文章编号：2096-1472(2021)-03-08-03
DOI:10.19644/j.cnki.issn2096-1472.2021.03.002

基于双排序原理的先进先出数据核销高速通用算法

王国忠

(上海交通大学电子信息与电气工程学院，上海 200240)
✉gzwang@sjtu.edu.cn

摘　要：本文按照软件中台的思想，设计了一个针对先进先出(First-In First-Out，FIFO)[1]数据核销的通用实现框架及其对应的入库数据模型、消费数据模型和匹配核销模型。同时，设计了基于双排序原理的先进先出数据核销高速实现方法。该方法按照匹配规则先对数据进行排序，然后对两个有序队列进行单循环匹配查找，避免了传统先进先出实现中的双循环操作，可以大幅度提高运算性能，大幅度节省CPU开销和存储开销，实现超大数据量的快速匹配，可以支持两个亿级数据库的快速先进先出匹配。

关键词：先进先出；存储过程；双排序；面向对象
中图分类号：TP31　**文献标识码**：A

High Speed Algorithm for First-In First-Out Data Verification based on Double Sorting Principle

WANG Guozhong

(*School of Electronic Information and Electrical Engineering*, *Shanghai Jiao Tong University*, *Shanghai* 200240, *China*)
✉gzwang@sjtu.edu.cn

Abstract: This paper proposed to design a general implementation framework for FIFO (First-In First-Out) data verification and its corresponding inbound data model, consumption data model and matching verification model based on idea of middle platform. At the same time, a high-speed realization method of FIFO data verification is designed based on the double sorting principle. This method first sorts data according to matching rules, and then performs a single-loop matching search on the two ordered queues, avoiding double-loop operation in traditional FIFO implementation. The proposed method significantly improves computing performance, and greatly saves CPU and storage overheads. As a result, it is capable of processing super large data volume, and can support fast FIFO matching of two databases with large data volume.

Keywords: First-In First-Out (FIFO); storage procedure; double sort; object-oriented

1 引言(Introduction)

先进先出是传统库存商品成本核算的一个通用方法，商品销售先选择最早采购的商品价格作为销售商品的库存成本。目前，国内很多企业对于各类消费数据都需要用先进先出的方法进行核算。随着互联网新零售大幅度崛起，消费数据从传统的商品批发零售扩大到很多场景，例如CRM[2](Customer Relationship Management)的积分核销[3]就是十分典型的一个应用，不同商家的积分产生和使用，需要相互结算。随着全渠道零售的兴起，各个结算主体之间也存在先进先出的结算需求。从软件实现上来看，数据库先进先出核销需要解决模块化和运算高效的技术难点。先进先出算法的缺陷会使系统运行性能变差，甚至导致系统瘫痪。

为了节省服务器的计算时间，需要设计一个通用的高速核销引擎，方便应用程序直接调用,大幅度降低普通程序员对出入数据核销实现的门槛。目前数据基本上都存储在数据库中，而程序开发工具大部分是Java或C语言等，通过JDBC/ODBC和数据库连接读取和写入数据。由于数据库和开发工具都具备运算能力，因此先进先出引擎可以利用开发工具能力，也可以利用数据库能力，以及两者混合能力来实现。

本文按照软件中台的思想，设计了一个针对先进先出数据核销的通用实现框架及其对应的入库数据模型、消费数据模型和匹配核销模型。同时，为了克服传统先进先出数据核销中的双循环操作瓶颈，设计了基于双排序原理的先进先出数据核销高速实现方法，通过先排序后先进先出，把先进先

出算法从乘法变成加法，实现超高速的计算引擎[4]，可以满足超大数据的先进先出需求。

2　先进先出通用模型设计(Design of FIFO general model)

为了实现通用的先进先出引擎，本文设计了一个先进先出通用实现框架[5]。如图1所示，该框架包含来源数据模型FIFO_in、消费数据模型FIFO_out和匹配核销模型FIFO三个模块。任何需要使用先进先出算法的系统，只要将数据按照要求存入两个出入表，然后调用先进先出引擎运行，即可得到结果。

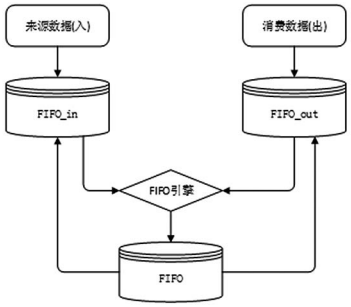

图1 先进先出通用实现框架

Fig.1 First-In First-Out universal implementation framework

入库数据模型FIFO_in，存储来源数据。例如对于商品，就是采购入库数据；对于顾客积分，就是顾客消费产生的积分。为了标识来源数据唯一性，该表需要加入来源数据的唯一主键信息，主要字段为：先进先出种类(用于区分应用场景，例如商品出入库、顾客积分等)、来源数据关键词(一般是入库单号+商品+日期或产生积分的销售小票单号+顾客信息+日期)、来源数据的数值(数量、金额、成本、积分等)、来源数据的优先级(解决FI的不同场景，例如后进先出，只需要调整优先级即可)、已核销的数值(被FO匹配到的数值)、未被核销的数值等。

消费数据模型FIFO_out，存储使用数据。对于商品，就是销售或批发数据；对于顾客积分，就是使用积分(包括积分抵现、退货、积分换券等用掉的积分)。为了标识目标数据唯一性，该表需要加入消费数据的唯一主键信息，主要字段为：先进先出种类(商品出入库或积分核销等场景)、消费数据关键词(一般是销售单号或积分使用单号，加上产品或顾客信息)、消费数据的数值(数量、金额、成本、积分等)、消费数据的优先级(解决哪些数据优先处理的问题)、已匹配数据(已经匹配到的数值)等。

匹配核销模型FIFO，存储匹配记录，即消费数据和来源数据核销关系。这需要通过运行先进先出引擎从FIFO_out循环，逐行从FIFO_in表寻找。FIFO匹配核销模型主要信息为：先进先出种类、出方关键词、入方关键词及核销数据。

3　五种先进先出算法(Five FIFO algorithms)

传统数据核销大部分是单个数据发生后，通过某些条件去寻找来源，这样实现比较简单。例如，顾客使用积分抵现，由于积分可能是其他商家产生的，这样就需要针对使用积分找到提供积分的商家销售单进行结算。实现单个数据寻找算法很简单，就是找到这个顾客以前的积分记录，把未核销的积分按照先进先出原则排序，逐个匹配，将匹配到的记录存在数据库中，并标记已经核销的记录，防止重复核销。然而当数据数量非常庞大时，先进先出引擎算法的效率直接决定了整个系统的运算性能，一个有缺陷的先进先出算法会使系统运行性能变差甚至导致系统瘫痪。

传统的先进先出数据核销采用双循环机制，运算量大，不适用于处理大规模的数据。为了解决这个问题，本文设计了基于双排序原理的先进先出数据核销高速实现方法。该方法按照匹配规则先对数据进行排序，然后对两个有序队列进行单循环匹配查找，避免了传统先进先出实现中的双循环操作，可以大幅度提高运算性能，节省CPU开销和存储开销，实现超大数据量的快速匹配，可以支持两个亿级数据库的快速先进先出匹配。本节从先进先出数据核销算法演进的角度，对以下五种算法的原理和优缺点进行说明。

(1)软件对象双循环匹配查找：基于面向对象的软件开发思想，先将来源数据和消费数据分别映射成开发工具环境的两个对象，通过应用软件(Java/C#)双循环进行匹配查找，然后把结果写到数据库中。

(2)软件调用数据库命令交互查找匹配：应用软件负责写SQL查找和修改语句，交互式调用数据库进行查找运算。这个方法中开发工具不存储大量数据，全部利用数据库的存储和SQL能力，进行循环调用。

(3)数据库内部双循环匹配查找：利用数据库的PL/SQL[6]能力，实现双循环匹配查找，不需要非数据库的开发语言代码。

(4)开发语言双排序单循环匹配查找：首先通过面向对象的开发语言排序，然后进行排序匹配查找，最后批量写到数据库中。

(5)数据库排序单循环匹配查找：通过PL/SQL，利用数据库的排序能力，直接在数据中排序快速匹配，不需要应用开发工具反复调用数据库计算能力。

3.1　方法1：软件对象双循环匹配查找

这是最传统的面向对象的方法，利用应用开发工具，例如Java、C++[7]或.net，将来源数据和消费数据分别映射成开发工具环境的两个对象，然后采用双循环方法，对两个对象进行匹配，最后将匹配结果存在FIFO对象并刷新FIFO_in和FIFO_out对象的已匹配数量，把对象保存到数据库的表中，如图2所示。

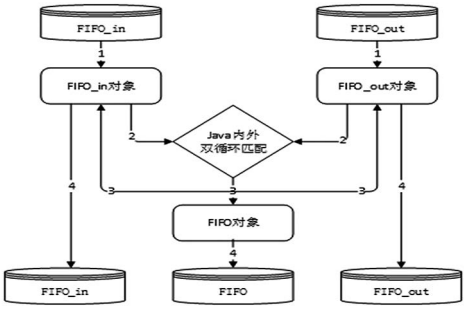

图2 软件对象双循环匹配查找

Fig.2 Software object double loop matching lookup

上述方法利用开发工具能力双循环，总体的计算次数为：出数据记录×进数据记录，所以，总体运算量比较大,效率很低。其优点是方便理解，调试运维效率高。

3.2 方法2：软件调用数据库命令交互查找匹配

本方法也是传统的算法，消费数据和来源数据不需要映射到对象，通过逐条把数据库消费数据读取到内存，然后根据条件从来源数据匹配，匹配到的直接写入数据库，如图3所示。这个方法的特点是开发工具部署的应用环境内存消耗很低，存储全部运用数据库的能力，指示命令都是应用程序下达给数据库的。这个方法类似传统的C/S架构，客户端发命令，数据库运算，并且一次命令对应一个匹配运算。

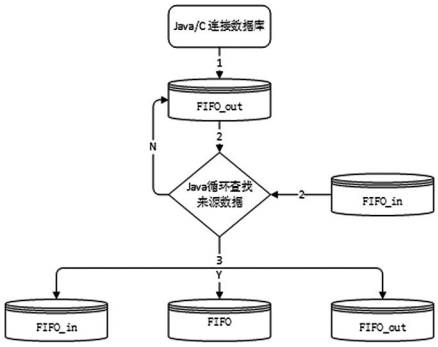

图3 应用软件调用数据库命令交互式查找

Fig.3 The application calls database commands for interactive lookups

这个方法的优点是调试方便，容易运维；缺点是运行效率很低，应用软件和数据库反复交互，每次SQL运行都需要编译，代价比较大。这个算法一般是系统开发初期使用，成熟后需要升级到方法3。

3.3 方法3：数据库内部双循环匹配查找

本方法是方法2的改进，将开发工具的循环代码通过PL/SQL[8]翻译成数据库的存储过程，在数据库内部实现双循环匹配查找，如图4所示。这样客户端和数据库只需要交互一次，具体循环全部在存储过程中，大幅度减少了网络交互的开销和SQL反复编译的代价。

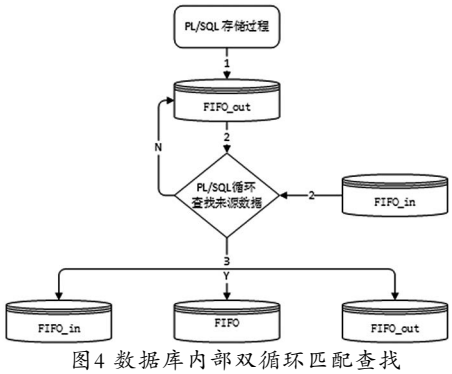

图4 数据库内部双循环匹配查找

Fig.4 Double-loop matching lookup within database

该方法存储过程的优点是运行效率很高，SQL代码都是预编译好的，大幅度节省了数据库编译SQL的时间，同时节省了开发语言和数据库交互的网络时延代价；缺点是出现问题时调试运维不方便，所以需要等方法2成熟后，再翻译成存储过程比较好。

3.4 方法4：开发语言双排序单循环匹配查找

前面三个方法的特点是运用内外双循环的传统算法，运算量都是消费记录和来源记录的乘积，对于数据量很大的场景运算非常慢。特别是匹配出现异常，需要重新匹配的时间很长，无法快速高效支撑大型企业的应用，所以理论上只能停留在实验室，不能投入生产系统。

先进先出本质上是两个大表的匹配，虽然不是Join，实际上可以利用排序Join的思想，将运算量从乘法变成加法，极大提高运算的效率，有力支撑企业先进先出各种场景的应用，包括BI(Business Intelligence)分析需要的先进先出应用。

方法4通过开发语言将数据进行排序,然后基于排序后的队列进行匹配查找，最后将结果批量写到数据库。图5首先将来源和消费两个大表数据映射成开发工具环境的两个对象(类似方法1)，然后分别按照匹配规则排序，变成两个队列。后续匹配算法变成了两个队列从头开始同时向下循环，把双循环变成单循环，运算量最大就是两个队列记录数的总和。

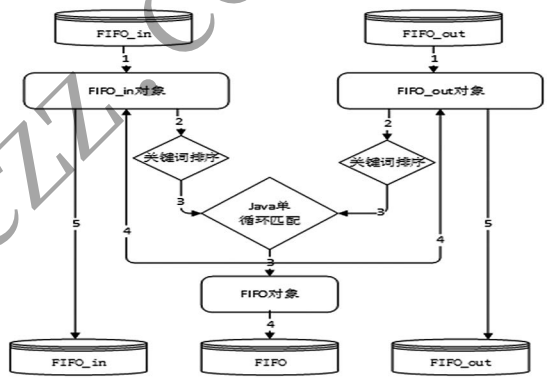

图5 软件排序单循环匹配查找

Fig.5 Software sorting single-loop matching lookup

排序是本方法的主要代价，需要利用已有的优秀排序算法进行快速排序，这样才可以将两个对象按照各自的排序进行单循环匹配。匹配过程是消费数据对象外循环，来源数据(进数据)为内循环。和双循环的主要区别是，这两个循环是同步进行的，谁匹配完一个就移到下一个，所以总的循环次数是两个相加。匹配完成后，需要将匹配结果的对象写回数据库。

3.5 方法5：数据库排序单循环匹配查找

方法4的缺点是需要将大量的数据读到应用服务器的内存，匹配完还需要将大量的匹配数据写回数据库，代价比较大。所以方法5通过PL/SQL将方法4的代码转换成数据库的存储过程，利用数据库的排序能力，可以大幅度提高运算性能，实现超大数据量的快速匹配，大幅度节省CPU开销和存储开销。

图6是数据库排序匹配方法的流程图，排序只需要数据库建一个索引，每秒可以处理十万条记录。然后定义两个Cursor,分别对应消费数据和来源数据，两个循环组成单循环，第一个循环和第二个循环同时进行匹配，整个计算量最多是两个表的记录数相加，所以可以支持两个亿级数据量的

快速先进先出匹配。

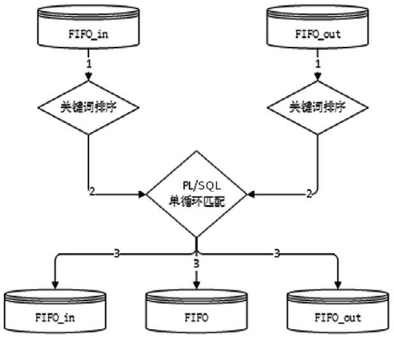

图6 数据库排序单循环匹配查找

Fig.6 Database sorting single-loop matching lookup

4 结论(Conclusion)

本文按照软件中台的思想，设计了一个针对先进先出数据核销的通用实现框架及其对应的入库数据模型、消费数据模型和匹配核销模型，并对五种先进先出数据核销的实现方法进行了阐述。对于企业应用，大部分采用方法5。先进先出作为企业级应用，模块化价值比较高，可防止不同应用场景重复开发，造成额外开发量以及性能潜在的风险。不同算法对系统的要求不一样，效果差异很大。普通程序员很容易使用前面运行效率低的算法，在系统开始运行几个月后，造成系统瘫痪。对于有一定规模的企业应用数据，数据库使用效率是十分重要的。先进先出只是数据库应用中比较常用的一种方法，其他人工智能的算法也需要充分利用数据库能力，合理设计算法。如果超过一个数据库能力，就需要用多个数据库以及大数据的方法解决。

参考文献(References)

[1] 阚运奇.营销零售行业先进先出算法设计[J].无线互联科技,2012(12):98.

[2] 李静.基于数据挖掘技术的电子商务CRM研究[J].现代电子技术,2015(11):126.

[3] 蒋文书.运用数据挖掘技术,精准把握客户需求[J].软件和集成电路,2018(Z1):20.

[4] 侯宁.大数据环境下并行化先进先出成本算法研究[J].软件导刊,2019,18(06):85.

[5] 许桂平.基于数据库的通用驱动程序自动编写算法研究[J].电子设计工程,2019(15):166-169;174.

[6] CJ Fernandez Candel, J Garcia Molina, FJ Bermudez Ruiz, et al. Developing a model-driven reengineering approach for migrating PL/SQL triggers to Java: A practical experience[J]. The Journal of Systems and Software,2019(151):38-64.

[7] 钟玲玲,刘冬雪,黄小平,等.基于C语言的学生信息管理系统设计与实现[J].河南科技学院学报(自然科学版),2019(04):62-67;78.

[8] 周岚.Oracle中基于Java的存储过程[D].合肥:安徽大学,2006.

作者简介：

王国忠(1971-)，男，硕士，讲师.研究领域：大型分布式数据库应用，企业级应用.

(上接第19页)

swarm optimization for multi-objective flexible job-shop scheduling problem[J]. The International Journal of Advanced Manufacturing Technology, 2013, 67(9):2885-2091.

[4] 余建军,孙树栋,郝京辉.免疫算法求解多目标柔性作业车间调度研究[J].计算机集成制造系统,2006,12(10):1643-1650.

[5] YAO B Z, YANG C Y, HU J J, et al. An improved ant colony optimization for flexible job shop scheduling problems[J]. Advanced Science Letters, 2011, 4(6):2127-2131.

[6] 张国辉,高亮,李培根,等.改进遗传算法求解柔性作业车间调度问题[J].机械工程学报,2009,45(7):145-151.

[7] 赵诗奎,方水良,顾新建.柔性车间调度的新型初始机制遗传算法[J].浙江大学学报(工学版),2013,47(6):1022-1030.

[8] 张立果,黎向锋,左敦稳,等.求解多目标柔性作业车间调度问题的两层遗传算法[J].计算机应用,2020,40(S1):14-22.

[9] 付建林,张恒志,张剑,等.自动导引车调度优化研究综述[J].系统仿真学报,2020,32(09):1664-1675.

[10] 戴敏,张玉伟,曾励.绿色作业车间机器与AGV的集成调度研究[J].南京航空航天大学学报,2020,52(03):468-477.

[11] 刘二辉,姚锡凡,陶韬,等.基于改进花授粉算法的共融AGV作业车间调度[J].计算机集成制造系统,2019,25(09):2219-2236.

[12] 徐云琴,叶春明,曹磊.含有AGV的柔性车间调度优化研究[J].计算机应用研究,2018,35(11):3271-3275.

[13] 贺长征,宋豫川,雷琦,等.柔性作业车间多自动导引小车和机器的集成调度[J].中国机械工程,2019,30(04):438-447.

[14] 张超勇,董星,王晓娟,等.基于改进非支配排序遗传算法的多目标柔性作业车间调度[J].机械工程学报,2010,46(11):156-164.

作者简介：

周　鑫(1991-)，男，硕士生.研究领域：智能调度算法，决策分析.

(上接第23页)

[8] 邹斌.基于生物行为特征及单分类算法的手机用户持续认证研究[D].重庆:西南大学,2019.

[9] 刘永帅.移动设备上基于生理行为特征的用户识别方法研究[D].成都:电子科技大学,2016.

[10] Ronao C A, Cho S B. Human activity recognition with smartphone sensors using deep learning neural networks-ScienceDirect[J]. Expert Systems with Applications, 2016, 59(10):235-244.

作者简介：

杨　帆(1982-)，男，本科，中级工程师.研究领域：支付技术，风控，人工智能.