银行管理信息系统服务器国产化过程中的 ETL 应用迁移①

张春瑞, 赵成坤, 吴 川, 陈 非

(中国建设银行新疆分行信息技术部, 乌鲁木齐 830000)

摘 要: 本文对某国有大型银行省级分行将 ETL 应用系统由访问进口服务器迁移至国产高端容错计算机环境的过程进行了描述. 通过基于应用域的数据拆分和基于视图/物化视图、数据同步、同义词的数据同步, 基本完成了分行应用数据从进口服务器环境往国产高端容错计算机环境的迁移过程. 在这个过程中, 我们发现由于客观条件的限制, 数据同步前后数据的一致性结果比较存在一定难度, 对应用迁移正确性的确认造成不利影响. 在实际的迁移过程中, 通过相应的措施保证, 顺利完成了迁移过程.

关键词: 服务器国产化; ETL; 数据迁移; 物化视图; 国产高端容错计算机

Process of ETL Application Migration to Server Localization on Bank Information Management System

ZHANG Chun-Rui, ZHAO Cheng-Kun, WU Chuan, CHEN Fei

(Information Technology Department, Xinjiang Branch of China Construction Bank, Wulumuqi 830000, China)

Abstract: In this paper, we describe a migration process of ETL system to localized server environment in a large state-owned bank branch. Using technologies of data separation based on application domain and data synchronization based on views/materialized views, data, data synchronization, Synonym, we mainly accomplish the process of migrating branch–related data from imported server environment to localized server environment. In this process, the cause of data incorrectness during application migration is found, which arises from insufficiently support of commercial database system to temporal data. This problem makes it difficult to develop programs to check data synchronization result. In the actual process of migration, through the corresponding measures to ensure that we successfully completed the migration process.

Key words: server localization; ETL; data migration; materialized view; domestic high-end fault-tolerant computer

1 引言

自斯诺登事件爆发后, 国家信息安全的重要性越为引发大众的关注, 这也越发证明政府在 2006 年即开始部署的“核高基重大专项”(核心电子器件、高端通用芯片及基础软件产品)所具有的战略前瞻性. 与当时环境略有不同, 现在国家间的信息战已暴露在公众视野中. 国产基础软件和基础服务器对保障国家信息安全的重要性已经在社会形成共识, 社会舆论不再只将国产基础软件和服务器作为向国外 IT 厂商压价的工具, 而是迫切希望国产基础软件与服务器能够得到真正应用, 并希望尽可能将国计民生相关的数据处理工作转移到国产高端容错计算机上. 银行业务系统安全是国家信息安全的重要组成部分, 社会更希望能在银行业务系统中尽早用上国产高端容错计算机和基础软件. 本文的工作就是在这样的背景下展开的.

本文的工作是国家科技支撑计划项目“国产高端容错计算机等国产化重大信息科技成果在银行业关键系统的示范应用及技术研究”的子课题“基于国产高端容错计算机的行业应用系统关键技术研究与应用示范”的组成部分. ETL(抽取、转换和导入)工作是数据

① 收稿时间:2015-05-12;收到修改稿时间:2015-08-10

仓库的基础, 是银行业务系统的基础环节, ETL 应用若能顺利访问国产高端容错计算机环境, 则说明国产高端容错计算机能在很大层面上支持银行业关键系统的运作. 而对ETL应用迁移中出现问题的解决方案, 对银行其他业务系统的迁移也能起到较好的参考作用.

本文关注于银行ETL应用在访问国产高端容错计算机环境过程中所遇到的实际问题, 通过比较和分析迁移方案的变迁过程, 对服务器国产化过程中应注意的共性问题进行简要讨论, 总结在银行业务系统中将制约服务器国产化的因素, 最后提出可能的解决方案.

随着大数据环境的普遍化, 人们的兴趣倾向于大数据环境下的数据迁移和分布式 ETL 的实现. 例如, 文献[5]总结了大数据环境下海量数据迁移的若干方法, 并结合上海社保系统数据迁移的实践, 对多种方法进行了尝试. 文献[6]将开源工具 Sqoop(可作为数据加载工具)和 Hive(可作为数据转换工具)相结合, 构建一个基于 Hadoop 平台的分布式 ETL 工具. 文献[4]的研究更为深入, 它对 Hadoop 底层源代码进行修改, 改善了 HDFS 数据块分配的合理性, 提高了大数据迁移的传输效率. 文献[10]和[11]研究的是云模型下数据迁移的策略和技术.

与我们的工作最为接近的是以下几个. 文献[9]也是面向银行计算机系统迁移方面的工作, 它的重点是双系统切换时的系统停机设计, 引入停机时间窗口内提供小范围服务的设计方案, 降低因系统切换带来的无法对外提供服务的损失. 文献[10]采用 Java 事务应用接口(JTA), 通过事务控制, 避免数据迁移过程中因意外中止造成的源与目标数据不一致. 这说明, 数据迁移过程中的数据不一致逐步被人们所关注, 但是本文的问题比以上情况更为突出, 因为本文面向持续的数据迁移, 而这两篇文献都面向一次性数据迁移的不一致问题.

本文的架构如下, 第一章描述系统环境; 第二章介绍在迁移过程中出现的问题及因此导致的对迁移方案修改; 第三章对迁移方案的调整进行比较和分析, 得出具有共性的问题; 第四章是对相关研究工作的介绍与比较; 最后一章是对全文的总结, 并描述下一步的工作方向.

2 测试环境描述

由于银行的应用环境错综复杂, 单个系统的成功应用示范难以充分验证国产高端容错服务器对于金融应用需求的满足程度, 在软硬件兼容性研究、应用系统范围等方面都存在较多的局限性, 需要通过示范应用系统的移植进一步扩大研究和应用范围. 本文的工作基于ETL应用系统向国产高端容错计算机迁移的过程, 提出适合业务环境需求的迁移方案, 以下将简要介绍测试环境.

2.1 系统架构

本次迁移工作中, 与本文相关的部分主要是基于服务器及进口服务器环境下ETL应用执行结果的比较, ETL 工作主要用于数据库增量的同步. 由此涉及到异构数据库主机环境下 ETL 应用运行一致性验证、增量数据动态获取、增量数据同步程序的开发及性能比较.

所谓 ETL 应用即(Extraction、Transformation、Loading, 抽取、转换和加载)的缩写, 基本原理就是从各种原始的业务系统(异构多源)中提取数据, 按照预先设计好的规则将抽取到的数据进行转换, 最后将转换完的数据按计划增量或全部导入到目标数据库中[1].

系统的架构图如图 1 所示, 系统分为 USER(用户)层、负载均衡(Linux 虚拟服务器)层、WEB 层、APP(应用)层、DB(数据库)层和存储层等 6 个层次, 本文的工作主要体现APP层次的任务调度模块和ETL集群模块. 除了 USER 层和负载均衡层以外, 每个层次都分成进口服务器(图中左侧)和国产高端容错计算机(图中右侧)两个环境. 这两个环境的程序运行逻辑是一致的, 即针对每日总行下发的数据和分行应用自身产生的数据, 在国外进口服务器和国产高端容错计算机的异构访问环境下运行相同的 ETL 过程, 将每日增量数据更新到各自对应的数据库中. 由于两个环境的 ETL 过程和数据都是一致的, 区别在于数据库管理系统的主机不同, 前者采用 HP 服务器作为主机, 而后者的主机是浪潮天梭高端容错计算机, 这正好形成一个天然的比较环境. 在 ETL 服务器上, 这两个环境都采用 Linux+DataStage EE 的软件架构. 在 DB 服务器方面, 国外进口服务器环境使用 HP UNIX 作为操作系统, 国产高端容错计算机采用 Linux 作为操作系统, 均采用 ORACLE 10g 为数据库.

从图 1 中可以看出, 两个环境是相互独立的, 各有各的服务器. 为了对两个环境进行全面比较, 初始化时, 先把历史数据从进口服务器环境全量导入到国产高端容错计算机环境. 之后, 针对每日新来的数据

文件, 两环境各自执行 ETL 作业, 将数据批量更新到各自的数据库服务器中. 然后, 比较相同的 ETL 程序在这两个环境各自的执行结果和性能, 以便对国产高端容错计算机能否支撑现有业务环境需求、以及如何支持业务环境需求提出有效的解决方案.

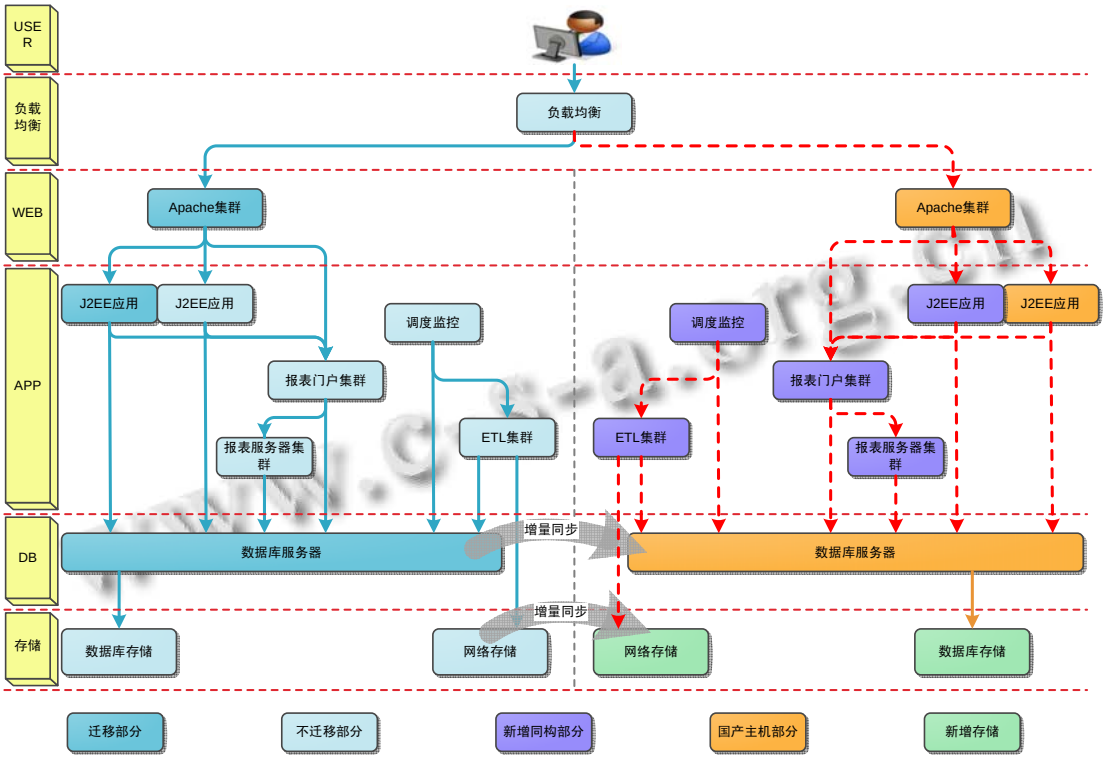

图 1 实验环境系统架构图

2.2 数据与作业描述

ETL 应用的输入数据分为总行下发数据和分行应用数据两类. 前者以下简称总推分数据, 后者简称分行数据, 两者的数据量大致相等. 每个环境中, 数据库内存储的全量数据达到 8T, 批量处理所需的日增量(及短期历史)的数据量达到 2T, 根据本项目存续的不同阶段, 对于存储的峰值要求为 10T.

批量数据处理类应用以后台数据处理作业为主, 每日日常运行作业数量约为 5000 个, 系统峰值出现在月初, 月初峰值每日运行作业数量将达到 7000 个左右.

2.3 约束和目标

与 ETL 应用迁移相关的约束主要有:

1)实现省级分行管理信息系统应用示范与上线运行, 系统连续稳定运行时间超过 90 天;

2)兼容 Oracle 数据库、Weblogic 中间件;

3)ETL 数据批量处理速度保证数据时间窗口的处理要求;

4)系统迁移后性能指标应该与迁移前相当;

具体目标包括:

1) 在国产主机数据库服务器环境下 ETL 应用能正常运行;

2) 多数据库环境下 ETL 应用运行一致性分析;

3)针对差异数据, 分析产生的原因, 并定位需要改造的内容, 如: 定位因 ETL 应用代码的原因造成的数据差异, 并提出解决方案.

3 迁移方案描述

3.1 理想方案(构建完全相同的比较环境)

为了能充分比较两个环境的数据处理准确性和运行效率, 最理想的方案是将所有数据(包含总推分数据和分行数据)都分别导入国产高端容错计算机环境和进口服务器环境, 两环境的数据都从外部文件获得, 避免环境间的数据交互, 降低环境间的耦合性. 具体的方案如图 2 所示, 左边是进口服务器环境, 右边是国产高端容错计算机环境. 初始时, 仅在进口服务器

环境上有 ETL 应用, 国产高端容错计算机环境上没有 ETL 应用. 然后, 将进口服务器环境的数据全量导入国产高端容错计算机环境. 两环境的数据一致后, 启动国产高端容错计算机环境上的 ETL 应用. 最后, 由自动化比对程序对两个环境的执行结果进行比较.

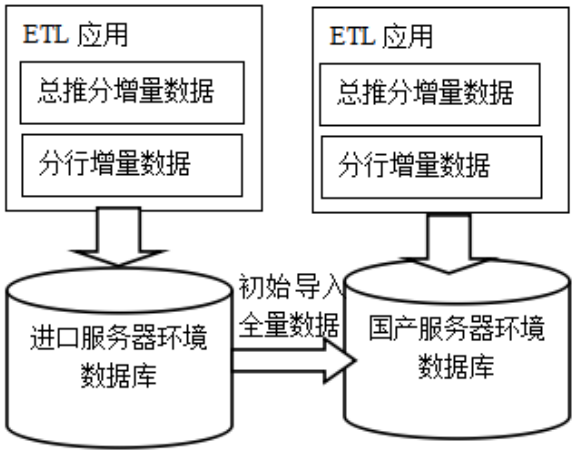

图 2 理想方案

但是在实际应用中, 该方案缺乏实现可能. 主要问题在于:

问题 1: 初始全量导入的数据量过大.

全量导入所需时间长达一周, 而导入过程中, 为避免双方数据不一致, 需要在进口服务器环境(源端)中停止数据更新, 这会导致原有环境长期不可用, 直接影响生产运营, 使得该方案不可行. 需要对迁移方案进行了改造, 降低初始数据导入耗用时间.

最简单的想法是将初始导入过程划分为多个批次, 每日导入一批数据, 通过一段时间的追加导入实现两环境数据一致. 其后, 将两环境分开, 各自执行 ETL 应用. 然而, 这种方法的问题在于批量导入的实现难度较大, 需要对原有系统进行较大改造, 方能实现.

实际的解决方案是将数据分成两部分(总推分数据和分行数据), 仅有分行数据迁移到国产高端容错计算机环境. 这样, 初始化的全量数据从 8T 减少为 3T 以下, 全量导入时间由预估的一周降低至 2 天以内. 该方案的详细介绍如下节所示.

3.2 初始方案(根据应用拆分数据)

由于初始导入数据量极大, 导入过程花费过长时间. 因此, 需要进行调整, 拟基于总推分应用和分行应用的不同, 将总推分数据相关的 ETL 应用放在进口服务器环境执行, 而将分行数据相关的 ETL 应用放在国产高端容错计算机环境执行. 为了充分利用系统资源, 拟基于两环境实现分行应用数据的互备与比较. 为此, 仍然需要在进口服务器环境中保持分行全量数据, 并每日运行 ETL 作业导入分行增量数据, 以便验证国产高端容错计算机环境下作业运行的正确性.

此迁移方案如图 3 所示. 在国产高端容错计算机环境仅保留分行全量数据, 初始化时将进口服务器环境的分行全量数据导入国产高端容错计算机环境数据库, 之后启动国产高端容错计算机环境的 ETL 应用, 国产高端容错计算机的ETL 应用也仅导入增量的分行应用数据. 其次, 由于分行应用数据的 ETL 应用需使用小部分总推分数据, 而国产高端容错计算机环境下缺乏这些数据, 因此需要在国产高端容错计算机环境的数据库中建立相应视图, 通过数据库链映射到进口服务器环境下具体存储数据的表.

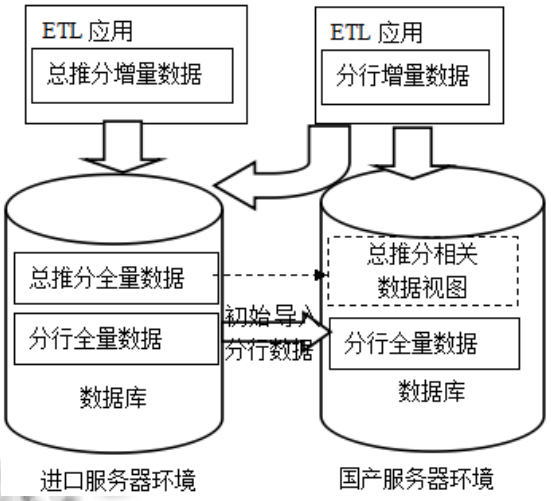

图 3 基于应用拆分数据的迁移方案

该方案解决了数据初始导入时间过长, 对生产系统影响较大的问题. 然而, 由于将数据分拆成两部分, 导致数据之间的关联被切断, 需要引入视图, 以保证 ETL 作业逻辑的正确性. 在小批量抽样数据(1%比例抽样)的测试环境中, 本方案经过验证无论从数据准确性和效率方面, 都没有问题. 但是, 在全量数据的测试环境中, 出现了由于视图的引入带来的跨库的数据查询性能低下问题(详见问题 2 描述).

问题 2: 跨库的数据查询性能低下

在国产高端容错计算机环境的 ETL 应用中, 经常需要关联总推分数据, 方能得到最终的结果. 由于总推分数据与分行数据存储分离, 每次查询时都需要将数据从

远端机器(进口服务器环境)传至本地(国产高端容错计算机环境), 如果表中数据超大, 则网络开销也很巨大, 导致ETL 作业处理的低效. 在全量数据测试环境的实验中, 出现未迁移前 1 分钟内(平均查询时间 53s)能完成的查询语句, 迁移后要花费一个多小时(平均查询时间 1h32m31s)的情况(这在系统使用过程中是不可接受的). 经过分析发现是由于巨型表间的关联引发的, 而在银行 ETL 应用中巨型表相当常见. 同时, 经过数据库执行计划分析发现, 耗时的另一个原因在于跨数据库数据访问引发索引失效(详见问题 3 描述).

问题 3: 跨数据库数据访问引发索引失效

当数据存储于不同的物理机器时, 查询优化需考虑数据跨库转移的开销. 适用 DBLINK 访问远端数据, 会导致在查询时, 由于数据分布存储的原因, 无法用上原本已建好的索引. 虽然, 可以通过在查询中指定 Hint 实现, 但可推广性不强(需要针对原有系统中所有的查询语句进行改造). 而如果采用物化视图的模式, 由于可以在物化视图上建立相关索引, 性能会得到大幅提升(我们采用问题 2 中描述的同样查询语句, 采用物化视图后在全量数据测试环境测得的平均查询时间为 49s).

因此, 本文对实现方案进行了改造, 通过物理视图方式将与分行 ETL 相关的总推分数据也存储一份在国产高端容错计算机环境中, 避免跨库查询的性能开销. 该方案如下节所示.

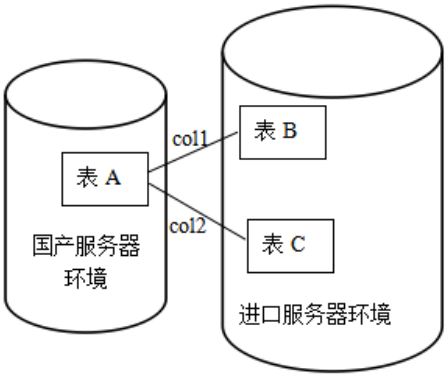

图 4　导致索引失效的查询图例

3.3 现实方案(通过同步保证一致)

本方案与预期方案的区别在于: 国产高端容错计算机环境中原先不实际存储总推分应用的数据, 仅仅存放一个指向具体数据表的视图. 现在为了避免跨库查询带来的网络开销及索引失效问题, 将这份数据也存储在国产高端容错计算机环境中. 为了保证两份数据的一致性, 通过采用 Oracle 物化视图机制, 将国产高端容错计算机环境中的总推分数据建成物化视图. 为避免数据实时刷新带来的网络开销, 物化视图采用延迟刷新的方式, 根据对应 ETL 作业结束信息, 调用 Oracle 刷新物化视图命令. 具体方案如下图所示, 与图 3 的区别在于将视图改成物化视图, 数据以物理形式保存.

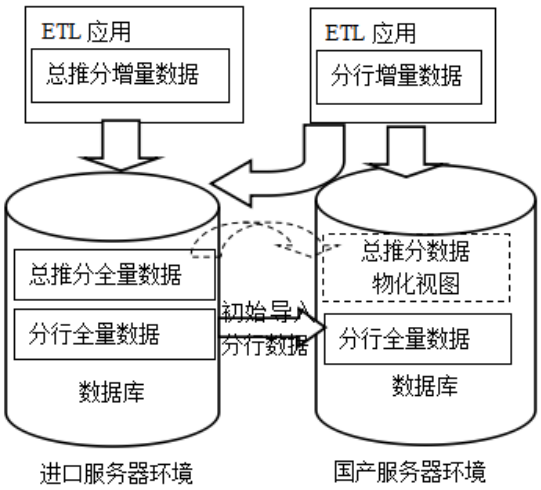

图 5　应用拆分而数据互备的迁移方案

本方案的问题在于两个:

问题 4: 多种数据导入方法;

问题 5: 物化视图带来的潜在数据不一致.

问题 4 在于, 日常运行中, 本方案需要两条数据传递路线, 一种是使用数据复制软件将初始的分行全量数据导入到国产高端容错计算机环境, 另一种是使用 Oracle 自带的物化视图刷新函数, 将总推分相关部分的增量数据刷新到国产高端容错计算机环境. 同是数据导入, 却使用两种不同的方式, 增加了系统运行环境的复杂性, 也对系统的维护带来不利影响. 之所以增加物化视图这种方案, 主要原因在于数据复制软件的用户友好性尚有提高的空间, 外部程序难以确知数据复制软件的数据刷新过程何时结束, 导致无法及时对外发布数据. 因此, 如果数据复制软件如能在这方面得到改善, 接下来需要考虑合并这两种方式.

问题 5 体现在两个方面, 一个是唯一索引的同步不一致性问题, 当源表具有唯一索引时, 如果在物化视图上也建唯一索引, 则会出现刷新时无法正常刷新

数据的出错信息. 解决方案是在物化视图上改建为普通索引, 然而这会导致潜在的数据不一致. 虽然在现有的场景下, 数据仅是单向同步, 不存在这样的问题, 但在复杂场景下仍会有可推广性问题. 另一个问题是检测程序的检测结果会出现不一致情况. 检测程序需要校验刷新后的物化视图是否与源表的数据一致, 然而刷新的过程比较耗时, 可能出现刷新时源表又出现数据更新的情况. 但是, 根据事务一致性原则, 这些之后的变化数据是应被更新到物化视图的. 这导致检验程序在刷新后检查源表与物化视图一致性时容易出错. 根据现实的运行情况, 在 397 张表的更新一致性检验过程中, 会报有大约 5 张表的数据不一致, 出错表的增量数据不一致率平均达到 1%左右. 解决方案就是调整刷新的时间点选择机制, 保证在刷新和检测这段时间内没有对源表的操作. 这个解决方案的问题也是可扩展性受限, 它对于以批量更新为特征的 ETL 应用是可行的, 但频繁更新的源表就难以确定更新时间点了.

在现实情形下, ETL 作业在国产高端容错计算机环境下的性能已达到期望, 甚至许多作业的性能还超过了在进口服务器环境下的表现. 当然, 究其原因部分是由于国产高端容错计算机环境的设备是后采购的, 较进口服务器为优.

在解决 ETL 应用在两环境下运行结果一致且性能满足需求的前提下, 期望将国产高端容错计算机环境推上生产系统, 仅保留一套分行 ETL 应用, 并将进口服务器环境改造为分行数据的备份系统. 具体方案如图 6 所示.

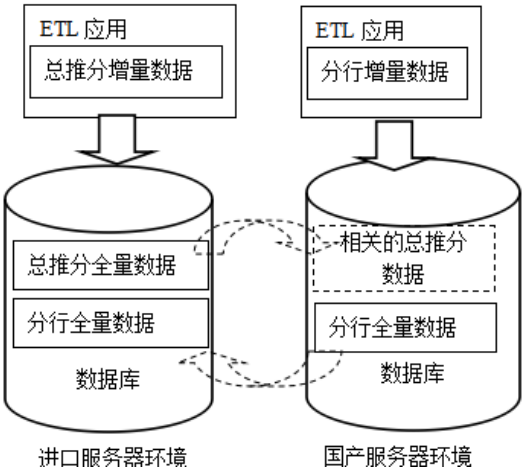

图 6 未来期望的数据架构

此时, ETL 应用仅将分行增量数据导入国产高端容错计算机环境, 为了保证进口服务器环境能备份分行全量数据, 将通过增量复制的方式将国产高端容错计算机环境中加工后的分行增量数据更新到进口服务器环境. 另外, 为统一数据导入方法, 期望通过数据复制软件将分行 ETL 相关的总推分数据每日增量刷新到国产高端容错计算机环境, 实现两环境的循环运作和数据互备. 这就接近我们最初的理想方案, 而且在数据一致的前提下, 通过数据增量复制的方法, 既实现了系统互备, 也避免了冗余的 ETL 应用.

4 分析和展望

在迁移过程中, 从理想方案, 到初始方案和现实方案, 经历了数次变迁. 变迁的原因在于海量数据带来的数据同步时间过长, 导致同步过程中跨库数据的一致性难以保证. 这也说明了应用系统往国产高端容错计算机迁移过程的复杂性. 但是通过实际迁移过程中, 在数据自动化比对、数据同步流程的优化的过程中进行的努力以及所取得的相应成果, 使得整体迁移过程逐步走向成功.

从 ETL 应用往国产高端容错计算机环境迁移的过程可以看出, 目前问题的症结在于软件生态并不能很好地配合服务器国产化的趋势, 服务器国产化需要合适的迁移软件相配合. 以本次迁移而言, 如果数据同步软件在性能和准确性上符合要求, 则迁移过程将会简单得多.

在未来的国产化过程中, 除非新上线的系统, 任何一个部门都不大可能一步到位, 需要有新旧环境并行运行与数据切换的步骤. 当存在并行运行的情况下, 就会出现海量数据的跨库迁移. 目前在海量数据环境下, 基于日志的数据增量同步软件对实时性迁移的性能支持不够好. 通过这次测试环境的搭建和实施, 我们越发感到数据迁移工作需要与作业调度工作相结合. 作业调度工具需要留下一个接口, 让用户进行配置, 使得迁移过程后用户能通过事务号进行查询, 还原迁移过程时的数据场景. 这属于时态数据库的研究范畴, 也就是说对于持续性的数据迁移场景, 传统关系数据库需要加强对时态信息的支持. 这样才有助于保障面向国产化环境迁移过程的准确性.

5 总结

本文对国有某大型银行省分行将ETL应用系统迁移至国产化服务器环境的过程进行了描述. 通过基于应用域的数据拆分和基于视图/物化视图的数据同步, 基本完成了分行应用数据从进口服务器环境往国产高端容错计算机环境的迁移过程. 在这个过程中, 我们面临着由于商业数据库对时态数据支持不充分的原因, 导致数据同步后结果比较存在一定难度, 特别是在源表出现意外变更时, 需要在数据库存储格式、数据同步软件或数据库作业调度层进行调整作进一步的支撑. 但是在实际过程中, 我们通过对自动化数据比对程序的开发以及数据同步程序效率提升的探索, 在很大程度上规避了相应问题的结果.

经过我们在此次ETL应用系统由访问进口服务器迁移至国产高端容错计算机环境的过程的探索, 更加坚定了我们对于国产高端容错计算机的信心, 而这个探索也必将对后续国产高端容错计算机在金融行业的广泛应用起到有力的助推作用.

参考文献

1 宋鹏,廉继.ETL 技术在复杂数据迁移项目中的应用.西安工程大学学报,2008,22(4):493–497.
2 丛慧刚,任庆东,李天阳,袁满.元数据驱动的大型数据库迁移工具实现.科学技术与工程,2011,11(10):2353 – 2356
3 陈园园,陶飞.社保信息系统中数据迁移的实现.苏州市职业大学学报,2011,22(2):27–30
4 何刚.基于 Hadoop 平台的分布式 ETL 研究与实现[学位论文].上海:东华大学,2014.
5 王刚,王冬,李文,李光亚.大数据环境下的数据迁移技术研究.微型电脑应用,2013,30(5):1–3
6 刘豹.一种分布式 ETL 工具的设计与实现.软件,2013,34(10):73–77
7 韩剑峰.可配置化数据迁移框架的研究与实现[学位论文].上海:上海交通大学, 2011
8 唐小新.基于 Unicode 字符集数据迁移的设计与实现.企业科技与发展,2011,(17):22–24
9 林卫华.银行计算机系统数据迁移与系统停机的研究与应用[学位论文].长春:吉林大学,2011.
10 张瑞藜.自治云服务模型与数据迁移策略的设计.科技研究,2012,(16):95
11 吴升启.基表加扩展表多租户数据存储及数据迁移的研究[学位论文].济南:山东大学,2012