron机制,使备份脚本每12 h自动运行,在相应的目录中创建备份文件.

实验2 cics region备份和恢复.

图5为交易中间件cics的备份脚本,通过export cics region并将其压缩成.Z的文件,存放在nfs/sharebkup下的/bkup.cics.list目录下. 为了节省文件系统的空间,每台Lpar的备份文件每12 h更新一次,新的备份文件将覆盖旧文件.

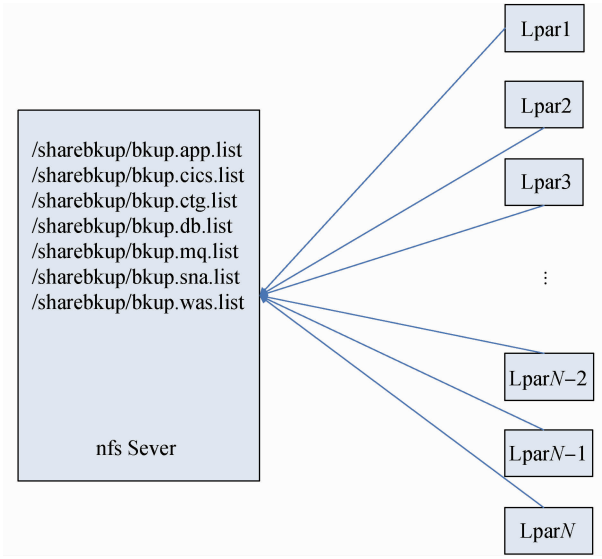

图 4　基于 nfs 的中间件、数据库及应用程序的灾备方案

Fig. 4　Disaster recovery strategy for the middleware, database and app based on the nfs

```
_logfile( ) {
    date6 = `date '+ %y%m%d %T'
    logdir = /usr/wys/base/`hostname`;  [ -d $logdir ] && mkdir-p $logdir
    logfile = $logdir/`basename $0`.`hostname`.log; touch $logfile 2 >/dev/null; chmod 666 $logfile 2 >/dev/null
  }
  _bkupdir( ) {
    df-gt |grep-w sharebkup  >/dev/null 2 >&1
    [ $? -ne 0 ] && { mkdir-p /sharebkup 2 >/dev/null; mount 22. 188. 129. 240:/sharebkup /sharebkup;}
    df-gt |grep-w sharebkup  >/dev/null 2 >&1
    [ $? -ne 0 ] && { echo "Erro:$date6 #$0 目标文件系统/sharebkup 无法 mount at `hostname`" ≫$logfile; exit 1;}
    bkupdir = /sharebkup/`basename $0`; mkdir-p $bkupdir 2 >/dev/null
  }

  _bkup_cics( ) {
    for list in `lssrc-a |grep-w cics |tr '.' ' ' |awk '{print $2}'`;do
      bkupfile = $bkupdir/`hostname`.$list. cicsdmp. Z
      cmd = "/usr/lpp/cics/bin/cicsexport-r $list |compress >$bkupfile"
      echo "Time:$date6 #$0— > $bkupfile" ≫$logfile; eval $cmd 2 >&1 |tee-a $logfile
    done
  }

  lable = cics; _logfile; _bkupdir; _bkup_$lable
```

图 5　cics 备份脚本(bkup. cics. sh)

Fig. 5　Shell for cics backup (bkup. cics. sh)

当因系统磁盘损坏造成 cics region 数据丢失时,可以用以下的方式恢复:

```
$ uncompress CICS_Z1. Z
$ smitty cics- > Import cics region
$ cicscp-v start region CICS Z1 StartType = cold
```

本实验中,RTO < 15 min,RPO = 0,说明这种方式是有效而且可靠的.

3.2.3　系统的日常灾备和恢复方案

为了提高操作系统的可靠性和安全性,定期进行系统备份和恢复演练是十分必要的. 当操作系统数据遭到破坏,系统无法正常启动时,就需要通过系统备份恢复系统. 当发生较大灾难时,系统遭到毁灭性破坏,此时就需要立即切换到灾备系统,而灾备系统是通过生产系统的系统备份数据实现系统同步的,可通过异步或者同步方式完成应用数据同步[11].

实验 3　生产中心与灾备中心系统切换.

如图 6 所示,假设生产环境的系统 LparA 遭到破坏,需要立即切换到灾备环境 LparB,该 Lpar 有数据库 rac 集群,磁盘信息保存完好,并且定期进行系统备份,保证 LparA 操作系统和 LparB 操作系统环境的一致性. 此时,系统的应急切换步骤如下:① 断开 HDS 远程拷贝(TC/HUR);② 通过远程 HMC 登录灾备孤岛环境;③ 在 LparB 上 varyonvg,包括 app,dmp vg;④ mount /oracle 和/arch1 文件系统;⑤ 启动 HA 并检查 concurrentvg rac_vg 状态;⑥ 在两节点执行 crsctl start crs 启动 oracle 数据库;⑦ crs_stat-t 检查数据库运行状态;⑧ 数据一致性验证;⑨ 实施网络切换,打开灾备孤岛环境接管原生产环境.

由于银行业数据中心操作系统关系到能否正常的进行日常业务操作,此处还推荐定期对系统进行磁盘备份,并且维护生产环境的磁盘信息. 在敏感的时间节点,例如节假日前后,对系统和应用进行磁盘备份是十分必要的.

4　结束语

灾难备份技术在信息化社会越来越受到高度的重视,而银行业面对日益庞大增长的业务数据,如何建设基于业务需求的容灾备份系统,已成为必须面对的严峻现实.

本研究首先介绍了关于银行业数据中心系统灾备的意义和目标,同时给出了灾备和恢复的指标体

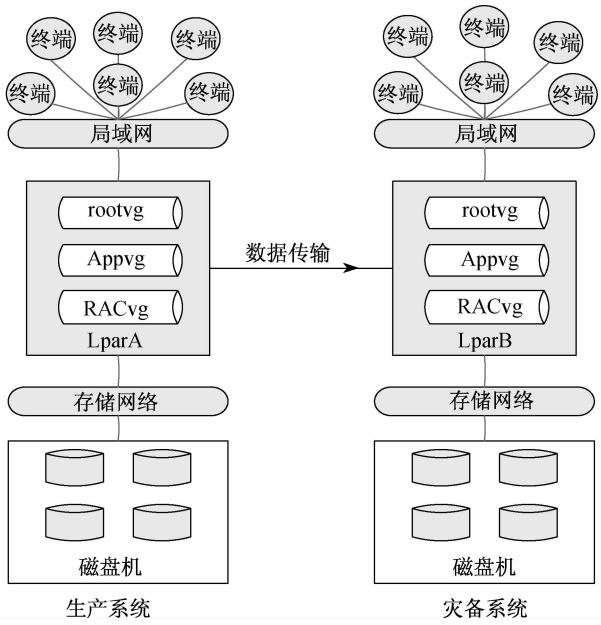

图6 生产中心与灾备中心系统切换

Fig. 6 Swiching the system from production center to disaster recovery center

系以及灾备的实现技术,并基于成本、技术、管理等多方面因素,设计并分析了一种能够适用于银行业数据中心的灾备架构方案. 其次,设计和验证了基于中间件、数据库和应用的备份方案. 这种方案基于nfs,自动完成备份任务,在单台系统发生故障时能快速恢复系统,并且这种方案的实现和维护是简单和灵活的. 最后,设计了基于 AIX P 系列小型机数据中心的应急切换方案,为切换演练和灾难发生时的系统切换提供参考依据. 这些灾备方案,无论是针对单点故障、逻辑灾难,还是在自然灾难的情况下,都可以实现数据及应用的实时备份和恢复,满足业务不中断的要求,达到了灾难备份的目的.

参考文献:

[1] 朱洪梅,孙玉华. 银行系统备份与恢复流程设计[J]. 金融电子化,2003(6):1-3.

[2] 杨鹏. 容灾备份系统中的同步策略研究及效率分析[D]. 成都:电子科技大学,2009:5-8.

[3] IBM. 容灾白皮书[R]. 北京:IBM 中国信息支持中心,2006.

[4] 王婷婷. 银行信息化灾备模式应因需而择[N]. 国际金融报,2004-10-12.

[5] NOCK C. 数据访问模式——面向对象应用中的数据库交互[M]. 北京:中国电力出版社,2004.

[6] IBM. 容灾备份方案建议书[R]. 北京:IBM 中国信息系统服务事业部,2002.

[7] 中国人民银行. JR/T 0044—2008 银行业信息系统灾难恢复管理规范[S]. 中国人民银行,2008.

[8] 汪琪. 灾难恢复发展趋势与变革[J]. 信息网络安全,2009(6):4-7.

[9] 李伟,来勤愉. 银行研发中心开放平台数据备份方案[J]. 计算机工程,2007(10):2-6.

[10] 张艳,李强,李舟军,等. 信息系统灾难恢复体系结构[J]. 计算机科学,2006,33(6):101-105.

[11] GARCLA-MOLINA H, HALIRM N, KING R P, et al. Overview of disaster recovery for transaction processing system [C] // Proceedings of IEEE 10th International Conference on Distributed Computing System. 1990:286-293.