驯服不规范数据，提升合规性

家银行机构总部位于加拿大多伦多，年收入约为 180 亿加元，他们发现自身面临一个大数据问题：整个组织内使用了十余年的多用途专有软件应用程序竟不再符合监管要求。该机构发现问题后，耗费 10 个月的时间进行评估和解决。

加拿大金融机构监管局 (OSFI) 和加拿大金融交易与报告分析中心 (FINTRAC) 制定了一套新的合规规范，旨在管理各种反洗钱 (AML) 和反恐怖主义融资 (ATF) 监管报告。在此之后，这个问题浮出水面。早在 2018 年，应用开发高级顾问 Bhavani Shankar Mudigonda 就发现银行未遵循新的标准。

一项内部调查显示，在现有的数据管理程序中，回收的客户识别码 (ID) 关联到多个存款和投资账户。因此，存在将客户信息暴露给未授权方并损害真实账户持有人隐私的风险。

MICK BRADY

自由技术传播员，拥有 20 多年的技术出版物编辑和写作经验。

对于为加拿大各地零售和商业客户（包括资本市场）提供全方位金融产品和财富管理功能的金融机构，这是个严峻的问题。为了避免潜在的法律、财务和声誉后果，这家银行希望迅速解决问题，保护客户账户并在进行任何外部审计之前满足合规。

评估挑战

第一步是确定错误的根本原因。对于从银行到电子通讯公司等零售和商业客户服务组织，创建唯一的 ID（即数字身份）是常规程序。这在追踪用途方面必不可少。在该银行，每次客户访问银行中心、进行数字交易或致电客户服务代表时，内部应用程序都会分配一个唯一的 ID。

“任何应用程序都会在数据库中占有一定的空间，”Mudigonda 指出。“因此，我们会先做一些内务处理，然后再重新调整一些现有帐号的用途。”银行系统在账户关闭或清除后回收客户 ID。

Mudigonda 还发现，在应用程序漫长的应用开发生命周期中，针对测试和其他目的创建了多个记录。结果是，银行除了有效数据外，还存在数千条虚拟记录。

由于即使重新调整唯一的客户 ID 用途，系统中仍然保留有虚拟记录，因此存在将同一帐号分配给多个客户的风险，导致客户 #1 能够查看客户 #2 的账户信息。

“该应用程序很大，记录了诸多客户信息。组织内部许多团队将该应用程序用于各种用途，包括对客户数据的各种类型操作。”Mudigonda 解释道。面临的挑战是既要保持真实数据的完整性，同时又要识别并清除虚拟记录，以保护客户隐私并完全符合当前的监管要求。

在了解问题的涉及范围之后，解决问题的核心团队随之组建。该团队的任务是调查可能的解决方案、确定最佳方法，并最终制定、实施和监督正在开展的工作。

制定解决方案策略

核心团队包括一名软件设计师，该设计师与其他团队成员多次召开头脑风暴会议，思考各种方法。该团队得出的结论是，对银行的内部应用程序应用补丁并且每晚多次运行清除作业就可以解决问题。

一名开发人员负责编码工作。另一名团队成员执行测试。Mudigonda 负责监督整个项目，并与银行的业务控制团队沟通，确保新功能能使应用程序合规。

有些业务规则可以编码到主应用程序中，将某项记录归类为虚拟记录，例如：

- 客户的名字、中间名、姓氏、街道地址、城市、省份或邮政编码由三个连续字符组成（如 AAA、111）。
- 客户的名字是非常通用的名字（如 John Doe）。
- 客户的名字、中间名、姓氏、街道地址、城市或省份的文本字符串为 "test" 或 "testing" 或 "test record"。
- 对加拿大邮政的实时地址验证 API 调用给予否定响应。

在头脑风暴会议期间提出的许多规则都遭到了否决。例如，对业务规则使用 OR 操作会导致合法记录被清除。“我们实施的解决方案使用了 AND 操作，因而非常稳健，”Mudigonda 解释道。

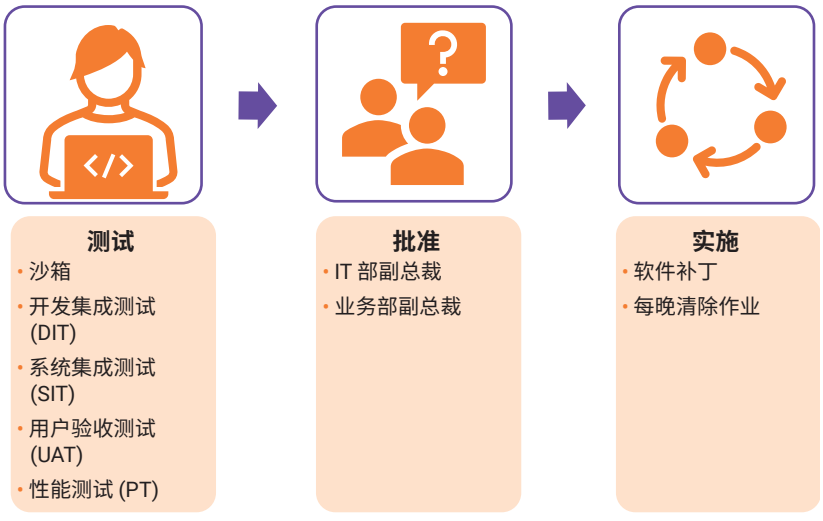
团队得出的结论是，对银行的内部应用程序应用补丁并且每晚多次运行清除作业就可以解决问题。

Mudigonda 表示，最终由负责内部审计的主题专家组成的内部控制团队制定了大约 24 条具体的业务规则。只有当某项记录满足规则中规定的全部 20 多项标准时，才会归类为虚拟记录。

除了将软件补丁编写到应用程序中，解决方案还要求每晚多次运行清除作业。为此，该团队可以重新使用所拥有的作业调度程序，这个程序在应用程序中运行许多其他作业。

Mudigonda 概述了该项目从沙箱到实施的开发过程（**图 1**）。

图 1
虚拟记录清除开发过程



测试、进一步测试和重新测试

为了与银行遵循完整软件开发生命周期的做法保持一致，在集成解决方案部署之前，针对其原型设计进行了多个阶段的测试。

沙箱阶段

该团队首先在本地机器（即沙箱机器）上开发代码，并进行了一轮单元测试。

“我自己做了一个测试，”Mudigonda 说。“我的系统没有连接到任何设备，只是一个独立的箱子。作为一名开发人员，我认为一切都很好，然后我将这些代码提升到 DIT 环境中。”

由于系统使用共享平台，银行能够自动实时向监管机构提供更正后的数据。

DIT 环境

DIT 环境增加了功能集成。当代码被提升到 DIT 环境时，开发人员能够使用其他集成系统测试补丁的功能。

由于存在大约 200 个依赖关系（约 100 个上游应用程序和 100 个下游应用程序），因此需要进行大量测试。“在 DIT 环境中，所有集成都是可用的，”Mudigonda 指出。“例如，我们的补丁是 ABC 应用程序，还有另一个名为 XYZ 的应用程序。在 DIT 环境中，ABC 连接到 XYZ。然而，在沙箱中，ABC 只是一台独立的机器。”

在相关系统集成并确信代码正常工作之后，软件补丁就会提升到下一个测试级别：SIT 环境。

SIT 环境

在 SIT 环境中，质量保证 (QA) 部门接管了测试过程。

“该部门提供专门的 QA 服务，因此会编写自己的测试案例，”Mudigonda 指出。“他们根据自己编写的测试案例，测试了部署在 SIT 环境中的那段代码。”

SIT 环境测试完成后，代码提升到 UAT 环境。

UAT 环境

在 UAT 环境中，部分用户（主要来自银行的业务方面）使用新应用程序。项目团队向抽取的部分用户提供了对该环境的访问权限，这些用户在其工作例行程序环境中测试了代码并签字。在 UAT 环境下通过审核后，该软件提升到 PT 环境。

PT 环境

在 PT 环境中，QA 部门的一个专门团队从性能角度重新测试了代码。测试人员考虑了一系列问题，例如中央处理器 (CPU) 是否因为代码而负担过重，或者代码是否占用太多内存。

获得高管批准

代码在每个测试环境中通过审核后，团队向 IT 和业务决策者介绍了拟议的解决方案，提请批准部署。

“我们必须得到 IT 部副总裁签字同意，”Mudigonda 回忆道。“我们必须先开发一个原型，然后向副总裁演示原型并获得他的认可。同时，我们还必须向业务部副总裁演示并得到他的认可，才能真正实施产品。”

这些高管批准了解决方案，然后进入到下一个项目阶段。

实施解决方案

到 2019 年秋季，业务规则编入到软件中，并且解决方案进入可用状态。该系统从与监管机构共享文件的登

录平台（文件夹）中获取 .CSV 文件，并运用规则。文件中满足所有规则的任何记录都归类为虚拟记录。然后所有虚拟记录合并到数据库的一个文件中，并输入到清除作业。清除过程将这些记录从核心应用程序移至历史应用程序。

“这项清除作业计划每晚运行五次，”Mudigonda 说道。“在大约 10 周的时间里，我们逐步清除了 95% 的虚拟记录，清理了系统并且使得我们生成的监管报告中没有虚拟记录。”

由于系统使用共享平台，银行能够自动实时向监管机构提供更正后的数据而无需多轮跟进。有错误的报告可以在执行定期安排的审计活动之前予以更正。

“我们比较了报告与生产数据库，并与监管团队核对，”Mudigonda 指出，因此新规则不可能将任何合法记录识别为虚拟记录。

Mudigonda 表示，由于该系统现在是数据管理程序的常规组成部分，因此银行对结果的准确性“有 100% 的信心”。“实施后，我们使用数据库中可靠的审计表验证了清除掉的客户记录。”

他解释道，数据库设有为每个事件创建唯一交易 ID 的审计表，允许开发人员端到端地跟踪每个交易的流程。

计算收益

对于银行而言，业务规则创新远不止是清理系统中发现的数千条虚拟记录的一次性解决方案。它已成为一种能够更稳健地维护数据库的长期战略性解决方案。现在，银行定期安排一项自动化作业，每晚分析五次客户 ID，并清除识别到的虚拟记录。

关于投资回报，“该解决方案的总成本（包括设计、开发、与业务部协调以及生产部署）约为 50,000 加元，”Mudigonda 说道。

我们可以合理地假设，如果银行未能满足监管要求的合规性，可能遭受的任何审计相关的处罚会比这个成本高得多。“我实在无法准确量化这一数字，但一般会有数百万加元之多，”Mudigonda 说道。

然而，由于数据管理不当可能产生深远的连锁反应，从某种意义上说，生成值得信任和可信的报告千金难买。“系统中有数百万个客户 ID，”Mudigonda 说。“清除系统中的虚拟记录非常重要，因为有很多其他产品系统依赖这些数据，许多发布报告的监管机构也是如此。”

由于数据管理不当可能产生深远的连锁反应，从某种意义上说，生成值得信任和可信的报告千金难买。

从日常银行业务到抵押贷款、商业银行业务、财富管理、投资产品和资本市场的企业交易，整个银行都依赖于客户数据。Mudigonda 指出，软件创新有助于银行减少隐私侵犯的行为，隐私侵犯问题如果不加以解决，可能会导致银行声誉受损。

他指出，每个业务领域都有合规标准，在各个业务领域运营的组织都可能遇到类似的问题。“实际上这不仅仅适用于金融机构，也可能适用于任何为客户提供服务的组织，”Mudigonda 说道。“可以是电信运营商，也可以是航空公司或者零售店，还可以是任何主要为客户提供服务的商店。例如，如果您是 Costco [总部位于美国的大型连锁店] 的会员，那么您每次到店都可能获得一个唯一的 ID。因此，任何为每位客户创建唯一 ID 的组织，都会面临这个问题。”

通过非常有限的投资以及依靠问题解决团队的关键创新，银行阻止了虚拟记录问题演变成大数据灾难。

ENJOYING THIS ARTICLE?

- Learn more about, discuss and collaborate on risk management in ISACA’s Online Forums. *https://engage.isaca.org/onlineforums*