CDP 技术在话音漫游清算容灾系统中的应用①

陈 勇

(中国移动(深圳)有限公司, 深圳 518048)

摘 要: 随着话音漫游业务的快速发展, 话音漫游清算涉及的领域不断扩大, 在重大事故, 自然灾害等突发事件发生时确保清算数据的可用性和业务连续性就显得尤为重要. 针对话音漫游清算系统的容灾需求以及有待改进的问题, 通过研究分析 CDP(Continuous Data Protection)容灾技术, 提出将 CDP 技术运用到话音漫游清算容灾系统的方法, 设计并构建了一套具有持续数据保护功能和全面数据恢复能力的容灾系统. 实践证明, 该方法能够在各种灾难场景中有效地保障话音漫游清算系统的数据完整性和业务连续性.

关键词: CDP; 话音漫游清算; 容灾系统; 数据保护

Application of CDP in Disaster Tolerant System for Voice Call Roaming Clearing

CHEN Yong

(China Mobile (ShenZhen) Limited, Shenzhen 518048, China)

Abstract: With the rapid development of voice call roaming, expanding in the areas of clearing, ensuring data availability and business continuity at the time of major accidents, natural disasters and other emergencies occurring is particularly important. According to the requirement and the problem needed to be improved of voice call roaming system, this paper thoroughly researches and analyses CDP disaster tolerant technology, proposes the method of applying the CDP technology to disaster tolerant system for voice call roaming clearing, designs and constructs the disaster tolerant system for Voice Call Roaming Clearing with continuous data protection and overalls data recovery. Practice has proved that this method can effectively protect data integrity and business continuity for voice roaming clearing systems in a variety of disaster scenarios.

Key words: CDP; voice call roaming clearing; disaster tolerant system; data protection

随着信息系统逐步成为各企业生产运行的核心, 数据的完整性和服务的连续性已成为保障生产的关键[1]. 目前, 各电信运营商都在加强对生产系统的容灾建设, 以求实现在各类灾难发生时能够快速恢复生产, 保障业务的连续性. 话音漫游清算系统原先采用传统的备份方式, 不仅备份恢复时间长, 且备份数据是否可用要通过恢复测试来验证. 由于对硬件环境要求较高, 这类验证常常是难以进行的. 因此, 如何利用现有容灾技术保护话音漫游清算系统, 在重大灾害或事故发生时确保清算业务的不间断成为一项重要的研究课题.

1 容灾系统介绍

容灾系统是指建立多套同样功能的 IT 系统, 这些系统往往建在相距甚远的两地. 当业务运行的系统遇到自然灾害、重大故障和人为错误等灾难时, 整个业务可以被切换到另一个系统运行, 从而保证业务不会长时间中断.

容灾系统的好坏主要通过恢复时间的长短和数据丢失量来衡量.

恢复时间目标, RTO(Recovery Time Objective), 反映的是恢复业务的及时性, 表示业务系统可以接受的服务停止的最长时间, 是衡量服务丢失的指标.

① 收稿时间:2015-07-09;收到修改稿时间:2015-10-09

数据恢复点目标, RPO(Recovery Point Objective), 反映的是恢复数据的完整性, 指业务系统可以接受的最大数据丢失量, 是衡量数据丢失的指标[2-3].

2 CDP数据保护技术研究

CDP 是 Continuous Data Protection 的缩写, 它将目标数据发生的所有变化记录和保存下来, 以实现在任何历史时间点上的数据恢复. 通过对目前出现的 CDP 产品分析, 基于数据块的 CDP 和基于文件的 CDP 是目前 CDP 技术的两种主流的实现模式[4-5]. 而基于磁盘的 CDP 技术使用更为广泛, 其采用的关键技术包括:

① MicroScan 精简复制技术

MicroScan 技术是一种精简复制技术, 它将最小传输数据单元缩小为 512 Byte, 而不是块定义的 4KB, 从而在远程镜像中节省了很大一部分网络传输带宽.

② 存储虚拟化技术

以完全开放的虚拟化平台为核心, 将物理存储设备进行逻辑化统一管理, 对主机操作系统屏蔽存储设备硬件的特殊性, 从而实现了存储整合和集中管理等功能.

③ 多时间点自动快照技术

多时间点快照技术能够记录业务系统数据的各时间点版本, 并将时间间隔很短的各数据版本保存下来[6]. 这些连续快照可大大地降低容灾系统的 RPO.

④ DB Agent 技术

数据库使用内存缓冲机制提高交易的性能, 数据在内存中处理后才会写入到磁盘. 一般的远程镜像技术是直接复制磁盘上的数据, 这将导致数据库由于数据一致性问题不能即刻打开, 从而影响系统的快速恢复. DB Agent 技术可以在指定的复制点和快照点对数据库副本进行校验, 确保数据库副本中的数据文件与日志文件的一致性, 从而能够快速完成数据库的启动和打开.

⑤ 读写优化技术

CDP 使用 HotZone 技术和 SafeCache 技术全面优化所管磁盘的读写性能.

HotZone 技术用来优化读性能. 该技术根据数据的读写频度定义热点区和非热点区, 并将热点区的数据块分配到高速磁盘中. 如果监控到一些数据块不再有频繁的读写, 则将其定义为非热点区, 将该区的数据块移到低速磁盘上.

SafeCache 技术则用来优化写性能. 该技术使用快速磁盘作为缓存区, 生产数据顺序写入到缓存区, 写入性能高于随机写方式, 然后按照 SafeCache 设置的策略, 将 Cache 中的数据再随机写入到后端存储中[7].

3 话音漫游清算系统CDP容灾方案的设计与实现

3.1 话音漫游清算系统现状分析

目前, 话音漫游清算系统的在线数据量已达到几十 TB, 且清算业务线不断增多, 数据量仍然逐年增长. 清算数据涉及到用户漫游位置和漫游资费都是高度敏感信息. 如何将 CDP 技术应用到话音漫游清算系统的容灾系统建设中, 以提高容灾系统的数据保护能力和容灾切换速度, 是一个迫切需要解决好的问题.

该系统的数据库和应用服务器采用的是 HP 安腾系列主机, 操作系统为 HP-UX 11.31. 生产中心的存储设备型号是 HP XP24000, 数据库使用的是 Oracle 11g Rac. 考虑到该系统现有的软硬件设备仅使用了不到两年, 在系统整体容灾设计中必须尽量利旧原有的主机、存储设备和数据库.

3.2 设计原则与目标

根据对系统重要性和现状分析, 话音漫游清算系统的容灾方案需要达到可防范各类灾难的目标, 做到没有防御死角. 容灾系统的设计原则制定为:

① 在各类灾难中, 中断的业务都能在最短的时间内恢复, 并且保证数据的最少丢失, 站点级别灾难的 RPO 和 RTO 不能高于 10 分钟.

② 构建一个通用、开放、与应用系统松耦合的容灾系统, 且系统应具备良好的功能扩展性、规模扩展性以及灵活的资源管理和应用管理扩展能力.

③ 非站点级事故无需切换到灾备中心.

④ 异地灾备的数据传输要在现有生产网络有限的带宽环境下进行.

⑤ 确保容灾数据的一致性和可用性.

⑥ 容灾方案的实施不能影响生产系统的运行连续性和稳定性, 也不需要改变原有应用结构.

⑦ 为了充分保护现有投资, 容灾方案必须能够兼容各种主流品牌的存储设备、主机设备和操作系统.

根据以上设计原则, 本文设计的数据保护和容灾目标可具体分解为:

① 实时镜像备份, 无备份窗口

容灾方案要克服传统备份方式的备份窗口过长和无法进行实时备份的缺陷.

② 备份数据立即可用

利用与原数据完全相同的实时镜像, 而非经过转换成备份格式的数据, 使得恢复的数据直接可用.

③ 多点快照, IO 级的历史数据恢复

灾备系统要实现既可在灾难发生时保护最新数据, 又可在历史数据丢失时恢复原有数据, 就需要实现多时间点快照, 且要实现颗粒度小到 IO 级的历史数据恢复, 以应对各种逻辑错误.

④ 针对不同层级故障的快速恢复能力

话音漫游清算系统的主机采用了冗余配置, 但是存储是单点, 所以在容灾的基础上, 还要实现对存储的冗余备份. 既可为避免数据丢失多提供一层保障, 也可在生产中心实现非站点级灾难的快速恢复.

⑤ 精简带宽的复制

传输时要做到精简带宽, 以确保在现有的网络环境完成正确的数据传输.

⑥ 一致性保证

保证灾备端的数据可用性, 灾备端能够快速启动并打开数据库或挂载文件系统以顺利恢复业务.

⑦ 实现方便灵活的容灾管理

实现管理集中化, 给维护人员提供友好的管理界面, 使用灵活的流程管理、事件管理以及故障告警管理来简化维护工作, 同时避免人为误操作.

3.3 话音漫游清算系统容灾架构设计与实现

3.3.1 容灾架构的设计

按照本文提出的设计原则和设计目标, 将话音清算系统的容灾架构设计如下.

该架构设计的核心是在生产中心和灾备中心分别部署一套 CDP 设备保护核心生产主机和容灾主机的数据. 保护数据的方法是先将生产存储上的数据通过镜像同步到 CDP 设备下挂的存储中, 生成生产数据的一个实时备份, 即图 1 中 A->B 的同步镜像过程. 生产中心 CDP 设备与灾备中心 CDP 设备通过以太网络连接. 生产中心 CDP 设备将其下挂存储中从生产中心镜像过来的数据远程复制到灾备中心 CDP 设备下挂的存储中, 即图 1 中 B->C 的异步复制过程, 从而实现远程数据级容灾.

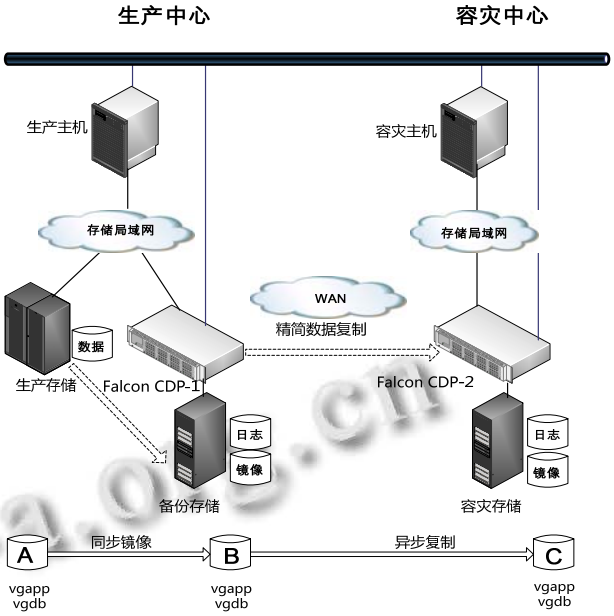

图 1　话音漫游清算系统 CDP 容灾架构示意图

3.3.2 容灾系统的实现

该架构的实现主要包括对生产中心的改造, 容灾中心的部署和两中心之间的数据同步.

① 生产中心的改造和容灾中心的部署

在原有话音漫游清算系统生产中心的存储局域网中, 部署一台存储设备作为原生产存储的备份存储. 考虑到成本因素, 备份存储可以采用比生产存储更为廉价的 HP 3PAR 磁盘阵列. 该存储通过光纤线连接到新部署的 CDP 设备, CDP 设备将备份存储经过虚拟化封装, 通过存储局域网挂载到生产主机. 生产存储上的数据通过操作系统或数据库的卷管理软件同步镜像到备份存储, 从而使备份存储上的数据与生产存储上的数据实时保持一致, 起到旁路备份的作用. 这种平滑的接入方式不需要对原有的生产存储局域网架构进行推倒重构, 可以在不中断业务系统的情况下进行改造. 将 CDP 设备下挂存储规划为不同的存储区, 分别使用超高速 SSD 固态盘作为 CDP 写入缓存和 I/O 日志区, 高速的 FC 光纤通道盘作为日志和核心业务数据库盘, 普通 SATA 硬盘作为应用和系统数据盘.

在容灾中心部署一套与生产中心结构类似的存储局域网. 在存储局域网中部署容灾 CDP 设备, CDP 设备连接新部署的一台 HP XP12000 磁盘阵列, 并将该存储进行虚拟化封装后通过存储局域网挂载到容灾主机.

整个部署实施分为以下几个大的步骤:

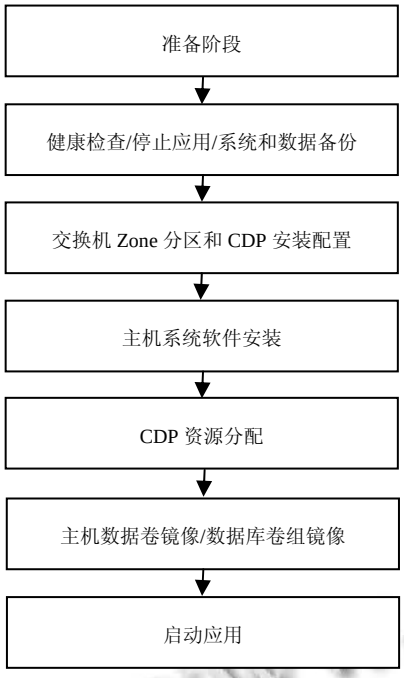

图 2 容灾部署实施步骤

② 规划 FC 交换机的 zoning 设置

CDP 的数据访问 FC 端口需要与 HP 主机的相对应访问端口建立一个专用的 Zone. 需要提前规划 zone 名称以及对应端口.

CDP 管理器的两个 FC 端口编号分别为 Qlogic100、Qlogic 101. 本次项目采用 Qlogic100 连接前端 HP 主机, 所以 Qlogic100 与 HP 主机的第一块 HBA 卡在一个 ZONE 中. 即:

CDPZone1: (HP1,CDP)

CDPZone2: (HP2,CDP)

图 3 FC 交换机端口分配示意图

③ CDP 资源分配

首先在目标主机上安装好 CDP ipstordisk rte 软件, 然后 CDP 上为主机准备容量, 并分配给应用主机.

在 CDP console 上为应用主机切分容量, 对应的每一个存储单元被称为 SAN resource. 所有 SAN resource 的切分容量都要略大于原生产盘的存储容量, 保证能完全存入原磁盘的所有数据, 以免发生溢出.

首先对 SAN 客户端进行如下的操作:

--在 Console 上创建应用主机客户端

--修改客户端属性和光纤通道的属性

然后, 设定主机光纤卡与 CDP 光纤端口的映射关系.

最后, 按照应用主机对应物理卷的大小和数量创建 SAN resource, 并分配给应用主机的客户端. SAN resource 形式如下:

vgcbs1_disk1

vgipp_disk1

以上分配的 CDP 资源必须在双机集群的备份主机上再次设置一次, 确保一旦主机发生双机切换, 能够使用相同的 CDP 资源.

④ 文件系统和数据库磁盘镜像的实现

1) 文件系统的磁盘镜像

首先, 在生产主机的 HP-UX 操作系统中识别 CDP 分配的虚拟化磁盘:

ioscan -fnC disk

```
H/W Path  Driver S/W State   H/W Type     Description
==========================================================
0/0/0/2/0/0/0.0.0           sdisk   CLAIMED     DEVICE       HP       LOGICAL VOLU
        /dev/dsk/c0t0d0   /dev/dsk/c0t0d0s3   /dev/rdsk/c0t0d0s2
        /dev/dsk/c0t0d0s1   /dev/rdsk/c0t0d0   /dev/rdsk/c0t0d0s3
        /dev/dsk/c0t0d0s2   /dev/rdsk/c0t0d0s1
0/0/0/7/0/0/0.1.220.0.0.0.0 sdisk   CLAIMED     DEVICE       FALCON   IPSTOR DISK
        /dev/dsk/c10t0d0   /dev/rdsk/c10t0d0
0/0/0/7/0/0/0.1.220.0.0.0.1 sdisk   CLAIMED     DEVICE       FALCON   IPSTOR DISK
        /dev/dsk/c10t0d1   /dev/rdsk/c10t0d1
0/0/0/7/0/0/0.1.220.0.0.0.2 sdisk   CLAIMED     DEVICE       FALCON   IPSTOR DISK
        /dev/dsk/c10t0d2   /dev/rdsk/c10t0d2
0/0/0/7/0/0/0.1.239.0.0.0.0 sdisk   CLAIMED     DEVICE       HP       OPEN-V
        /dev/dsk/c6t0d0   /dev/rdsk/c6t0d0
```

图 4 操作系统识别出的 CDP 磁盘

图中输出结果中带有“FALCON IPSTOR DISK”字样的为 CDP 虚化生成的磁盘.

用 insf -e 命令将创建 CDP 磁盘的设备文件.

然后, 将 CDP 分配的磁盘创建为物理卷.

pvcreate /dev/rdsk/c11t0d0

最后, 将原有卷组中的逻辑卷镜像到 CDP 分配的虚拟磁盘上.

lvextend –m 1 /dev/vgap/lvap /dev/dsk/c10t0d0

2) 数据库的磁盘镜像

由于 Oracle 11g Rac 采用 ASM 进行磁盘管理, 所以不能直接通过 HP-UX 操作系统进行磁盘镜像, 需要通过对 ASM 中的 diskgroup 创建 failgroup, 从而配置 ASM mirror, 主要步骤和命令如下.

增加 diskgroup 的 failgroup, 并在 failgroup 加入 CDP 虚拟化磁盘:

SQL> CREATE DISKGROUP mirrordg NORMAL

```
REDUNDANCY
    FAILGROUP control_1 DISK
      '/dev/rdsk/c6t0d0', '/dev/rdsk/c6t0d1'
    FAILGROUP failure_group_2 DISK
      '/dev/rdsk/c10t0d0', /dev/rdsk/c10t0d1' ;
```

由于原有主磁盘阵列为性能更高的 HP xp24000 磁盘阵列, 将其数据库读优先级设置为最高:

```
SQL> alter system set ASM_PREFERRED_READ_FAILURE_GROUPS ='MIRRORDG.CONTROL_1';
```

如果有多个 diskgroup, 可以设置

```
SQL> alter system set ASM_PREFERRED_READ_FAILURE_GROUPS='DATA.DFG1',' FRA.FFG1';
```

设置读优先级别之后, 在主磁盘无故障的情况下, 数据库就只从主磁盘读取数据.

⑤ 两中心之间的数据同步

生产中心与灾备中心的 CDP 设备通过 TCP/IP 网络连接, 并开启从生产中心到灾备中心的远程复制. 由于两中心距离较远, 为了防止由于传输时延对生产系统处理性能造成影响, 远程复制使用异步复制模式. 同时, 采用压缩传输模式以节省带宽, 打开数据加密模式以保证数据传输过程中的安全.

复制策略如下.

持续时间间隔: 每 10 分钟复制一次.

容量的变化量: 新数据超过 5MB 就开始复制.

3.4 灾难恢复的实现

使用以上容灾方案可实现话音漫游清算系统的各类灾难恢复, 具体恢复方法如下.

① 数据库表级别的丢失或损坏

面对数据库表级别的灾难, 可在主机上挂载 CDP 提供的历史快照, 该快照包含完整的记录条目, 然后再使用数据库命令将丢失或损坏的记录导入到生产数据库中.

② 文件丢失或损坏

如果应用系统中的文件损坏或丢失, 首先提取出没有发生数据丢失的时间点的快照, 将其分配给主机, 然后主机扫描快照盘并恢复出丢失的文件. 在特别紧急或者丢失的文件数目较多的情形下, 还可以直接将快照磁盘加入生产系统临时接管生产应用.

③ 文件系统不能挂载和数据库不能启动

当生产系统的文件系统出现故障无法挂载或数据库崩溃不能启动时, 可以将 CDP 的历史快照提取出来并分配给对应的主机. 主机用这些磁盘来挂载文件系统或启动数据库.

④ 生产存储发生故障

当生产中心主存储的磁盘出现错误无法工作时, 生产系统可以自动使用 CDP 网关下挂的冗余备份存储中的镜像盘运行业务, 业务不会出现任何中断迹象, 实现 RPO=0 和 RTO=0 的理想状态. 待主存储的故障磁盘修复好以后, 新数据会自动从镜像盘重新同步到修复好的磁盘上.

⑤ 站点级灾难

当在生产中心发生站点级别的灾难时, 可通过紧急启用灾备中心的容灾系统来接管业务运行. 灾备中心的 CDP 把从生产中心复制过来的容灾数据分配给灾备中心的主机, 即可启动灾备端应用和数据库进而恢复业务, 也为生产中心的系统恢复赢得了足够的时间. 业务在灾备中心运行所产生的新数据可以在生产中心系统恢复后通过灾备中心 CDP 反向复制回生产中心, 待两中心数据完全同步后便可在适当的时候将业务系统回切回生产中心.

4 结语

本文通过将 CDP 技术应用到话音清算容灾系统, 实现了本地的应用系统保护和恢复以及更可靠的异地容灾系统, 获得了备份和容灾的双重效果. 该容灾方案可以充分利旧原有存储、主机、协议及软件, 且带宽占用率低. 从而大大节省了企业的成本. 同时, 本文的 CDP 容灾架构是一次构建成的开放架构, 在基础架构不需要做任何改动的情况下即可完成应用和设备的扩展, 进而满足未来业务扩展的需求.

参考文献

1 刘栋,夏清国,张砚耕.一种远程容灾系统的设计与实现.计算机与现代化,2012,(2):149−152.

2 周强,赵海峰,纪允.基于 CDP 技术的地市级烟草公司通用型灾备系统方案.计算机安全,2012,(10):79−84.

3 溪利亚,顾兵.CDP 技术在数据容灾系统中的应用.计算机与现代化,2011,(10):37−39.

4 陈鹏,杨频,赵奎.远程容灾系统的设计与实现.计算机工程与设计,2011,(10):3247−3250.

5 谢舟,金政哲.基于 CDP 的数据容灾系统设计与实现.广州大学学报:自然科学版,2012,(5):78−81.

6 彭晔,刘晓垒,杜敏.CDP 异地灾备系统的研究与实现.中国管理信息化,2013,(21):49−51.

7 杨栋.CDP 技术在医院灾备系统中的应用.中国医疗设备,2011,(7):68−70.

8 纪芳.CDP 在电力容灾系统中的应用.东北电力大学学报,2010,30(6):95−98.