一种应用级容灾系统的设计与实现

万瑾慧，李 涛，胡晓勤，卢正添
（四川大学 计算机学院，四川 成都 610065）

摘 要：该文设计并实现了一种应用级容灾系统。该系统采用模块化的设计方法，通过多策略的数据备份和数据恢复技术，保护了数据的完整性和高可用性；使用数据复制的方法确保了主备应用系统之间的数据一致性；在灾难发生时，通过服务切换的方法使得备用应用系统能迅速地接替主系统继续提供服务，保持了应用的业务连续性，实现了应用级容灾。

关键词：应用级容灾；业务连续性；数据复制；服务切换

Design and Implementation of an Application Level Disaster Recovery System

WAN Jin—hui, LI Tao, HU Xiao—qin, LU Zheng—tian
(School of Computer, Sichuan University, Chengdu 610065, P.R. China)

Abstract: An application level disaster recovery system is designed and implemented in this paper. The method of modular design is adopted by this system, multi—strategy backup and recovery of data is used to protect the integrity and high—availability of data. Data replication ensures the data consistency between the primary and backup system. When disaster happens, business continuity is maintained by switching the services, so the disaster recovery of application level is achieved.

Key words: application level disaster recovery; business continuity; data replication; service switch

随着信息技术的发展，多数企业越来越依赖于IT系统管理生产、提供服务。但各种灾难及突发事件造成数据的丢失或业务长时间的停顿，都可能使企业的生产停滞、声誉受损和客户流失[1]，给企业带来严重损失。因此，容灾技术逐渐为人们所关注和研究。根据容灾的定义和分类[2]，应用级容灾可以保证企业的关键应用在灾难情况下的业务连续性[3](Business Continuity)，将灾难对系统的破坏和造成的损失降到最低。

基于此，本文设计并实现了一种应用级容灾系统。该系统架构于互联网上，通过采用数据备份、数据恢复、数据复制和服务切换技术，保证了灾难情况下的数据完整可用和应用的连续运行。

1 系统设计

1.1 设计思想

应用级容灾是指实施容灾的系统不论在何种情况下都能提供不间断的用户应用服务。如果企业只有一套应用系统，在受灾时就会由于单点失败而使应用服务中断，所以，在异地建立和应用系统相同的备用系统[4]，是实现业务连续性的必要保证。

因此要实现应用级容灾，首先，通过异地IT系统的部署在远程建立一套以上完整的、与本地系统相当的且可以互为备份的应用系统；其次，主备系统间通过数据备份和数据复制技术实现数据级容灾，构筑应用级容灾的基础；最后，通过服务切换使得备用系统能接替主系统对外提供服务，保证服务不中断。

本系统要求实施应用级容灾的应用采用数据库存储业务数据。因为在此基础上，可以避免考虑上层应用及其处理流程的复杂性和多样性，不必为消除应用间的差异而改建原有的应用，而且能使用数据库现有的一些成熟技术。

1.2 系统结构

系统结构如图1所示，该容灾系统可划分为应用服务器、管理控制中心和管理配置端3个部分。

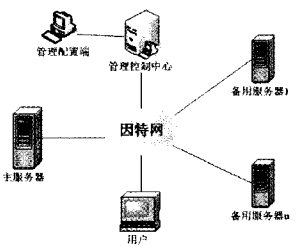

图 1 容灾系统结构图

管理控制中心负责容灾策略的发布和管理及建立主备应用服务器间的通信。管理配置端提供给系统管理员制定各种容灾策略，对系统和各个容灾任务进行管理、配置以及查看反馈信息。应用服务器包含一个主服务器和多个远程备用服务器，它们运行应用系统提供服务，用数据库存储业务数据。

1.3 构成模块

从功能上看，系统主要包括数据备份模块、数据复制模块、数据恢复

基金项目：国家863计划(2006AA01Z435)；国家自然科学基金(60573130,60502011)；教育部新世纪优秀人才计划(NCET—04—0870)

模块、服务切换模块、通信模块和管理模块。模块构成如图 2 所示。

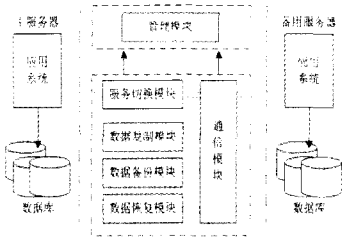

图 2 系统模块图

数据复制模块可以对主备数据库进行同步以保证两地数据的一致性。

数据备份模块和数据恢复模块能以多种策略对主数据库、表按时间点进行数据备份与恢复，保证了数据的完整性和高可用性。

服务切换模块能在主备服务器之间进行服务转移，保障了业务的连续性。

通信模块为系统中数据及控制命令提供了传输接口。

管理模块能配置系统、发布和管理各种容灾任务、记录日志等。

1.4 容灾过程

系统的一个典型容灾过程为：在受灾前，系统在主备数据库间进行数据复制，如有需要还可以按时间点对主数据库的数据进行备份。在灾难发生时，系统通过服务切换激活经数据同步过的备用服务器来提供服务。在灾难后的重建过程中，原主服务器通过数据复制可以达到最新的状态，也可以通过数据恢复还原到过去的某个时间点。

2 具体实现

2.1 数据复制模块

数据复制模块对主备数据库进行同步以保证数据的一致性。由数据容灾和应用容灾的关系可知，只有主备的数据保证了一致，两地的应用才能保持连续。

复制过程中，初始时，它要求主备数据库完全同构且内容相同。然后，它采用“写”操作重放的思想进行数据复制[5]，即对主数据库中的数据变化进行监控，把捕获到的数据变化实时地传递给备用数据库并进行重做。这样，主备数据库的数据进行了相同的改变，因而它们始终保持了一致。整个复制过程见图 3。

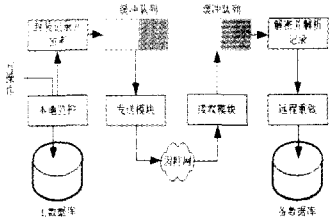

图 3 数据复制过程图

已有较多的数据库产品提供了数据库实时复制技术，如 IBM DB2 HADR、Informix HDR 和 Oracle Data Guard 等。在系统中，可以结合这些技术来完成数据复制。

经同步的备用数据库平时只能对主数据库传递过来的变化进行重做，但不接受上层应用的“写”操作，它可供上层应用进行查询。在灾难发生时，备用数据库被激活后，备用应用系统才能接替主系统提供完全的服务从而保证应用的业务连续性。

2.2 服务切换模块

服务切换实现了主备应用系统之间的服务转移，保证了应用的业务连续性，是实现应用容灾的重要手段。

服务切换模块包括 3 个职能模块：失效检测模块、服务激活模块和应用请求转发模块。它们协同作用，完成服务的切换，过程如图 4 所示。

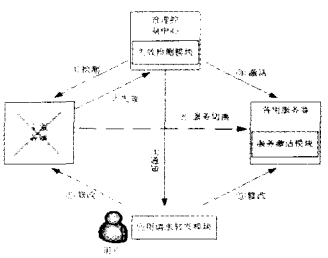

图 4 服务切换过程示意图

(1) 管理控制中心的失效检测模块对主系统的服务进行失效检测。

(2) 若检测到主系统的服务失效，将开始服务切换过程。

(3) 管理控制中心选择一台数据库经过数据同步的备用服务器进行激活。被选中的备用服务器调用服务激活模块激活数据库，使备用应用系统可接替主系统来提供服务。

(4) 激活成功后，管理控制中心告知用户端服务切换已经发生，并告知新的主服务器 IP。

(5) 用户如需服务，应用请求转发模块将对其发往原主服务器的数据报进行修改，使之发往新主服务器。

(6) 转发完成，服务成功从主系统切换到备用应用系统，实现应用容灾。

服务切换采用 IP 重定向技术以达到转移服务节点的目的，具体到系统中，数据包的修改是通过 NDIS（网络驱动接口规范）中间层驱动程序实现的。NDIS 中间层驱动程序位于传输层驱动程序和网卡驱动程序之间，在驱动级别上做修改，既高效，又能截获所有的数据包，而且对上层应用不造成影响。

2.3 数据备份 / 恢复模块

数据备份模块和数据恢复模块能以多种策略，比如网络或本地、手动或定时、数据库或数据表等对主数据库的数据进行备份和恢复。区别于数据复制，数据的备份和恢复只对主数据库、表进行操作，它适合于对由病毒、误操作等引起的灾难进行挽救。

数据库备份包括全备份、增量备份和差异备份，系统采用联机热备的方式实现数据库全备份，增量备份和差异备份在全备份的基础上对数据的变化进行备份。数据表备份分为整表备份和结构备份，结构备份只对表结

构的定义进行备份，整表备份在此基础上还对表中的记录进行备份。

数据恢复模块按恢复策略和时间点通过管理模块查询备份日志可找到备份数据集的存放地，然后调用通信模块将所需数据集传到待恢复服务器进行恢复。

对数据库进行恢复时，先进行全备份恢复，用 T_0 时刻备份生成的全备份数据集恢复数据库，通常是用它覆盖现有的数据库文件，之后进行增量或差异恢复。

用 T_n 时刻的增量备份集 $I(T_n)$ 进行增量备份恢复，要考虑两种情况：

(1)T_0 到 T_n 之间没有做过差异备份，如图 5 中①所示，应依次对 $I(T_1)$、$I(T_2)$ 直到 $I(T_n)$ 中的数据变化进行重做以完成恢复；

(2)T_0 到 T_n 之间曾做过差异备份，如图 5 中②所示，其中 T_m 是 T_n 之前最近的差异备份的时刻。这时，先对差异备份集 $D(T_m)$ 中的数据变化进行重做，然后依次对 $I(T_{m+1})$ 到 $I(T_n)$ 各个增量备份集中的数据变化进行重做即可。

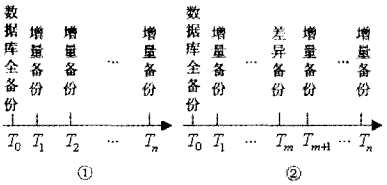

图 5 数据备份时间轴

用 T_m 时刻的差异备份集 $D(T_m)$ 进行差异备份恢复，只需在全备份恢复的基础上对 $D(T_m)$ 中的数据变化进行重做即可。

数据表的恢复与数据库全备份的恢复方法相似，此处不再赘述。

数据备份模块和数据恢复模块可以通过使用数据库提供的备份恢复管理器实现，也可以使用数据库提供的接口编程实现。

2.4 通信模块

通信模块采用 ICE（互联网通信引擎）中间件实现，使用 SSL 安全传输协议。通信模块分为发送模块和接收模块，它们驻留在应用服务器上。发送子模块从缓冲队列中取出记录，对其加密并构造成数据报发送到目标主机；接收子模块则接收数据报，验证数据报的完整性，从中分离出记录。

2.5 管理模块

在管理配置端采用 JPF（Java 插件框架）框架实现了图形化的界面，提供给管理员发布各种命令、查看反馈信息。在管理控制中心上负责对服务器配置信息和任务信息进行管理，对备份和恢复任务进行日志记录。

3 实验分析

在不同的网络带宽下，对系统进行了测试。测试的应用服务器环境为：2.4G 的 CPU，512M 内存，操作系统是 Windows 2000 Server，数据库采用 Oracle 9.2。

表 1 数据备份及恢复耗时表

（带宽：1Mbps）

备份类型	备份量大小	备份用时	恢复用时
库全备份	1617.8MB	35 分 13 秒	25 分 44 秒
增量备份	2.02MB	44 秒	35 分 24 秒
差异备份	2.23MB	44 秒	30 分 1 秒
整表备份	74MB	3 分 28 秒	4 分 28 秒
表结构备份	234KB	1 分 43 秒	2 分 2 秒

由表 1 得知，数据库大小在 1.5GB 时，全备份和恢复整个数据库耗时在半小时左右；对某个用户表的恢复，则只需几分钟。

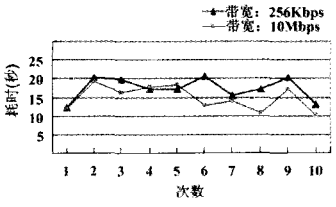

图 6 服务切换用时图

从图 6 可以看到，系统服务切换耗时在 10 秒到 20 秒之间，且基本上不受网络带宽的影响。

以上实验证明，系统有较快的数据备份和恢复速度及很短的服务切换时间。实验中，在服务切换后，用户能继续访问应用系统获得服务，系统实现了应用级的容灾。

4 小结

本文为企业提供了一种安全、高效、廉价的应用级容灾系统。当然，企业要在灾难中生存下来，除了部署容灾系统之外，还应该结合自身的资源，确定合理的容灾方案和制定完备的容灾恢复计划。在容灾系统建立之后，还需要建立完善的管理机制，才能最大限度地发挥容灾系统的作用。

参考文献：

[1] 李涛．网络安全概论[M]．北京：电子工业出版社，2004.

[2] 王树鹏，云晓春，余翔湛等．容灾的理论与关键技术分析[J]．计算机工程与应用，2004，(28)：54-58.

[3] Menkus，Belden．New importance of 'business continuity' in data processing disaster recovery planning[J]．Computer & Security，1994，13(2)：115-118.

[4] 周焕平，张士平．远程应用级容灾系统架构设计与研究[J]．计算机工程，2006，32(10)：277-279.

[5] Lars Frank．Evaluation of the Basic Remote Backup and Replication Methods for High Availability Database[J]．Software：Practice and Experience，1999，29(15)：1339-1353.

作者简介：万瑾慧（1982-），男，湖南衡阳人，硕士，研究方向：网络安全技术及应用。

收稿日期：2007-08-14