文章编号：2096-1472(2021)-11-07-03
DOI:10.19644/j.cnki.issn2096-1472.2021.011.002

基于多数据库的商户清分负载均衡分片算法研究

张 晖

(银联商务股份有限公司，上海 201203)
✉hz-job@163.com

摘 要：在商户清分批处理作业中，应用程序的高并发数和单机设备的资源利用率总是有一定上限，数据库本身的处理能力和通讯带宽也对批处理作业有一定的约束。根据清分数据商户原子性特征，设计了一种负载均衡的分片算法，实现清分批处理任务在多应用、多数据库间分布式、高并发协同完成，集群节点还可以线性扩展。通过实验测试，对比使用该算法前后的负载均衡性能，分片算法能够在保证商户原子性的情况下有效均衡清分流水，显著提高清分服务器集群的并发读写性能，从而证明了该算法的有效性。

关键词：分布式；高并发；商户原子性；清分
中图分类号：TP312 **文献标识码**：A

Research on Load Balancing Partition Algorithm of Merchant Sorting based on Multi-database

ZHANG Hui

(*China UnionPay Merchant Services Co*., *Ltd*., *Shanghai* 201203, *China*)
✉hz-job@163.com

Abstract: In the merchant sorting batch tasks, high concurrency of applications and resource utilization of single equipment always have a certain upper limit. Processing capacity and communication bandwidth of the database also have certain constraints on batch jobs. According to the atomicity of merchant sorting data, this paper proposes to design a load balancing partition algorithm to realize the distributed and high-concurrency collaborative completion of sorting batch processing tasks among multiple applications and databases. The cluster nodes can also be linearly expanded. Experimental test is made by comparing the load balancing performance before and after using the algorithm. Results verify that the proposed partition algorithm can effectively balance the sorting pipeline under the condition of ensuring the atomicity of merchants, and significantly improve the concurrent read-write performance of the sorting server cluster, which proves its effectiveness.

Keywords: distributed; high concurrency; merchant atomicity; sorting

1 引言(Introduction)

数据清分系统通常要面对庞大的、多方的交易数据，根据一定的业务、勾兑和计算规则对数据进行清洗、拼接和计算等处理，从而得到各利益参与方的资金分配结果，为下游的资金结算和划付提供依据。以上要求决定了清分系统每次处理的数据量都是千万级别的，大型支付机构甚至达到亿级的流水处理量，并且通常要勾兑三方以上的源流水，业务逻辑复杂，处理时效要求高，处理时间段集中在各方联机交易系统完成日切之后的几个小时内，还要为异常情况下重跑批的补救工作留有足够的冗余时间。在互联网数据爆炸式增长的大背景下，传统的应用单机(IBM小型机)高并发已经凸显出IO和CPU瓶颈，应用和数据库分离也存在吞吐量[1]和网络带宽[2](千兆网络交换机)的性能瓶颈。为此，需要借助一定的负载均衡算法将庞大的业务处理均衡地分散到多节点(应用+数据)，以便多机(PC服务器)集群[3]并行分布式处理，实现处理能力的线性扩展。这种集群技术可以用最少的投资获得接近于大型主机的性能，可以满足不断增长的负载需求[4]。由于具有横向扩展性，增加PC机的组合在性价比上已经远超IBM小

型机了，该方案的实施具有较好的经济效益。

2 支付公司商户清分系统一般模型(General model of merchant sorting system of payment company)

商户清分系统的主要功能是汇聚清算组织(银联、网联)和其他第三方渠道清算流水，再对流水进行预处理，根据商户/机构结算、分组规则，分润计算规则，费用收取计算规则等信息，对清算数据进行处理，计算应收发卡转接机构金额、商户本金、手续费、分润、费用等信息，并按要求生成商户、机构对账报表、清分结果等文件。清分系统的功能定位决定了系统的模型如图1所示。

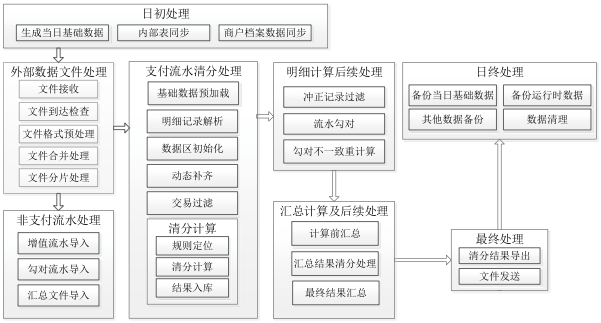

图1 商户清分系统业务模型

Fig.1 Business model of merchant sorting system

清分处理的关键步骤是收集各方清算流水，按照设置的规则勾兑、计算，最终按各方要求出具各类明细和汇总报表。要在规定时间处理千万甚至上亿的流水就需要分布式并发处理，而并发处理的前提是流水的分片，合理的分片能达到良好的负载均衡效果，充分利用每个分布式节点的计算能力。

3 清分数据分片算法研究(Research on partition algorithm of sorting data)

支付领域中商户清分系统的业务特点和清分系统的模型构造，天然需要运用分片思想来达到分布式并行处理的效果。分片原理就是要把一个原始问题分成若干子问题，先递归求得子问题的解，然后把这些解合并起来，得出原始问题的解。在确保商户原子性的前提下，商户清分业务需要分片架构和与之适应的分片算法协同完成。

3.1 分片算法思想

分片(Sharding)是指将内存中的数据拆分成不同的块，分别存储到不同机器上的过程[5]。本文的分片是将商户清分数据按一定的算法拆分成不同的子文件，分别发送到不同的机器上的过程。通过分割数据到不同的服务器，让数据集的不同部分分别由不同的服务器负责，使得单个机器上的请求数减少，系统总负载得到提高，总存储空间也得到提高[6]。

在商户清分业务中，支付机构给每个商户都分配了唯一的商户编号。对同一个商户的交易要求在一个清分处理单元内完成，这就是商户的原子性。商户原子性是本文负载均衡算法的出发点，同时要解决如下几个问题：

(1)撤销交易查找原交易问题、多方文件勾兑问题；

(2)差错交易查找历史原交易问题；

(3)联机退货交易查找历史原交易问题；

(4)特殊商户按照交易汇总金额清算问题；

(5)商户档案信息、配置信息共享问题；

(6)结果文件合并过程中非商户维度问题、部分接口文件存在科目汇总问题；

(7)多库数据中的唯一性冲突和归集问题。

相应的解决方式为：通过改良的倒序贪婪平均分配算法[7]解决(1)中的问题，通过差错交易独立文件解决(2)中的问题，通过开发跨进程的历史库查询服务解决(3)中的问题，通过特殊商户分组解决(4)中的问题，通过OGG[8]数据同步方式解决(5)中的问题，(6)、(7)中的问题通过在汇总服务器中二次加工来解决。先快速在流水文件中提取当日当批次发生交易的商户号及对应的交易笔数，根据清分流水条数和服务器节点数来确定每台服务器的负荷；接着对数据进行切片，每个切片优先分配笔数多的商户，达到临界点时用交易量少的商户进行微调，确保每个商户的完整交易都在一个切片中，并且每个节点设备的交易笔数均衡。

该分片方案在确保商户原子性的基础上，充分发挥了清分系统分布式高并发的计算效能，在实践中达到了如下目标：

(1)清分分布式处理可以实现多台设备交易量的负载均衡，保证商户记录的原子性。通过多台机器的分摊，将总交易量放到多台数据库上分布式执行，增加系统的吞吐量；加强了网络数据处理能力，提升了网络的灵活性和可用性；同一商户的交易在同一个切片中处理，可以确保流水之间的关联不被破坏。

(2)清分分布式处理可以线性扩展处理节点，应对海量数据的处理。随着交易量的增加，可以线性扩展数据库和应用服务器的数量；单节点故障可以临时替补备份机，也可以重新分配执行计划，具有高容错性。

(3)清分分布式处理可以提升各类接口、报表的生成时效性。

3.2 分片集群架构

批处理作业由于在实时计算过程中要传输大批量的流水，如果应用和数据库分离就会出现网络传输瓶颈；如果应用和数据库不分离，在高并发下就会出现CPU、IO瓶颈。为了解决这个两难境地，将源数据分片，每一片数据在一个应用和数据库共同体节点下计算处理，然后将各节点计算结果进行合并处理，出具商户和机构要求结果报表。设计分片集群逻辑架构如图2所示。

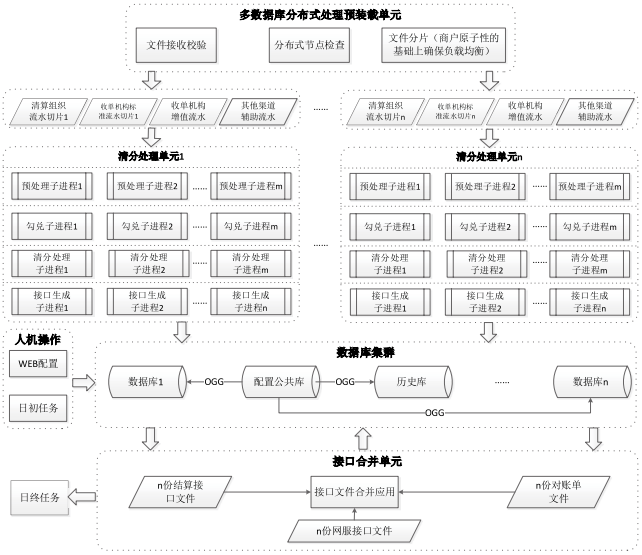

图2 分片集群架构逻辑图

Fig.2 Logical diagram of partitioned cluster architecture

在该架构中，清分流水按照一定算法完成分片，每一个分片在一个节点中完成，一个节点的应用和数据库部署在一

台物理服务器上。计算规则通过公共配置库实时同步到每个节点数据库，历史库也作为公共库，为每个节点提供差错交易查找原交易信息之用。当所有节点计算完成后，接口合并单元汇集每个节点的数据结果进行汇总计算，完成最终的清分报告。本文节点数据库选用Oracle，也可以选用其他商用数据库，对于公共库的访问通过设计自定义数据库访问服务来完成，数据库之间的信息同步使用OGG方式。日初和日终任务用于系统的监控、资源的分配和回收等。

分片集群架构中，公共库(配置库和历史库)的数据无法放入每个集群节点中，本文设计了一套高效的访问机制，如图3所示。每个清分处理单元对应的PC机上，在需要查询公共库信息时，通过调用查询公共库client类(DBClient)中的查询方法来实现历史库的查询。考虑到公共库中交易数据存放的完整性和原始性，DBClient只支持查询(select)，不支持更新(update)和插入(insert)等更改数据的操作方法。查询公共库client类(DBClient)中的方法使用短连接方式访问数据库。按照目前清分系统查询公共库的要求，DBClient封装有很多种不同的查询方法，每次调用DBClient的一个查询方法，返回一条查询结果(找到原交易或未找到原交易)，若找到原交易，将原交易的查询结果返回给查询方法。

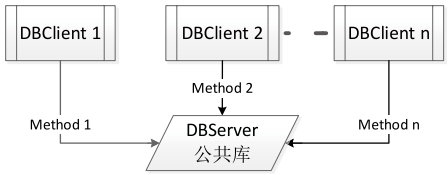

图3 分片计算中公共库访问机制

Fig.3 Access mechanism of shared library in partitioned computing

公共库所在服务器上部署访问数据服务DBServer，它是常驻公共库系统的守护进程，通过监控指定的访问端口port，用来响应客户端类DBClient的各种select查询数据库访问方法，并将查询结果返回给DBClient。考虑到客户端类DBClient的并发查询，该数据服务DBServer需支持多进程访问数据库。

各分片节点完成计算后，要进行文件合并、对账单多维度重构操作。对于后续的结算划付文件，通过各子机器上传的接口文件追加合并完成；对于标准化的商户对账单明细和汇总文件，也可以追加方式进行合并，多序号文件还要进行压缩处理；对于个性化的分析报表，需要将子节点结果数据导入中间表，辅助公共库中的配置信息进行二次加工分析。

3.3 关键算法分析

首先遍历清算组织下发的流水，比如银联下发的为ACOMN流水文件，统计出总交易条数S，发生交易的商户数为m及每个商户对应的交易量$L_i(1\leqslant i\leqslant m)$。如果设备分布式节点的个数为$n$，$a=S/n$，则分配给每个节点处理的条数$S_j(1\leqslant j\leqslant n)$的范围为$[a-d,\ a+d]$，$d$可根据情况进行调整。在进行分片拆分前，将特殊分组的商户挑出来先分配到一台机器上(假设是编号最后的一台)，对剩余的商户可采用快速排序法，按照商户交易量进行降序排序，排序后的商户交易量序列为$L_i(1\leqslant i\leqslant m)$，$L_m$最大，$L_1$最小。为了便于取商户，设两个指针变量，尾巴指针$x=1$，头部指针$y=m$，用二维数组$Q_{ji}(1\leqslant j\leqslant n,1\leqslant i\leqslant m)$存储商户分配结果，以下是商户分配算法的伪代码。

```
//清算流水切片算法
for(j=1;j≤n;j++)
{
  Sj=0；i=0；
  //在下限以下，优选分配大交易量商户，从顶部取
  while (Sj<a−d)
  {
     Sj=Sj+Ly；Qji=My;y−−;i++；
     If (Sj>a+d)
     { Sj=Sj−Ly； y++； i−−； Qji='\0'； break;}
  }
  //在上限以下，用小交易量商户微调，从底部取
  while(Sj<a+d)
  {
     Sj= Sj + Lx；Qji=Mx；x++；i++；
     If(Sj>a+d)
     {
        Sj=Sj−Lx；x−−；i−−； Qji='\0'；
        break；
     }
  }
}
```

商户分配完之后，检查指针x和y的位置，若存在未分配的商户，需要将剩余商户分配给S_j最小的节点；然后根据Q_{ji}中的商户，切割拆分各渠道的清算流水，并把拆分的流水推送到对应的机器节点上。通过该算法就将清分系统待处理的数据源按照商户记录数切分成多片文件，由每个独立的清分处理单元完成每个切分文件的数据批处理。批处理完成后将每个单元的批处理结果发送到归集服务器，经归集服务器合并和多维重构后，还原成下游结算、划付需要的各类接口文件、商户对账文件和报表分析类文件。

在整个清分批处理的过程中，切片和合并的处理是单节点，各个切片的清分计算在多节点分布式集群下处理。由于本切片算法和合并处理都是基于文件操作，可以在处理前将文件导入内存，因为很少涉及对数据库的访问，因而切片和合并没有IO的瓶颈障碍，可以通过应用高并发和扩展CPU数量来提升性能。

4 算法性能评估(Algorithm performance evaluation)

测试环境使用1 台分片服务器，1 台报表合并服务器，8 台分布式计算服务器，每台服务器内存256 GB，CPU为Intel Xeon E5−2680 v4 14C ，千兆网口。操作系统是Suce Unix，数据库使用Oracle 11g。分别测试了在应用本分片算法前后清分跑批的时间消耗，如表1所示。

表1 清分跑批性能测试结果

Tab.1 Test results of sorting batch performance

流水笔数(万)	改进前耗时(分钟)	改进后耗时(分钟)	性能提升百分比(%)
200	8	6	20
300	13	10	22
400	14	11	23
500	16	12	24
1,000	23	15	34

没有使用该算法前，由于商户的原子性，应用是8 个计算节点，数据库只能是一个节点。这种情况下数据库的CPU和IO利用率最高只能到达75%，应用和数据库间带宽上、下限的流量都达到800 Mbs时，加大应用并发量也不能提升了，此时带宽成了性能瓶颈。而使用本文的分片算法保证商户原子性后，既能分布式处理，又能把应用和数据库放在一个物理节点，不存在网络传输瓶颈，每台服务器的CPU和IO利用率可以压测到95%以上，充分使用设备资源。

通过测试数据比对也发现，设备资源得到充分释放后，性能提升20%以上，并且随着数据量的增加，性能提升效果更加明显。这是由于分片算法克服了清分业务商户原子性的问题，又能根据设备节点数负载均衡地分配任务，将来计算节点可以横向扩展，应对亿级的清分流水量。

5 结论(Conclusion)

商户清分系统作为支付机构的核心业务系统，是一种典型的批处理系统，既不同于实时响应的联机交易系统，也不同于高挖掘价值的大数据分析系统。公有云PASS层虽能提供各种分布式工具，但金融数据安全性存疑，搭建私有云又具有较高的成本。本文提供的负载均衡分片算法能够灵活自主地实现商户清分高效分布式处理，保证了商户原子性，兼顾了安全性和实时性，为业界千万级以上交易量的清分提供了一种思路。由于篇幅所限，本文对多节点的调度管理和个别节点异常情况下的恢复措施不再详细探讨。

参考文献(References)

[1] DONG H, SI H, ZONG H, et al. Unstructured mesh generation based on Parallel Virtual Machine in cyber-physical system[J]. EURASIP Journal on Wireless Communications and Networking, 2019(1):1-11.

[2] ZHE S J. Cloud-computing load-balancing mechanism dependent on marine environmental information[J]. Journal of Coastal Research, 2019, 98(sp1):137-140.

[3] SHARMA D. Response time based balancing of load in web server clusters[C]// 2018 International Conference on Reliability. Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO). Noida, India: IEEE, 2018:471-476.

[4] 周莹莲,刘甫.服务器负载均衡技术研究[J].计算机与数字工程,2010,38(4):11-14,35.

[5] 陈敬静,马明栋,王得玉.MongoDB负载均衡算法优化研究[J].计算机技术与发展,2020,30(03):88-92.

[6] 李朝奎,严雯英,杨武,等.三维城市模型数据划分及分布式存储方法[J].地球信息科学学报,2015,17(12):1442-1449.

[7] 黄景修,刘清堂,吴林静.一种面向多终端服务的学术会议管理系统设计与实现[J].计算机应用与软件,2016,33(07):68-71,101.

[8] 贾海军.一种基于OGG方式进行数据迁移的研究[J].软件,2015(05):140-144.

作者简介：

张 晖(1981-)，男，硕士，高级工程师/经济师.研究领域：软件工程.

(上接第14页)

统计结果明显优于前者，其他缺失比例不插补效果则与模拟数据无异。因此也可以说，在实际缺失数据的插补预测中，选择哪种插补方法进行预测研究是数据容量、缺失比例、运算速度和数据分布等因素共同作用的结果，要针对具体情况制订具体方案。

参考文献(References)

[1] 杨晟.基于数据挖掘技术的用户异常用电检测系统的研究与实现[D].北京:北京邮电大学,2019.

[2] 熊中敏,郭怀宇,吴月欣.缺失数据处理方法研究综述[J].计算机工程与应用,2021,57(14):27-38.

[3] 张松兰,王鹏,徐子伟.基于统计相关的缺失值数据处理研究[J].统计与决策,2016(12):13-16.

[4] 朱苗苗.基于时间序列模型的网络流量预测研究[D].西安:西安工程大学,2017.

[5] VAZIFEHDAN M, MOATTAR M H, JALALI M. A hybrid bayesian network and tensor factorization approach for missing value imputation to improve breast cancer recurrence prediction[J]. Journal of King Saud University—Computer and Information Sciences, 2019, 31(2):175-184.

[6] 陈雁声.时间序列中缺失数据的处理方法综述[J].信息与电脑(理论版),2020,32(10):19-22.

[7] 张昕.不完备信息系统下空缺数据处理方法的分析比较[J].海南师范大学学报(自然科学版),2008(04):444-447.

[8] 黄樑昌.kNN填充算法的分析和改进研究[D].桂林:广西师范大学,2010.

[9] 朱高培,朱乐乐,孟马承,等.基于Monte Carlo模拟的四种完全随机双变量缺失数据处理方法的比较[J].中国卫生统计,2018,35(05):707-709.

[10] 林进钿.基于深度学习的电力系统扰动后动态频率特征预测[D].成都:西南交通大学,2019.

作者简介：

徐鸿艳(1997-)，女，硕士生.研究领域：社会经济统计学.

孙云山(1980-)，男，博士，教授.研究领域：信号与信息处理.本文通讯作者.

秦琦琳(1997-)，女，硕士生.研究领域：时序预测，深度学习.

朱明涛(2001-)，男，本科生.研究领域：通信信息处理.