中国银行澳门分行灾备建设简介

龚存
中国银行澳门分行
信息科技部
系统运维团队
Email: gong_cun@bocmacau.com

Abstract—银行信息系统的安全运行是金融业稳定运行和长远发展的基础，并关系到民生和社会稳定，做好灾难备份体系建设是银行业信息安全的重要保障和关键环节。本文主要介绍了中国银行澳门分行在金融业务多元化发展、IT 应用环境日趋复杂的情况下，如何通过技术和架构的演进，将灾备建设从业务种类单一的灾备外包服务，发展成涵盖本地所有重要系统的云数据中心。

1 背景

中国银行澳门分行（以下简称澳门分行）是澳门特区的发钞行、政府公库代理行、人民币业务清算行，主流业务约占本地市场份额的 40%，因此澳门分行的灾备建设不仅保障了本行业务的可持续性，对特区经济的稳定和信息化建设都有重要的参考意义。

澳门分行自 2009 年开始启动开放平台灾备建设，其中历经了 2009 ~ 2016 年灾备恢复外包，2016 ~ 2017 年灾备机房自建和部分灾备系统迁移，以及 2017 ~ 2018 年大规模灾备系统建设等三个阶段，以下将详细介绍这三个阶段的灾备架构和技术演进。

2 技术演进

在灾备恢复方面，RTO 和 RPO 是业界公认的衡量灾备系统的两个重要指标 [1]：

- RTO（Recovery Time Objective）即恢复时间目标，指灾难发生后，信息系统从停顿到必须恢复的时间要求。
- RPO（Recovery Point Objective）即恢复点目标，指灾难发生后，数据必须恢复到的时间点要求。

RTO 和 RPO 实际反映的是业务连续性目标和数据一致性目标，澳门分行各个阶段的灾备架构和技术演进，都围绕这两个目标展开。

2.1 灾备外包服务（2009~2016）

澳门分行 2009 年 7 月开始启动开放平台灾备建设，与 IBM 签署灾备外包服务，由 IBM 提供灾备场所、硬件设施、通讯线路等资源及服务（Business Continuance Recovery Service, BCRS）。灾备中心位于澳门凼仔的澳门电讯（以下简称 CTM）机房 [1]。灾备中心成立之初只覆盖柜员终端、网上银行等基本交易系统，之后又陆续增加了人民币清算系统、证券系统等重要业务系统。

由于澳门分行主中心和备份中心之间的专线带宽只有 100 Mbit/s，因此主中心和备份中心之间的存储复制采用了 Global Mirror with Change Volumes (GM/CV) 技术，GM/CV 是一种异步存储复制技术，源存储通过 CV 周期性地复制增量数据，而目标存储通过 CV 为上一个周期保留完整增量数据。假如灾难发生导致存储复制中断，且数据不一致，目标存储将可以通过 CV 回滚到最后一次检查点，以保证数据一致性。因此理论上 GM/CV 的 RPO $\leq 2 \times$ Cycle，由于带宽受限，周期 Cycle（默认为 10 分钟）实际是一个变值，假如数据量较大（如夜晚批处理作业时段），一次 Cycle 可能需要数小时，这样 RPO 就有可能达不到业务及监管要求。

由于灾备外包服务签署之前，澳门中银已经在电服中心（另一独立场所）设立了网络备份节点，该节点上联亚太信息中心，下联各支行，因此灾备的网络仍然使用该备份节点，形成了一个较复杂的网络环境：

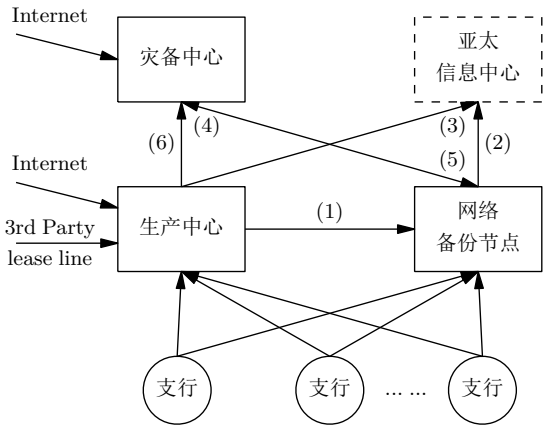


图 1: BCRS 网络架构

其中对应的连接含义为：

标记	含义
(1)	生产线路切换到备份线路
(2)	从备份线路连接亚太信息中心
(3)	从生产线路连接亚太信息中心
(4)	发生灾备时，支行网络通过备份线路连接灾备中心
(5)	发生灾备时，灾备中心通过备份线路连接亚太信息中心
(6)	生产和灾备中心通过 FCIP 协议进行存储复制

1. 由 IBM 向 CTM 租用

由于整体架构非常复杂，造成每次切换演练的难度很大，难度体现为：

1) 系统切换后有大量手工配置要进行。为了缩短 RTO 时间，我们尽量将所有数据包括操作系统都通过 GM/CV 远程复制到灾备中心，但要让操作系统在不同型号和配置的硬件上启动，需要做很多适配，甚至有一些老版本操作系统不支持在不同设备上直接启动，需要做一些特殊处理。
2) 完成系统启动后，由于 RPO $\neq 0$，应用还要进行数据和交易一致性检查，修复应用启动中的种种问题。
3) 由于网络环境复杂，生产系统切换到灾备中心后，网络需要做大量调整，如地址转换、防火墙开通等。

由于切换难度大，不仅造成了 RTO 较长，而且在演练回退时也带来一定的风险，如存储复制的方向，路由和防火墙要重新调整等。

2.2 灾备机房自建（2016~2017）

随着加入灾备的重要系统不断增加，IBM 所提供的灾备场所无论服务器还是实际操作的空间都已经不能满足要求，澳门中银于 2016 年启动了北安灾备机房自建工程，将位于凼仔的北安仓库改建为灾备机房（以下简称北安灾备中心）。是次自建的目标是：

1) 为业务系统进行可用性分级，将本地重要灾备系统新建或迁移到北安灾备中心。通过业务恢复需求评估，对于缺失 DR 能力的系统，到底多大程度地影响本地重要业务的连续性，并在北安灾备中心建设过程中弥补缺失的环节。
2) 重新进行架构设计，简化灾备操作步骤，缩小 RPO、RTO 时间。在满足灾备需求的基础上，尽量复用资源，提高投资回报率。

2.2.1 业务影响分析

业务影响分析（Business Impact Analysis, BIA），包括：

1) 业务可用性分级
   澳门分行通过合规部牵头，统筹各业务部门评定业务系统可用性等级，科技部对于分级过程及结果给予技术评估和反馈意见。
2) 业务关联分析
   根据业务可用性分级，柜员终端、网上银行、证券投资、支付清算、移动应用等系统具有较高的优先级，但实际上，有些系统可能会在业务可用性分级中遗漏，但却和重要业务系统之间有很大的依赖关系，从而影响到重要业务系统的可用性。这就需要通过业务关联分析进行补充。举例来说，人民币清算行系统（以下简称 CCS）的应用关联分析如图 2 所示。

从上图可以看出，一个复合系统由于牵涉多个系统，系统间交互多且复杂，有些非业务系统如企业总线、文件传输等系统，由于对用户透明，但如果缺失则可能引起灾备失效。因此，业务关联性分析必须由业务和科技部门共同参与，针对典型场景和重要交易，分析清楚每一笔交易从业务发起到结束所经过的所有环节。

经过以上业务影响分析，澳门分行重要业务系统及其支撑系统共 24 个。可以看出，仅一个重要业务系统（人民币清算行系统），就有如此复杂的关联关系，旧的架构方案可能由于复杂度而失效，因此必须进行灾备架构的重新设计。

2.2.2 架构设计

如前所述，新架构的目标是减少灾备方案复杂度，缩小 RPO、RTO 时间，同时在满足灾备的基础上，尽量复用资源，提高投资回报率。为实现这个目标，最自然的方案就是通过在数据链路层打通网络，将生产中心的网络直接延伸到灾备中心，这样网络架构的简化不仅将减少 RTO 时间，同时对应用架构和交互关系的影响也会减至最低。同时，利用两中心之间的高带宽线路，做同步存储复制将 RPO 减至最低。

我们通过以下步骤来实施以上方案：

1) 租用多条宽带专线，通过 密集波分复用（*Dense wavelength division multiplexing* , DWDM）技术，提高带宽传输率。DWDM 通过放大光信号来提高光纤承载量，从而提高带宽传输率。澳门分行通过 DWDM 承载 2 条带宽为 10gib/s 的网络光纤和 4 条带宽为 8gib/s 的存储光纤：

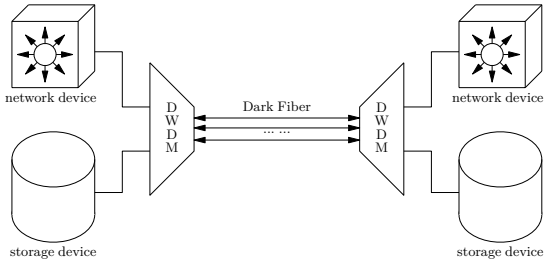


图 3: DWDM 架构

   生产和北安灾备中心的地理距离约 3km，线路距离约 10km，在 DWDM 上测得的两端往返延迟（round-trip times, RTT）$\leq$ 0.1ms。
2) 实时存储复制：当生产和灾备中心实现高带宽低损耗传输后，对于重要系统，使用实施存储复制（Metro Mirror, MM）代替异步存储复制，使得 RPO 降为 0。由于同步存储复制的写操作（Write）需要异地存储的 Ack 同步确认，因此同步存储复制对性能会产生一定的延迟影响（在“效能”章节将分析影响）：

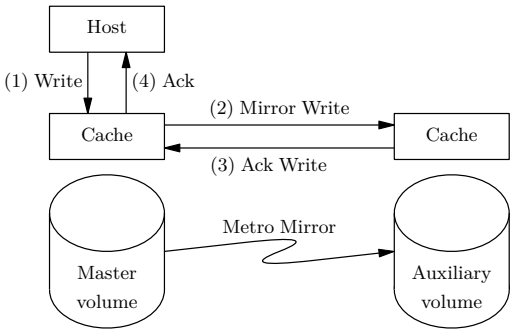
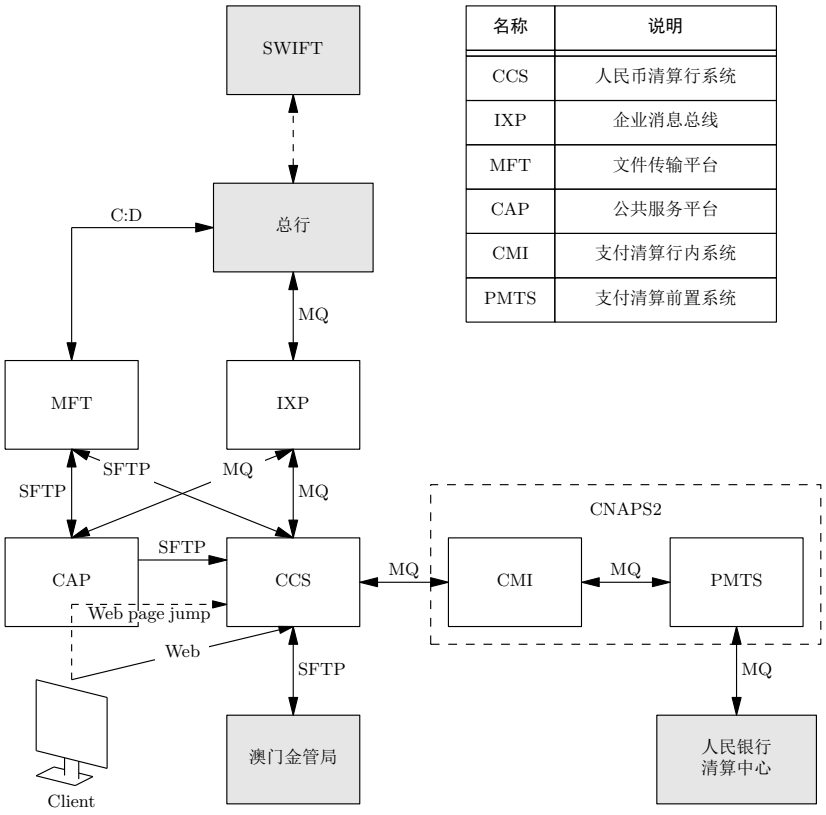


图 4: 实时存储复制

3) 通过 Cisco 迭加传输虚拟化（Overlay Transport Virtualization, OTV）技术在数据链路层（Data Link）打通生产



名称	说明
CCS	人民币清算行系统
IXP	企业消息总线
MFT	文件传输平台
CAP	公共服务平台
CMI	支付清算行内系统
PMTS	支付清算前置系统

图 2: CCS 应用关系图

和灾备中心网络连接。OTV 是一项“MAC in IP”技术，通过使用 MAC 地址路由规则，提供一种迭加（overlay）网络，能够在 IP 层建立二层连接的虚拟隧道：

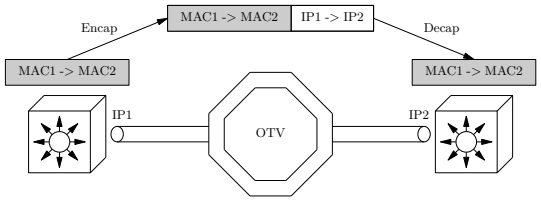


图 5: OTV 架构

实现网络二层打通之后，针对灾备的网络配置如 IP 路由等大大简化，也为种种依赖于二层网络连接的集群应用或虚拟化技术等打下了基础。

4) 通过二层和三层网络连接组合实现部分组件的双活。并不是任何类型的网络都必须打通二层连接，比如面向互联网服务请求的 Web 服务器。由于前期应用系统已经将处理互联网请求的 Web 服务器、处理业务逻辑的应用服务器以及进行数据处理的数据库服务器等进行了合理分层，因此我们通过在灾备中心新增面向互联网的 Web 服务器，并和生产中心的应用服务器进行三层连接，通过全局负载均衡器派发请求，就已经在 Web 层达到了生产和灾备多活。全局负载均衡器（Global Traffic Manager, GTM）通过向已知页面发送 HTTP/HTTPS 请求来监控是否正常工作，并动态派发 DNS 和 Web 请求。如下图所示，当正常运行时，位于灾备中心的 Web 服务器也能处理应用请求：

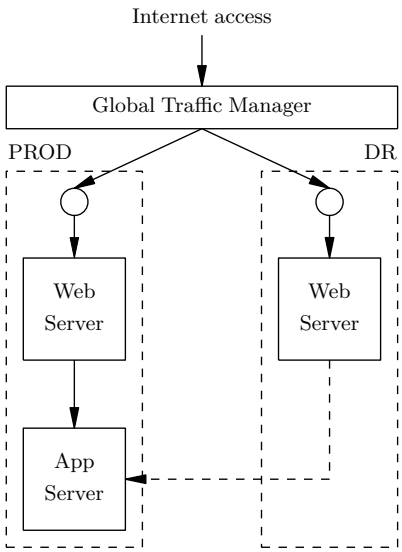


图 6: GTM(Normal)

当启动灾备时，应用服务器将通过存储复制在灾备中心运行，全部应用交易将发生在灾备中心。

2.2.3 效能

根据以上架构，效能的考虑主要在于生产中心和灾备中心之间存储复制或网络传输对应用造成的延迟影响。图 7 是通过不同的块大小（block size）写 2GB 文件，远程同步复制下的磁盘 I/O 和本地磁盘 I/O 的对比。可以发现，随着 block size 的增大，传输速率也随着增大，虽然两者的差距也随之增大（远程复制要比本地慢），但并不是一个严重的问题，这是由于 $传输速率 = 文件大$

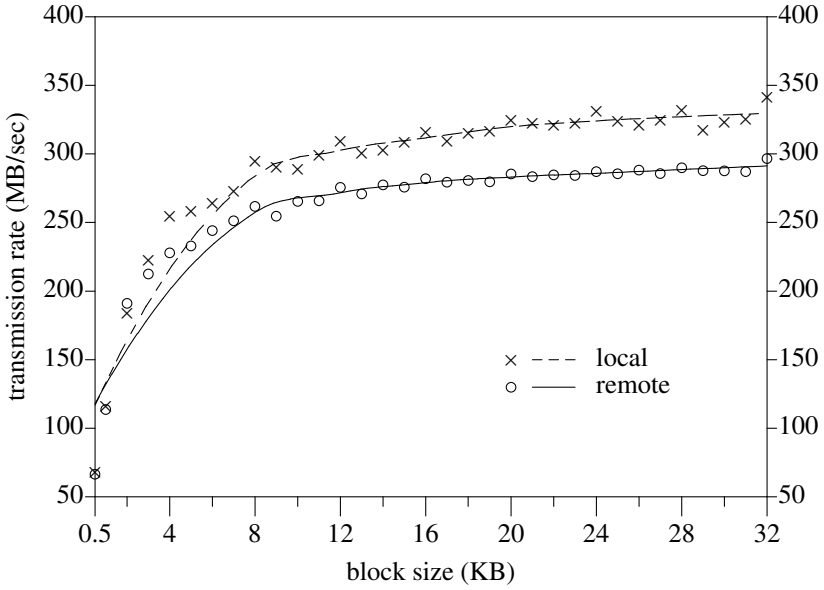


图 7: 远程复制下的磁盘写速率和本地磁盘对比

小/传输时间, 因此在文件大小固定的情况下，传输速率的大小是由传输时间决定的，而传输时间又是由存储复制的延迟和次数决定的。

由图 8 可以看出，在传输窗口之内，传输时间是由首比特的网络传输时延和传送整个文件的主机处理时延组成。因此当文件大小固定（2GB），block size 越大，传输的次数越少，网络传输的滞后效应就越明显，反之则滞后效应越不明显 [2]。对于大部分应用来说，数据传输可以抽象为没有起始也没有结束的字节流，首比特的网络传输时延可以忽略，窗口大小和拥塞重传就成了影响滞后的关键 [3]。因此，除了采用高带宽低损耗传输减少拥塞重传外，还可以通过调大网络或光纤传输的滑动窗口来补偿滞后[2]：

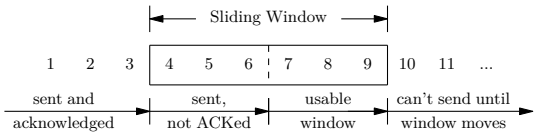


图 9: 滑动窗口

从上图可以看出，只要在滑动窗口内接收到对方的确认报文（Ack），就可以将滑动窗口不断向右推进，从而形成一条不间断流水线，避免停等带来的时延。

2.3 云数据中心（2017~2018）

当实现了网络二层打通之后，就可以在生产和灾备中心之间使用各种依赖于二层网络连通的虚拟化技术，比如 IBM Live Partition Mobility (LPM) 技术或 VMware VMotion 技术，这两种技术都允许在共享二层网络和存储的物理服务器之间，动态迁移其上运行的虚拟服务器，并保持进程、内存及网络状态的一致。因此如果将远程同步复制的存储合并为一个虚拟的共享存储，并同时挂载在生产和灾备服务器上，就可以实现生产和灾备中心之间应用系统的动态迁移。通过 IBM SAN Volume Controller Stretched Cluster (SVC-SC) 技术，可以达到这一目的：

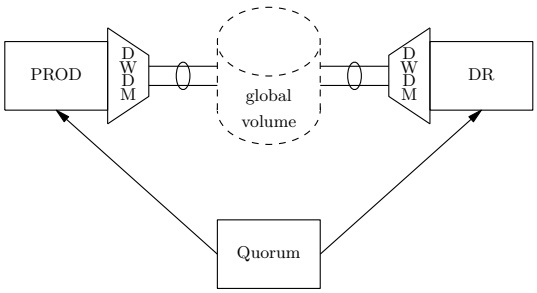


图 10: SVC Stretched Cluster

如上图所示，SVC-SC 将图 4 所示的显式复制的生产和灾备磁盘，变为隐式复制的镜像卷（Mirrored Volume），一对镜像卷作为 一个虚拟磁盘同时挂载给生产和灾备服务器，这种冗余性保证了无论生产还是灾备任何一个位置丢失了磁盘，另一边仍能访问磁盘。但是，仍然要注意防范当中间线路中断可能引起的“裂脑”（split-brain）风险 — 如果中间线路中断，而两边均能访问本地磁盘，就会造成数据不一致。因此，除了为中间线路租用多条不同的 ISP 专线之外，还要搭设仲裁节点[3]，当发生中间线路中断时，仲裁节点将决定在哪个场所能够访问磁盘，并阻断另一个场所对磁盘的访问。

SVC-SC 的实现意味着虚拟化的进一步提升，在原有网络打通的基础上，生产的某一服务器集群，可以动态迁移其中一部或几部到灾备中心运行，而不会对业务有任何影响，并且在任何时候，生产和灾备中心的硬件资源都可以动态调动，真正实现了“灾备 = 多活 + 高可用”的云数据中心。

2. 调大 TCP 的 *receive buffer* 或光纤信道的 *buffer credits* 。当然除了调大滑动窗口，也可以调大主机 CPU 以减少处理时延。

3. 新架构下将原灾备中心（CTM 机房）设置为仲裁节点。

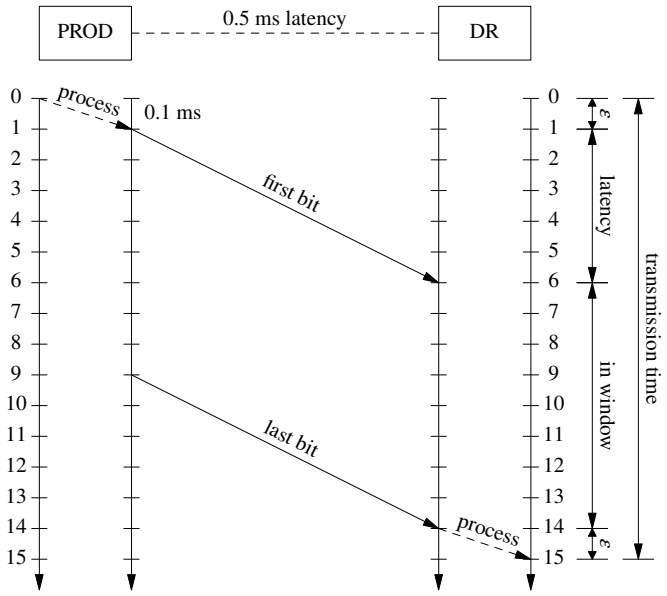


图 8: 传输时延

3 展望

从以上的技术演进可以看出，澳门分行在基础设施架构上做了大量工作，在应用开发基本无需改造的情况下，实现了多活的虚拟数据中心，但仍存在一些问题有待解决。

3.1 通过自动化简化灾备流程

由于 24 个灾备系统数量多，系统间依赖关系复杂，大量系统的切换不仅需要手工操作，在组织流程上也有很大难度。多活灾备中心在 2018 年底建设实施完成，当年 11 月份进行了完成建设后的第一次灾备演练，为了能够顺利实施切换验证，我们将整体工作步骤分解，并成立多级调度进行指挥协调，其中一级调度负责总体指挥，二级调度负责统筹各团队的具体实施内容、并对切换中可能遇到的问题进行组织决策（图 11）。

通过实际演练，证明了多级调度和问题管理机制，对于大规模场景下的的灾备演练是非常有效的，但是大量的手工操作，特别是应用启停和对依赖关系的判断，都非常依靠特定技术人员，不仅影响了 RTO 指标，也影响实际切换的可操作性。因此，未来应该将应用间的关联关系抽象为以下有限自动机：

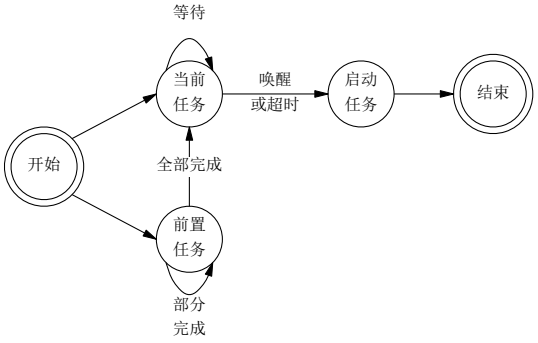


图 12: 应用启动状态图

我们可以将有限自动机翻译为以下代码，从而变为可行的自动化流程：

```
state = s1; { start }
while state in s1, s2, ..., sn do
  case state in
    s1:
      case action in
      a1:
        advance the action
        state = s2
      a2:
        ...
      end case
    s2:
      ...
  end case
end while
```

3.2 增强灾备监控和应急

由于新架构下的灾备中心实际成为了生产中心的延伸，因此在灾备中心的监控上需要在未来做进一步的加强，包括对跨中心的网络和存储状态的监控、对灾备中心的互联网入口的入侵检测等，并准备好相应的应急预案。值得注意的是，虽然 SVC-SC 跨域存储允许任何一个节点丢失对磁盘的访问，但是一旦丢失了任一节点，意味着图 4 中的缓存（Cache）缺失，对磁盘的写访问将变为 write-through mode，从而影响到磁盘 I/O 效能，因此未来应该考虑通过增加多个 I/O group，当其中某一 I/O group 的单节点失效时，可以动态将磁盘迁移到另一 I/O group。

3.3 向分布式架构发展

目前的灾备架构，虽然虚拟化程度已经达到了云数据中心的要求，但灾备切换方案还是基于传统的存储复制方案，没有充分利用分布式架构在容灾方面的天然优势，事实上我们应该在图 6 的基础上，将应用服务器做类似于 Web 服务器的分布式扩展，将数据库做主副拷贝，生产中心部署主数据库，灾备中心则是只读副本，发生灾备切换时赋予灾备中心写数据库权限 [4]：

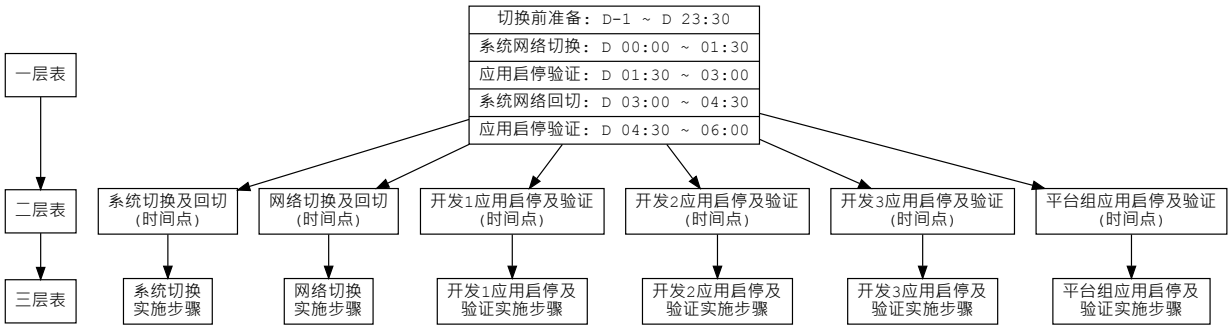


图 11: 灾备演练的组织架构

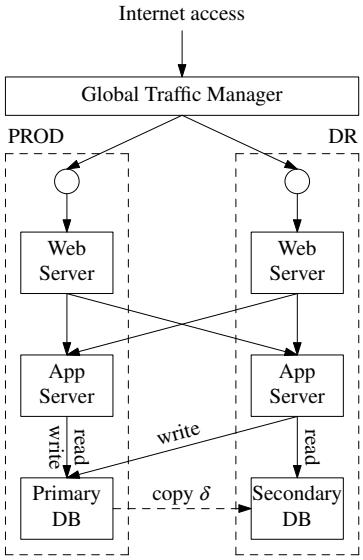


图 13: 分布式架构

通过分布式扩展，不断将生产和灾备中心解耦合、去存储复制（包括 SVC-SC），灾备的流程才能最简化，对性能的影响也会降至最低。分布式扩展也需要应用系统的改造，比如尽量将无状态（stateless）组件剥离前置，以允许整体的快速迭代、简化部署，通过 分散/聚集 I/O 、批量操作、应用层缓存等以减少来回、增加有效带宽等。总之，灾备系统向分布式系统转换需要系统思维，任何改动都需要由用户面临的或运营需要的问题来驱动并做全局设计，拘泥于局部的优化只能平添复杂性。可以预见，未来从灾备中心真正转型到云平台的过程中，架构设计和工程实践的华尔兹将一直跳下去。

References

[1] 中国人民银行 (PBC), 银行业信息系统灾难恢复管理规范, 中国人民银行发布, JR/T 0044—2008, 2008。

[2] W.Richard Stevens, *TCP/IP Illustrated, Volume 3*, 1st Edition, Addison-Wesley, 1996: 附录 A.3 详细讲述了迟延和带宽的关系。

[3] W.Richard Stevens, *TCP/IP Illustrated, Volume 1*, 1st Edition, Addison-Wesley, 1993: 第 20 章第 3 节详细讲述了 TCP 的滑动窗口机制。

[4] Rajesh Nishtala *et al.*, *Scaling Memcache at Facebook*, 10th USENIX NSDI’13, 2013: 介绍了 Facebook 分布式系统的工程实践，其中第 5 部分讲述了多数据中心部署的主副拷贝，以及如何缩短副本数据库滞后主数据库的时间。