(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 119645982 A
(43)申请公布日 2025.03.18

(21)申请号 202411740028.9

(22)申请日 2024.11.29

(71)申请人 中国农业银行股份有限公司
地址 100005 北京市东城区建国门内大街69号

(72)发明人 毕景新 宋厚岩 侯敏

(74)专利代理机构 北京集佳知识产权代理有限公司 11227
专利代理师 杨倪月

(51)Int.Cl.
G06F 16/215(2019.01)
G06F 16/22(2019.01)

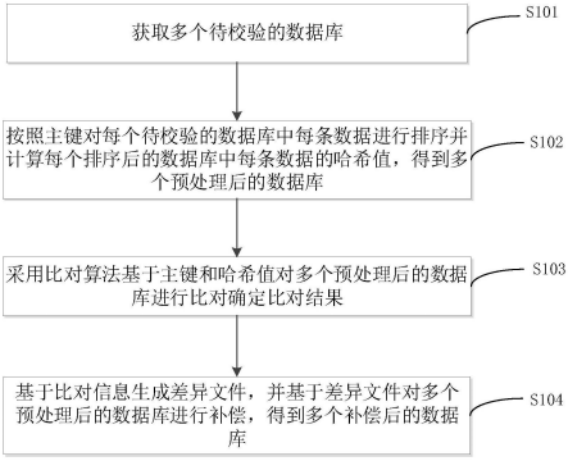
权利要求书2页 说明书12页 附图8页

(54)发明名称

数据库一致性比对与补偿方法、装置、设备、介质及产品

(57)摘要

本申请公开了一种数据库一致性比对与补偿方法,所述方法包括:获取多个待校验的数据库;按照主键对每个待校验的数据库中每条数据进行排序并计算每个排序后的数据库中每条数据的哈希值,得到多个预处理后的数据库;采用比对算法基于主键和哈希值对多个预处理后的数据库进行比对确定比对结果;所述比对结果为第一结果或第二结果;所述第一结果表征多个预处理后的数据库无差异;所述第二结果表征多个预处理后的数据库存在差异;所述第二结果包括存在差异结论和比对信息;基于比对信息生成差异文件,并基于差异文件对多个预处理后的数据库进行补偿,得到多个补偿后的数据库。本申请提高了检测效率、降低了存储成本及资源开销。



CN 119645982 A

1.一种数据库一致性比对与补偿方法,其特征在于,所述数据库一致性比对与补偿方法包括:

获取多个待校验的数据库;

按照主键对每个待校验的数据库中每条数据进行排序并计算每个排序后的数据库中每条数据的哈希值,得到多个预处理后的数据库;

采用比对算法基于主键和哈希值对多个预处理后的数据库进行比对确定比对结果;所述比对结果为第一结果或第二结果;所述第一结果表征多个预处理后的数据库无差异;所述第二结果表征多个预处理后的数据库存在差异;所述第二结果包括存在差异结论和比对信息;

基于比对信息生成差异文件,并基于差异文件对多个预处理后的数据库进行补偿,得到多个补偿后的数据库。

2.根据权利要求1所述的数据库一致性比对与补偿方法,其特征在于,所述采用比对算法基于主键和哈希值对多个预处理后的数据库进行比对确定比对结果,具体包括:

将多个预处理后的数据库中的任意一个数据库作为标准数据库,将剩余数据库作为待比对数据库;

基于主键和哈希值比对所述标准数据库中每条数据与待比对数据库中每条数据生成比对结果。

3.根据权利要求2所述的数据库一致性比对与补偿方法,其特征在于,所述基于主键和哈希值比对所述标准数据库中每条数据与待比对数据库中每条数据生成比对结果,具体包括:

判断每个待比对数据库中每条数据的主键和哈希值是否均满足第一预设条件,得到第一判断结果;所述第一预设条件为待比对数据库中数据和标准数据库对应的数据的主键一致且哈希值一致;

若所述第一判断结果为是,则表征多个待比对数据与标准数据库一致,则比对结果为第一结果;

若所述第一判断结果为否,则表征比对结果为第二结果,基于比对规则基于主键和哈希值比对标准数据库中每条数据与待比对数据库中每条数据生成第二结果中的比对信息。

4.根据权利要求3所述的数据库一致性比对与补偿方法,其特征在于,所述基于比对规则基于主键和哈希值比对标准数据库中每条数据与待比对数据库中每条数据生成第二结果中的比对信息,具体包括:

判断多个待比对数据库中不满足第一预设条件的待比对数据库中的数据是否满足第二预设条件,得到第二判断结果;所述第二预设条件为待比对数据库中数据和标准数据库对应的数据的主键一致且哈希值不一致;

若所述第二判断结果为是,则将满足第二预设条件的待比对数据库中对应数据的比对信息记录为不一致;

若所述第二判断结果为否,判断多个待比对数据库中不满足第一预设条件和第二预设条件的待比对数据库中的数据是否满足第三预设条件,得到第三判断结果;所述第三预设条件为待比对数据库中数据的主键小于标准数据库对应的数据的主键或待校验数据的数据总条数大于标准数据的数据总条数;

若所述第三判断结果为是,则将满足第三预设条件的待比对数据库中对应数据的比对信息记录为冗余;

若所述第三判断结果为否,则将不满足第三预设条件的待比对数据库中对应数据的比对信息记录为缺失。

5.根据权利要求4所述的数据库一致性比对与补偿方法,其特征在于,所述基于差异文件对多个预处理后的数据库进行补偿,得到多个补偿后的数据库,具体包括:

解析差异文件,得到解析文件;

基于解析文件确定每个待比对数据库中每条数据的补偿方式,并基于预设补偿方式对每个待比对数据库进行补偿,得到多个补偿后的数据库。

6.根据权利要求5所述的数据库一致性比对与补偿方法,其特征在于,所述基于解析文件确定每个待比对数据库中每条数据的补偿方式,并基于预设补偿方式对每个待比对数据库进行补偿,得到多个补偿后的数据库,具体包括:

基于所述解析文件,判断每个待比对数据库中每条数据是否存在不一致情况,得到第四判断结果;

若所述第四判断结果为是,则基于标准数据库对不一致的数据进行更新;

若所述第四判断结果为否,则判断每个待比对数据库中每条数据是否存在冗余情况,得到第五判断结果;

若所述第五判断结果为是,则删除冗余的数据;

若所述第五判断结果为否,表征多个待比对数据库中存在缺失,则基于标准数据库对缺失的数据进行新增。

7.一种数据库一致性比对与补偿装置,其特征在于,所述数据库一致性比对与补偿装置包括:

获取单元,用于获取多个待校验的数据库;

预处理单元,用于按照主键对每个待校验的数据库中每条数据进行排序并计算每个排序后的数据库中每条数据的哈希值,得到多个预处理后的数据库;

比对单元,用于采用比对算法基于主键和哈希值对多个预处理后的数据库进行比对确定比对结果;所述比对结果为第一结果或第二结果;所述第一结果表征多个预处理后的数据库无差异;所述第二结果表征多个预处理后的数据库存在差异;所述第二结果包括存在差异结论和比对信息;

补偿单元,用于基于比对信息生成差异文件,并基于差异文件对多个预处理后的数据库进行补偿,得到多个补偿后的数据库。

8.一种计算机设备,包括:存储器、处理器以及存储在存储器上并可在处理器上运行的计算机程序,其特征在于,所述处理器执行所述计算机程序以实现权利要求1-6中任一项所述的数据库一致性比对与补偿方法。

9.一种计算机可读存储介质,其上存储有计算机程序,其特征在于,该计算机程序被处理器执行时实现权利要求1-6中任一项所述的数据库一致性比对与补偿方法。

10.一种计算机程序产品,包括计算机程序,其特征在于,该计算机程序被处理器执行时实现权利要求1-6中任一项所述的数据库一致性比对与补偿方法。

数据库一致性比对与补偿方法、装置、设备、介质及产品

技术领域

[0001] 本申请涉及数据库技术领域，特别是涉及一种数据库一致性比对与补偿方法、装置、设备、介质及产品。

背景技术

[0002] 随着系统业务规模的不断增长，业务的稳定性备受关注，大多企业选择部署多套数据库服务来提高系统的可用性和容灾能力，这样当一个机房出现故障，业务可迅速切换到另外的机房正常工作，因此，各个机房数据库之间的数据一致性校验尤为重要。但伴随着交易量的逐年上升，不可避免的会出现多个数据库之间数据不一致的问题，且出现数据差异时很难被及时发现，使得数据比对、校验的工作变得越来越困难。在此情况下，若执行数据库切换操作会影响业务系统的正常运行，一些情况下，还可能造成严重的生产事故。

[0003] 为了解决数据库一致性问题，一般采用数据副本的方式，利用多个数据副本和多个历史数据副本进行比对，得到多个修改记录集，基于多个修改记录集进行数据比对和补偿。但该方式资源开销大，消耗较多的存储空间，且效率较低，不适合于实际使用。

发明内容

[0004] 基于上述问题，本申请提供了一种数据库一致性比对与补偿方法、装置、设备、介质及产品，目的是提高检测效率、降低存储成本及资源开销。

[0005] 本申请实施例公开了如下技术方案：

[0006] 本申请第一方面提供了一种数据库一致性比对与补偿方法，该方法包括：

[0007] 获取多个待校验的数据库；

[0008] 按照主键对每个待校验的数据库中每条数据进行排序并计算每个排序后的数据库中每条数据的哈希值，得到多个预处理后的数据库；

[0009] 采用比对算法基于主键和哈希值对多个预处理后的数据库进行比对确定比对结果；所述比对结果为第一结果或第二结果；所述第一结果表征多个预处理后的数据库无差异；所述第二结果表征多个预处理后的数据库存在差异；所述第二结果包括存在差异结论和比对信息；

[0010] 基于比对信息生成差异文件，并基于差异文件对多个预处理后的数据库进行补偿，得到多个补偿后的数据库。

[0011] 可选地，所述采用比对算法基于主键和哈希值对多个预处理后的数据库进行比对确定比对结果，具体包括：

[0012] 将多个预处理后的数据库中的任意一个数据库作为标准数据库，将剩余数据库作为待比对数据库；

[0013] 基于主键和哈希值比对所述标准数据库中每条数据与待比对数据库中每条数据生成比对结果。

[0014] 可选地，所述基于主键和哈希值比对所述标准数据库中每条数据与待比对数据库

中每条数据生成比对结果,具体包括:

[0015] 判断每个待比对数据库中每条数据的主键和哈希值是否均满足第一预设条件,得到第一判断结果;所述第一预设条件为待比对数据库中数据和标准数据库对应的数据的主键一致且哈希值一致;

[0016] 若所述第一判断结果为是,则表征多个待比对数据与标准数据库一致,则比对结果为第一结果;

[0017] 若所述第一判断结果为否,则表征比对结果为第二结果,基于比对规则基于主键和哈希值比对标准数据库中每条数据与待比对数据库中每条数据生成第二结果中的比对信息。

[0018] 可选地,所述基于比对规则基于主键和哈希值比对标准数据库中每条数据与待比对数据库中每条数据生成第二结果中的比对信息,具体包括:

[0019] 判断多个待比对数据库中不满足第一预设条件的待比对数据库中的数据是否满足第二预设条件,得到第二判断结果;所述第二预设条件为待比对数据库中数据和标准数据库对应的数据的主键一致且哈希值不一致;

[0020] 若所述第二判断结果为是,则将满足第二预设条件的待比对数据库中对应数据的比对信息记录为不一致;

[0021] 若所述第二判断结果为否,判断多个待比对数据库中不满足第一预设条件和第二预设条件的待比对数据库中的数据是否满足第三预设条件,得到第三判断结果;所述第三预设条件为待比对数据库中数据的主键小于标准数据库对应的数据的主键或待校验数据的数据总条数大于标准数据的数据总条数;

[0022] 若所述第三判断结果为是,则将满足第三预设条件的待比对数据库中对应数据的比对信息记录为冗余;

[0023] 若所述第三判断结果为否,则将不满足第三预设条件的待比对数据库中对应数据的比对信息记录为缺失。

[0024] 可选地,所述基于差异文件对多个预处理后的数据库进行补偿,得到多个补偿后的数据库,具体包括:

[0025] 解析差异文件,得到解析文件;

[0026] 基于解析文件确定每个待比对数据库中每条数据的补偿方式,并基于预设补偿方式对每个待比对数据库进行补偿,得到多个补偿后的数据库。

[0027] 可选地,所述基于解析文件确定每个待比对数据库中每条数据的补偿方式,并基于预设补偿方式对每个待比对数据库进行补偿,得到多个补偿后的数据库,具体包括:

[0028] 基于所述解析文件,判断每个待比对数据库中每条数据是否存在不一致情况,得到第四判断结果;

[0029] 若所述第四判断结果为是,则基于标准数据库对不一致的数据进行更新;

[0030] 若所述第四判断结果为否,则判断每个待比对数据库中每条数据是否存在冗余情况,得到第五判断结果;

[0031] 若所述第五判断结果为是,则删除冗余的数据;

[0032] 若所述第五判断结果为否,表征多个待比对数据库中存在缺失,则基于标准数据库对缺失的数据进行新增。

[0033] 本申请第二方面提供了一种数据库一致性比对与补偿装置,该装置包括:

[0034] 获取单元,用于获取多个待校验的数据库;

[0035] 预处理单元,用于按照主键对每个待校验的数据库中每条数据进行排序并计算每个排序后的数据库中每条数据的哈希值,得到多个预处理后的数据库;

[0036] 比对单元,用于采用比对算法基于主键和哈希值对多个预处理后的数据库进行比对确定比对结果;所述比对结果为第一结果或第二结果;所述第一结果表征多个预处理后的数据库无差异;所述第二结果表征多个预处理后的数据库存在差异;所述第二结果包括存在差异结论和比对信息;

[0037] 补偿单元,用于基于比对信息生成差异文件,并基于差异文件对多个预处理后的数据库进行补偿,得到多个补偿后的数据库。

[0038] 本申请第三方面提供了一种计算机设备,该设备包括:存储器、处理器以及存储在存储器上并可在处理器上运行的计算机程序,所述处理器执行所述计算机程序以实现上述所述的数据库一致性比对与补偿方法。

[0039] 本申请第四方面提供了一种计算机可读存储介质,该介质上存储有计算机程序,该计算机程序被处理器执行时实现上述所述的数据库一致性比对与补偿方法。

[0040] 本申请第五方面提供了一种计算机程序产品,该产品包括计算机程序,该计算机程序被处理器执行时实现上述所述的数据库一致性比对与补偿方法。

[0041] 相较于现有技术,本申请具有以下有益效果:

[0042] 本申请按照主键对每个数据库进行数据排序,保证了在后续比对和补偿时以数据库之间比对的基础是一致的,实现对不同环境、不同数据库类型的数据进行差异比对,后分别计算待校验的数据库中每一条数据的哈希值,通过逐步比对每一条数据的主键和哈希值确定比对结果。比对结果为第一结果或第二结果,若比对结果为第一结果则表征待校验的数据库之间不存在差异,若比对结果为第二结果则表征待校验的数据库之间存在差异,且第二结果还能反映出比对信息,方便实现数据库之间的补偿,因此本申请在比对和补偿过程中均不需要多个修改记录集以及数据副本,检测效率高、存储成本低且资源开销小。

附图说明

[0043] 为了更清楚地说明本申请实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本申请的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动性的前提下,还可以根据这些附图获得其他的附图。

[0044] 图1为本申请实施例提供的数据库一致性比对与补偿方法的流程图;

[0045] 图2为本申请实施例提供的数据库一致性比对与补偿方法的比对算法第一示意图;

[0046] 图3为本申请实施例提供的数据库一致性比对与补偿方法的比对算法第二示意图;

[0047] 图4为本申请实施例提供的数据库一致性比对与补偿装置的结构图;

[0048] 图5为本申请实施例提供的预处理单元和比对单元的工作流程图;

[0049] 图6为本申请实施例提供的补偿单元的工作流程图;

[0050] 图7为本申请实施例提供的控制单元的工作流程图；

[0051] 图8为本申请实施例提供的异常监控单元的工作流程图；

[0052] 图9为本申请实施例提供的数据库一致性比对与补偿装置的工作流程图。

具体实施方式

[0053] 正如前文描述，目前的多数据库比对和补偿资源开销大，消耗较多的存储空间，且效率较低，不适合于实际使用。发明人经过研究发现按照主键对待校验的数据库进行排序，并分别计算待校验的数据库中每一条数据的哈希值。通过逐步比对每一条数据的主键和哈希值确定比对结果，可实现检测效率高、存储成本低且资源开销小。

[0054] 为了使本技术领域的人员更好地理解本申请方案，下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范围。

[0055] 相关术语解释：

[0056] 数据补偿：当数据库中某些数据丢失或错误时，恢复丢失或错误的数据，确保数据的准确性和完整性。

[0057] 容灾：确保企业业务和应用在遭受自然灾害、人为破坏、网络攻击等突发事件时，能够快速且有效地恢复并继续正常运营的策略和措施。

[0058] 数据备份：将数据库中的数据复制到其他存储介质上的过程。

[0059] 数据源：指存储和提供数据的地方或来源。

[0060] 时间复杂度：用来描述算法在处理不用规模数据时的运行时间。

[0061] 空间复杂度：用来描述算法在运行过程中所需额外存储空间的量度。

[0062] 目前一些已知的数据库一致性比对与补偿的技术方案一般是：

[0063] 从数据源模块中选中目标数据库并采集数据，对采集的数据进行一致性检测并生成差异化脚本。此方法虽支持比对多数据库之间的结构性差异，支持一键生成差异化脚本，但直接对源数据进行比较，当数据规模较大时对系统的开销占用也较大，且在数据补偿时通过比对生成的差异化脚本自动补偿数据，难以保证差异化脚本的准确性。

[0064] 或者是，配置要进行数据比对的数据源，比对源端与目的端数据的2张表，进行总行数计算，再对源端与目的端数据的2张表进行哈希计算，将表中每行的数据取出，每行数据生成唯一的哈希值，如果目的端的哈希值在源表中不存在，再将此行数据标记为不一致，生成差异化脚本手工数据补偿。此方法虽然简化了比对数据，但不支持对不同环境、不同数据库类型的数据进行差异比对，且缺少对源数据的预处理，对数据量较大的数据表比对效果较差，对于差异需要手动补偿占有额外人力。

[0065] 或者是，将第一数据副本与对应的第一历史数据副本进行比对，得到第一修改记录集；将第二数据副本与对应的第二历史数据副本进行比对，得到第二修改记录集；根据历史比对结果、第一修改记录集和第二修改记录集，得到第一数据副本与第二数据副本之间的数据比对结果。此方法虽克服了数据差异性，但采用数据副本的方式，需维护多个文件，后续当再次进行数据比对时，还需再次生成副本文件，并对源、副本文件进行比对分析，空间复杂度较高，对系统开销占用较大，大大增加了资源的浪费与空间的消耗。

[0066] 针对上述现有技术的缺点,发明人进行了相关研究,并提出了一种数据库一致性比对与补偿方法,该方法克服了现有技术中存在的技术问题,现参见图1所示的数据库一致性比对与补偿方法的流程图,对一种数据库一致性比对与补偿方法进行说明:

[0067] S101、获取多个待校验的数据库。

[0068] 实际应用中,可以选择同一企业的多个机房数据库作为待校验的数据库,从而实现对同一企业的多个机房数据库的比对和补偿,使得在该企业其中一个机房出现故障时,通过应用本申请技术方案提供的数据库一致性比对与补偿方法,可以将业务迅速切换至另一个机房数据库进行正常工作,而不会影响该业务的正常运行。

[0069] S102、按照主键对每个待校验的数据库中每条数据进行排序并计算每个排序后的数据库中每条数据的哈希值,得到多个预处理后的数据库。

[0070] 先确定每个待校验的数据库中每条数据的主键,再对待校验的数据库中的每条数据按照主键进行排序,保证了在后续比对和补偿时以数据库之间比对的基础是一致的,实现对不同环境、不同数据库类型的数据进行差异比对。

[0071] 计算排序后的数据库中每条数据的哈希值,是为了在比对过程中比对每条数据的哈希值便可,无需比对每条数据中的每个数据,避免在比对过程中占用过多资源,且本申请无需在比对过程中生成副本进行比对,降低对系统的开销占用,有效节省运行资源与存储空间。

[0072] S103、采用比对算法基于主键和哈希值对多个预处理后的数据库进行比对确定比对结果。

[0073] 本申请将比对结果分为两种,分别为第一结果或第二结果。若比对结果为第一结果,则表征多个预处理后的数据库之间没有差异,则无需进行数据库之间的补偿,因此则无需生成差异文件进行补偿。若比对结果为第二结果,则表征多个预处理后的数据库之间存在差异,则需要采用比对规则对多个预处理后的数据库进行比对,从而确定差异文件,后根据差异文件对多个预处理后的数据库进行数据补偿,因此第二结果需包括存在差异结论和比对信息,此差异结论为存在差异的结论,此比对信息则可以理解为差异文件中的比对信息,差异文件是基于比对信息构建的。

[0074] 为克服现有技术的问题,本申请比对对方式进行了改进,现对本申请的比对算法进行说明:

[0075] 将多个预处理后的数据库中的任意一个数据库作为标准数据库,将剩余数据库作为待比对数据库。

[0076] 本申请的比对算法采用了以一方为准的方式进行比对,以一方为准即以多个数据库中任意一个数据库为基准对其他数据库进行比对和补偿,因此本申请需选出标准数据库作为比对和补偿的基准,而本申请标准数据库的选择主要采用随机选择和自定义两种方式。

[0077] 若多个预处理后的数据库距离上次比对和补偿间隔时间较短或者多个预处理后的数据库之间存在差异的概率较小,则可以采用随机选择的方式确定标准数据库。

[0078] 若多个预处理后的数据库距离上次比对和补偿间隔时间较长或者多个预处理后的数据库之间存在差异的概率较大,则可以自定义任意一个数据库为标准数据库,例如将当前正在运行的数据库作为标准数据库。

[0079] 基于主键和哈希值比对所述标准数据库中每条数据与待比对数据库中每条数据生成比对结果。

[0080] 通过比对标准数据库中每条数据的主键与每个待比对数据库中每条数据库的主键,以及比对标准数据库中每条数据的哈希值与每个待比对数据库中每条数据库的哈希值生成比对结果。

[0081] 判断每个待比对数据库中每条数据的主键和哈希值是否均满足第一预设条件,得到第一判断结果。所述第一预设条件为待比对数据库中数据和标准数据库对应的数据的主键一致且哈希值一致。

[0082] 若所述第一判断结果为是,则表征多个待比对数据与标准数据库一致,则比对结果为第一结果。

[0083] 若满足第一预设条件,则表征每个待比对数据库中数据均与标准数据库数据一致,无需生成差异文件,则比对结果为第一结果。

[0084] 若所述第一判断结果为否,则表征比对结果为第二结果,基于比对规则基于主键和哈希值比对标准数据库中每条数据与待比对数据库中每条数据生成第二结果中的比对信息。

[0085] 本申请比对规则克服了现有技术中存在的占用资源过多的问题,现对本申请的比对规则进行说明:

[0086] 判断多个待比对数据库中不满足第一预设条件的待比对数据库中的数据是否满足第二预设条件,得到第二判断结果;所述第二预设条件为待比对数据库中数据和标准数据库对应的数据的主键一致且哈希值不一致。

[0087] 若所述第二判断结果为是,则将满足第二预设条件的待比对数据库中对应数据的比对信息记录为不一致。

[0088] 若所述第二判断结果为否,判断多个待比对数据库中不满足第一预设条件和第二预设条件的待比对数据库中的数据是否满足第三预设条件,得到第三判断结果;所述第三预设条件为待比对数据库中数据的主键小于标准数据库对应的数据的主键或待校验数据的数据总条数大于标准数据的数据总条数。

[0089] 若所述第三判断结果为是,则将满足第三预设条件的待比对数据库中对应数据的比对信息记录为冗余。

[0090] 若所述第三判断结果为否,则将不满足第三预设条件的待比对数据库中对应数据的比对信息记录为缺失。

[0091] 本申请的比对规则主要是通过比对标准数据库和待比对数据库中每条数据的主键和哈希值是否一致。

[0092] 若标准数据库中数据的主键和待比对数据库的主键一致而哈希值不一致,则说明待比对数据中存在数据不一致的问题,因此则需在比对信息中记录该待比对数据库中存在不一致。

[0093] 若待比对数据库的主键大于标准数据库中数据的主键或者待比对数据库的数据总条数大于标准数据库的数据总条数,则说明待比对数据库中存在数据冗余的问题,则需在比对信息中记录该待比对数据库中存在冗余。

[0094] 若待比对数据库的主键小于标准数据库中数据的主键或者待比对数据库的数据

总条数小于标准数据库的数据总条数,则说明待比对数据库中存在数据缺失的问题,则需在比对信息中记录该待比对数据库中存在缺失。

[0095] 本申请提供了比对算法的具体实施例,利用指针指向待比对数据和标准数据库进行比对,更加直观的展现比对算法,现参数图2所示的一致性比对与补偿方法的比对算法第一示意图和图3所示的数据库一致性比对与补偿方法的比对算法第二示意图,比对对算法进行说明:

[0096] 在主键有序的前提下,如图2所示,将标准数据库定义为图2所示的数据库1,将待比对数据库定义为图2所示的数据库2至数据库n,其中key为主键,value为哈希值,比对算法流程如图3所示:

[0097] (1)指针p1指向数据库1主键列表的主键key11,指针p2指向数据库2主键列表的主键key21,指针pn指向数据库n主键列表的主键keyn1,以数据库1为依据,逐表比对主键和该主键所对应哈希值是否相同。

[0098] (2)以指针p1指向的主键key11为依据,若某一表指针pi对应的主键keyi1小于指针p1的主键key11时,说明该条数据在数据库1中不存在,该数据库需进行数据删除操作,比对信息记录为冗余,并将指针pi向下移动。

[0099] (3)若指针pi指向主键keyi1与指针p1指向主键key11一致,且指针pi指向主键对应的哈希值valuei1与指针p1指向主键对应的哈希值value11一致,则说明数据一致,依次比较其他数据库指针所指向的主键。

[0100] (4)若指针pi指向主键keyi1与指针p1指向主键key11一致,但指针pi指向主键对应的哈希值valuei1与指针p1指向主键对应的哈希值value11不一致,则以数据库1为依据,该数据库需进行数据更新操作,比对结果记录为不一致,依次比较其他数据表指针所指向的主键。

[0101] (5)若指针pi指向主键keyi1大于指针p1指向主键key11时,说明该条数据在该数据库中存在缺失情况,该数据表需进行数据新增操作,比对结果记录为缺失,依次比较其他数据表指针所指向的主键。

[0102] (6)若此轮已经比较完最后一个数据库,则指针p1向下移动,重复(2)-(5)。

[0103] (7)若指针p1已遍历完数据表1,指针pi已遍历完数据表i,则需对数据库中第i+1条到数据中第n条剩余未比较的数据进行删除操作,比对信息记录为冗余。

[0104] (8)若指针p1已遍历完数据表1,其他数据库表指针均已遍历完成,则比对流程结束。

[0105] S104、基于比对信息生成差异文件,并基于差异文件对多个预处理后的数据库进行补偿,得到多个补偿后的数据库。

[0106] 本申请的具体补偿过程如下:

[0107] 解析差异文件,得到解析文件。

[0108] 基于解析文件确定每个待比对数据库中每条数据的补偿方式,并基于预设补偿方式对每个待比对数据库进行补偿,得到多个补偿后的数据库。

[0109] 补偿方式包括新增、更新和删除。

[0110] 基于所述解析文件,判断每个待比对数据库中每条数据是否存在不一致情况,得到第四判断结果。

[0111] 若所述第四判断结果为是,则基于标准数据库对不一致的数据进行更新。

[0112] 更新表征待比对的数据库中存在与标准数据库不一致的数据,则需要基于标准数据库对存在不一致的数据库进行更新,即删除不一致数据并将标准数据库中对应的数据复制粘贴至删除部分。

[0113] 若所述第四判断结果为否,则判断每个待比对数据库中每条数据是否存在冗余情况,得到第五判断结果。

[0114] 若所述第五判断结果为是,则删除冗余的数据。

[0115] 删除表征待比对的数据库除了与标准数据库一致的内容外,还存在标准数据库中不存在的数据,则需要对此数据删除。

[0116] 若所述第五判断结果为否,表征多个待比对数据库中存在缺失,则基于标准数据库对缺失的数据进行新增。

[0117] 新增为基于标准数据库对缺失数据的数据库进行数据的补充,即在标准数据库中寻找与缺失数据的数据库中存在缺失的部分对应的部分的数据,并将此部分的数据复制粘贴至缺失数据的数据库中。

[0118] 在具体应用时,数据库一致性比对与补偿方法,还可以包括:

[0119] 校验用户权限,当收到用户数据一致性比对配置请求时,校验用户应用权限。

[0120] 只有当用户有应用权限时,才可以进行数据库比对和补偿,可以保证数据库的安全性。

[0121] 设置数据一致性比对流程的执行时间,定期比对,确保多个数据库持续保持数据一致性,为比对与补偿正常运行提供可靠的数据支持。

[0122] 在数据源中灵活选择比对的数据库及数据实体。

[0123] 当监听到异常时,产生告警并记录异常日志并根据异常日志判断异常类型分析异常日志,关闭自动补偿流程结束补偿,方便网关系统运维支持人员排查问题,当没有监听到异常时,则结束。

[0124] 本申请支持比对多数据库之间的数据差异,通过比较哈希值的形式对数据进行一致性比对,检测效率高,存储成本低,内存占用率低,无需直接对源数据进行比对操作,当数据规模较大时对系统的开销占用也较小。支持不同环境、不同数据库类型差异比对,增加对源数据预处理操作,降低数据比对的成本,可根据生成的差异文件分情况对数据进行自动补偿,减少人力操作,增加了补偿操作的灵活性。无需存储数据副本,且无需重复对多表进行比对计算操作,减少对系统资源的浪费与空间的消耗,存储成本低,对系统开销占用较小。

[0125] 本申请还提供了一种数据库一致性比对与补偿装置,现参见图4所示的数据库一致性比对与补偿装置的结构图,对数据库一致性比对与补偿装置进行说明:

[0126] 获取单元,用于获取多个待校验的数据库。

[0127] 预处理单元,用于按照主键对每个待校验的数据库中每条数据进行排序并计算每个排序后的数据库中每条数据的哈希值,得到多个预处理后的数据库。

[0128] 预处理单元包括数据源处理器和排序器。

[0129] 数据源处理器,用于将输入数据进行预处理,解决因不同数据库之间的数据模型、数据存储、数据排序等方面存在差异导致的相同的数据处理及表示方式不同的问题,提高

数据质量。

[0130] 排序器,用于对预处理后的数据库按照主键排序与哈希计算进行预处理。

[0131] 比对单元,用于采用比对算法基于主键和哈希值对多个预处理后的数据库进行比对确定比对结果。

[0132] 所述比对结果为第一结果或第二结果;所述第一结果表征多个预处理后的数据库无差异;所述第二结果表征多个预处理后的数据库存在差异;所述第二结果包括存在差异结论和比对信息。

[0133] 比对单元包括差异文件生成器和比对器。

[0134] 比对器,用于执行数据比对流程,以一方数据库为准,依据比对算法进行数据一致性校验。

[0135] 差异文件生成器,用于在比对结果为第二结果时,记录多库之间的差异数据并生成差异文件。

[0136] 比对器包括分类模块和比对模块,现对分类模块和比对模块的功能进行说明:

[0137] 分类模块,用于将多个预处理后的数据库中的任意一个数据库作为标准数据库,将剩余数据库作为待比对数据库。

[0138] 比对模块,用于基于主键和哈希值比对所述标准数据库中每条数据与待比对数据库中每条数据生成比对结果。

[0139] 比对模块包括第一判断一级模块和第二判断一级模块,现对第一判断一级模块和第二判断一级模块的功能进行说明:

[0140] 第一判断一级模块,用于判断每个待比对数据库中每条数据的主键和哈希值是否均满足第一预设条件,得到第一判断结果。所述第一预设条件为待比对数据库中数据和标准数据库对应的数据的主键一致且哈希值一致。若所述第一判断结果为是,则表征多个待比对数据与标准数据库一致,则比对结果为第一结果。

[0141] 第二判断一级模块,用于若所述第一判断结果为否,则表征比对结果为第二结果,基于比对规则基于主键和哈希值比对标准数据库中每条数据与待比对数据库中每条数据生成第二结果中的比对信息。

[0142] 第二判断一级模块还包括第一判断二级模块和第二判断二级模块,现对第一判断二级模块和第二判断二级模块的功能进行说明:

[0143] 第一判断二级模块,用于判断多个待比对数据库中不满足第一预设条件的待比对数据库中的数据是否满足第二预设条件,得到第二判断结果;若所述第二判断结果为是,则将满足第二预设条件的待比对数据库中对应数据的比对信息记录为不一致。

[0144] 所述第二预设条件为待比对数据库中数据和标准数据库对应的数据的主键一致且哈希值不一致。

[0145] 第二判断二级模块,用于若所述第二判断结果为否,判断多个待比对数据库中不满足第一预设条件和第二预设条件的待比对数据库中的数据是否满足第三预设条件,得到第三判断结果;若所述第三判断结果为是,则将满足第三预设条件的待比对数据库中对应数据的比对信息记录为冗余。若所述第三判断结果为否,则将不满足第三预设条件的待比对数据库中对应数据的比对信息记录为缺失。

[0146] 所述第三预设条件为待比对数据库中数据的主键小于标准数据库对应的数据的

主键或待校验数据的数据总条数大于标准数据的数据总条数。

[0147] 预处理单元和比对单元的工作流程如图5所示,数据源处理器对选定的比对数据库进行预处理,排序器对预处理后的数据库按主键排序,对排序后的数据库进行哈希计算,对数据库进行比对,生成比对结果。

[0148] 补偿单元,用于基于比对信息生成差异文件,并基于差异文件对多个预处理后的数据库进行补偿,得到多个补偿后的数据库。

[0149] 补偿单元包括差异文件解析器和数据补偿器。

[0150] 补偿单元,用于在比对出数据库间数据不一致之后,执行数据补偿操作。首先解析差异文件,根据差异清单中记录的不同差异类型执行对应操作,保证数补偿的正确性。

[0151] 补偿单元的工作流程如图6所示,解析生成的差异文件,进行自动数据补偿,补偿过程中是否发生异常,若是则生成问题清单,进行异常分析,若否则结束。

[0152] 差异文件解析器,用于解析差异文件,得到解析文件,数据补偿的操作类型根据比对流程可以分为“增删改”三类。

[0153] 数据补偿器,用于根据差异文件自动完成补偿过程。基于解析文件确定每个待比对数据库中每条数据的补偿方式,并基于预设补偿方式对每个待比对数据库进行补偿,得到多个补偿后的数据库。

[0154] 数据补偿器包括第一补偿模块、第二补偿模块和第三补偿模块,现对第一补偿模块、第二补偿模块和第三补偿模块的功能进行说明:

[0155] 第一补偿模块,用于基于所述解析文件,判断每个待比对数据库中每条数据是否存在不一致情况,得到第四判断结果。若所述第四判断结果为是,则基于标准数据库对不一致的数据进行更新。

[0156] 第二补偿模块,用于若所述第四判断结果为否,则判断每个待比对数据库中每条数据是否存在冗余情况,得到第五判断结果。若所述第五判断结果为是,则删除冗余的数据。

[0157] 第三补偿模块,用于若所述第五判断结果为否,表征多个待比对数据库中存在缺失,则基于标准数据库对缺失的数据进行新增。

[0158] 在数据补偿器工作过程中可能会出现补偿错误的问题,因此在具体应用时,数据库一致性比对与补偿装置还可以根据实际情况设置控制单元和异常监控单元。通过控制单元对整个装置进行控制。通过异常监控单元对补偿过程进行异常监控单元,并在发现异常情况时发送的异常告警信息,中止数据补偿过程,进行异常分析,待异常解决后,再次开启数据补偿过程。

[0159] 现参见图4所示的数据库一致性比对与补偿装置的结构图,对控制单元和异常监控单元进行说明:

[0160] 控制器包括权限校验器、实体选择器、定时器和补偿启停器。

[0161] 控制单元主要用于协调控制整个数据库一致性比对与补偿装置,通过权限校验器的用户配置定时器开启时间、选择数据一致性比对的数据库及实体类型,在数据自动补偿过程中接收异常发生时产生的异常告警信息,补偿启停器中断自动补偿过程。

[0162] 权限控制器用于校验用户权限,当收到用户数据一致性比对配置请求时,由权限校验器校验用户应用权限。

[0163] 定时器用于控制数据一致性比对流程的执行时间,定期比对,能够确保多地数据库持续保持数据一致性,为数据库一致性比对与补偿装置正常运行提供可靠的数据支持。

[0164] 实体选择器用于在数据源中灵活选择比对的数据库及数据实体。

[0165] 补偿启停器主要用于开启/停止数据自动补偿流程,当补偿过程中接收到异常监控单元发送的异常告警信息时,补偿启停器停止数据补偿,以防止在数据同步进程中,存在数据未同步成功,造成多库间数据不一致。

[0166] 控制单元的工作流程图如图7所示,用户权限控制,实体选择器选择比对的数据及实体,定时器开启数据比对,执行数据补偿过程中产生异常告警,补偿启停器停止数据补偿,异常分析完成后,补偿启停器开始数据补偿。

[0167] 异常监控单元包括异常分析器、异常告警器和异常监听器。

[0168] 异常监控单元用于实时监控数据补偿的执行情况,当监听到有异常发生时将产生异常告警,补偿启停器中断自动补偿过程,同时记录异常原因生成问题清单,便于错误排查处理。

[0169] 异常监控单元实时监听执行自动补偿时的异常情况,异常监控单元的工作流程如图8所示。

[0170] 异常监控单元与控制单元同步启动,异常监听器是否监听到异常,当监听到异常时,产生告警并记录异常日志并根据异常日志判断异常类型,控制单元接收到异常告警并分析异常日志,由补偿启停器关闭自动补偿流程结束补偿,方便网关系统运维支持人员排查问题,当没有监听到异常时,则结束。

[0171] 数据库一致性比对与补偿装置的工作流程如图9所示:

[0172] 控制单元控制数据比对定时器开启时间,通过权限校验的用户选择需要进行一致性比对的数据源及实体类型。

[0173] 预处理单元首先对输入数据进行预处理操作,避免由于各数据库产品对特殊值的存储方式不同及不同数据库数据排序不一致被误判定为差异数据的问题导致输入参数不一致,然后对预处理后的数据库按照主键排序并计算哈希值解决不同数据库间因数据排序不同被误判定为差异数据的问题。

[0174] 比对单元执行数据比对流程,以一方数据库为准,依据比对规则进行数据一致性校验,记录多库之间的差异数据并形成差异文件。

[0175] 补偿单元解析生成的差异文件,自动执行数据补偿流程,实现多数据库间数据一致性补偿修复。

[0176] 异常监控单元实时监听数据补偿的执行情况,当监听到有异常发生时,异常告警器产生异常告警,向控制单元发送错误信息,补偿启停器中断补偿过程并生成问题清单,对问题清单进行异常分析,判断其异常类型执行后续相应操作。

[0177] 一种计算机设备,包括:存储器、处理器以及存储在存储器上并可在处理器上运行的计算机程序,所述处理器执行所述计算机程序以实现数据库一致性比对与补偿方法。

[0178] 一种计算机可读存储介质,其上存储有计算机程序,该计算机程序被处理器执行时实现数据库一致性比对与补偿方法。

[0179] 一种计算机程序产品,包括计算机程序,该计算机程序被处理器执行时实现数据库一致性比对与补偿方法。

[0180] 需要说明的是,本说明书中的各个实施例均采用递进的方式描述,各个实施例之间相同相似的部分互相参见即可,每个实施例重点说明的都是与其他实施例的不同之处。尤其,对于设备及系统实施例而言,由于其基本相似于方法实施例,所以描述得比较简单,相关之处参见方法实施例的部分说明即可。以上所描述的设备及系统实施例仅仅是示意性的,其中作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元提示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。本领域普通技术人员在不付出创造性劳动的情况下,即可以理解并实施。

[0181] 以上所述,仅为本申请的一种具体实施方式,但本申请的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本申请揭露的技术范围内,可轻易想到的变化或替换,都应涵盖在本申请的保护范围之内。因此,本申请的保护范围应该以权利要求的保护范围为准。

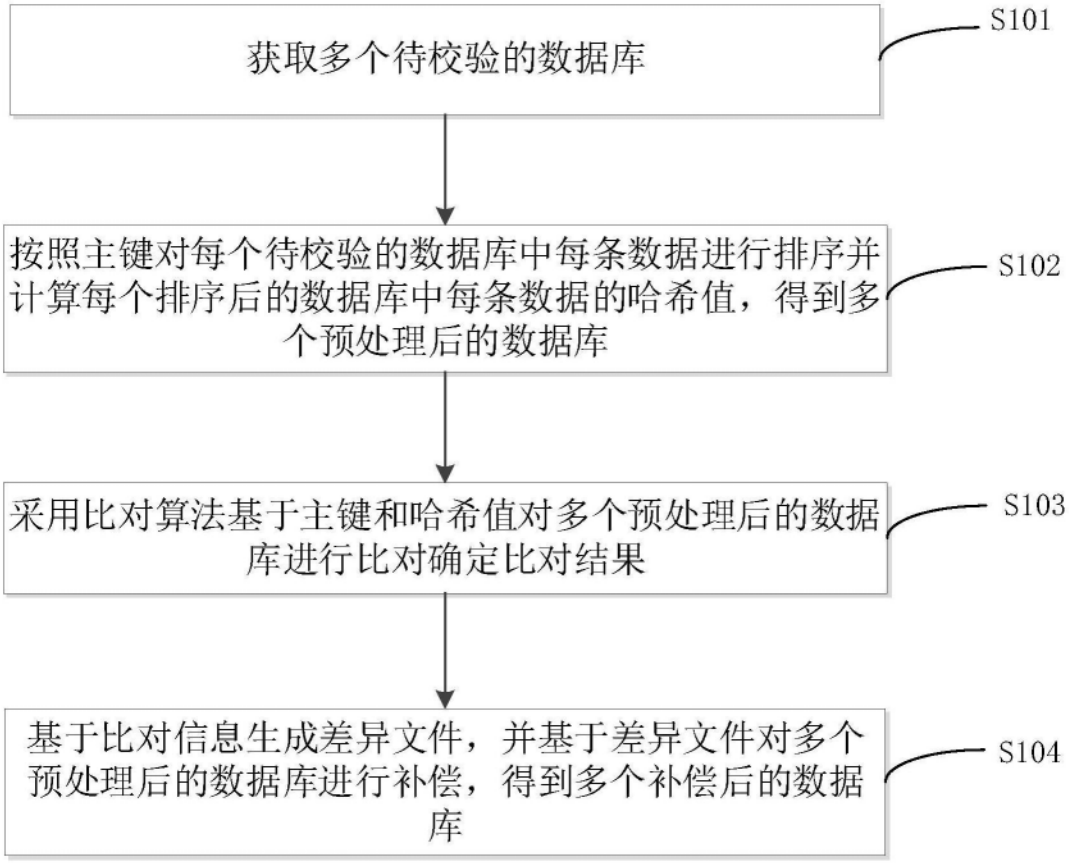

获取多个待校验的数据库
S101
按照主键对每个待校验的数据库中每条数据进行排序并计算每个排序后的数据库中每条数据的哈希值，得到多个预处理后的数据库
S102
采用比对算法基于主键和哈希值对多个预处理后的数据库进行比对确定比对结果
S103
基于比对信息生成差异文件，并基于差异文件对多个预处理后的数据库进行补偿，得到多个补偿后的数据库
S104

图1

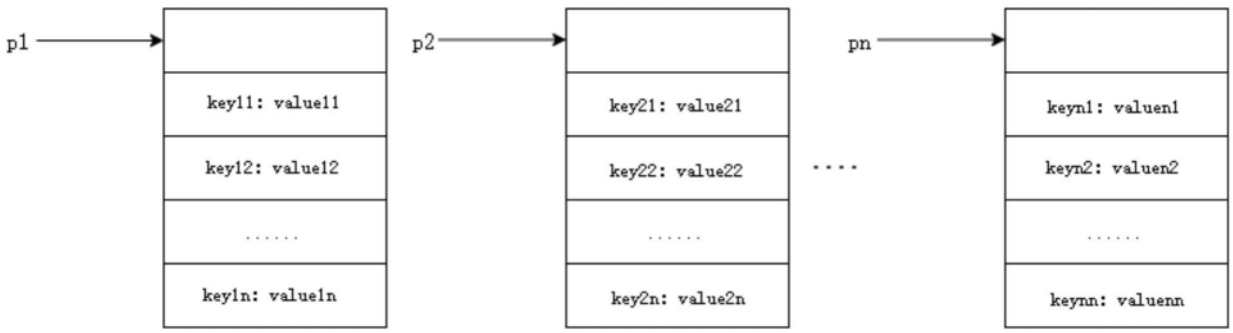

p1
key11: value11
key12: value12
.....
key1n: value1n
p2
key21: value21
key22: value22
.....
key2n: value2n
....
pn
keyn1: valuen1
keyn2: valuen2
.....
keynn: valuenn

图2

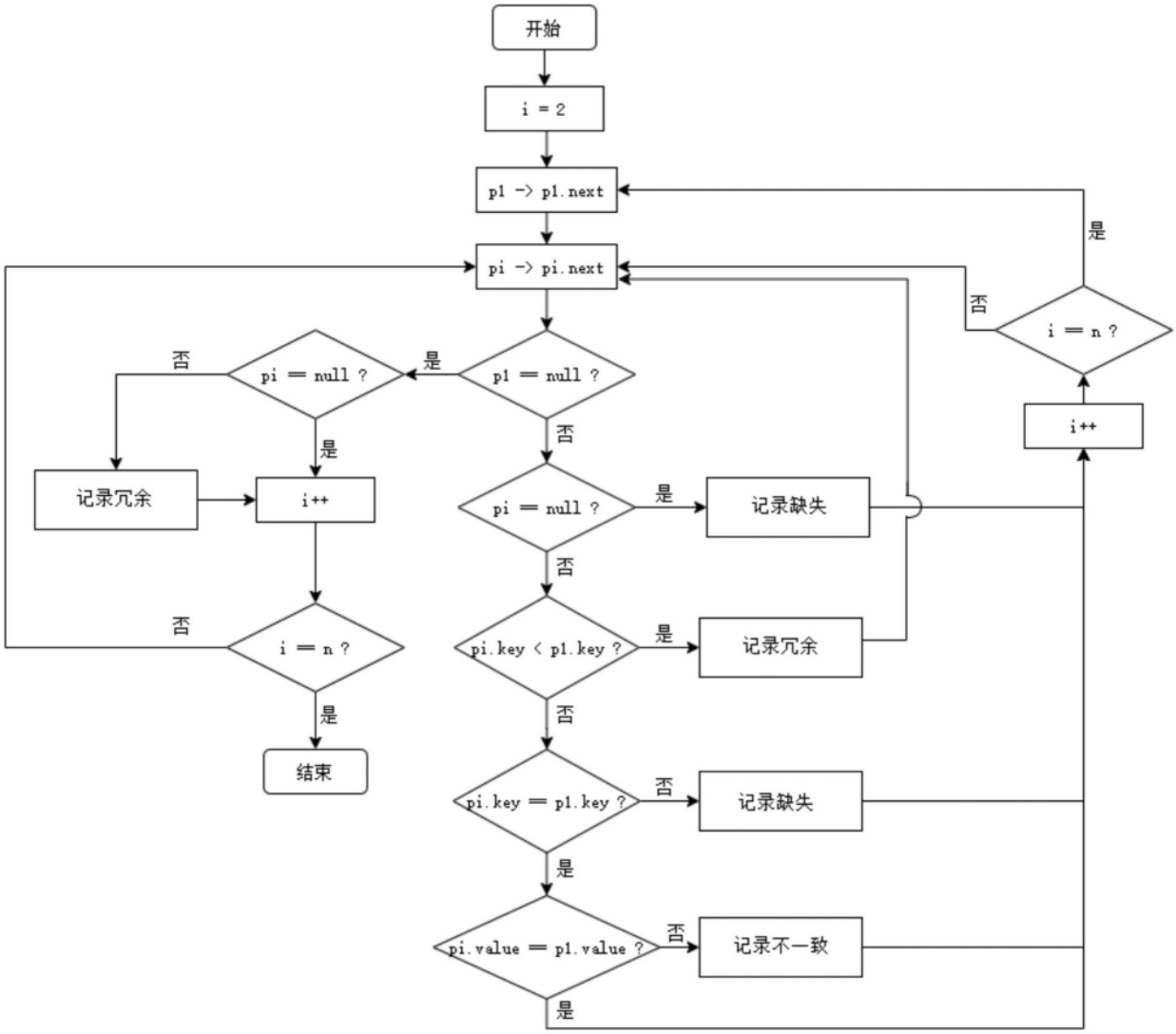

开始
i = 2
p1 -> p1.next
pi -> pi.next
p1 == null ?
是
否
pi == null ?
是
否
记录冗余
i++
pi == null ?
是
记录缺失
否
i == n ?
是
否
pi.key < p1.key ?
是
记录冗余
否
结束
pi.key == p1.key ?
否
记录缺失
是
pi.value == p1.value ?
否
记录不一致
是
i == n ?
是
否
i++

图3

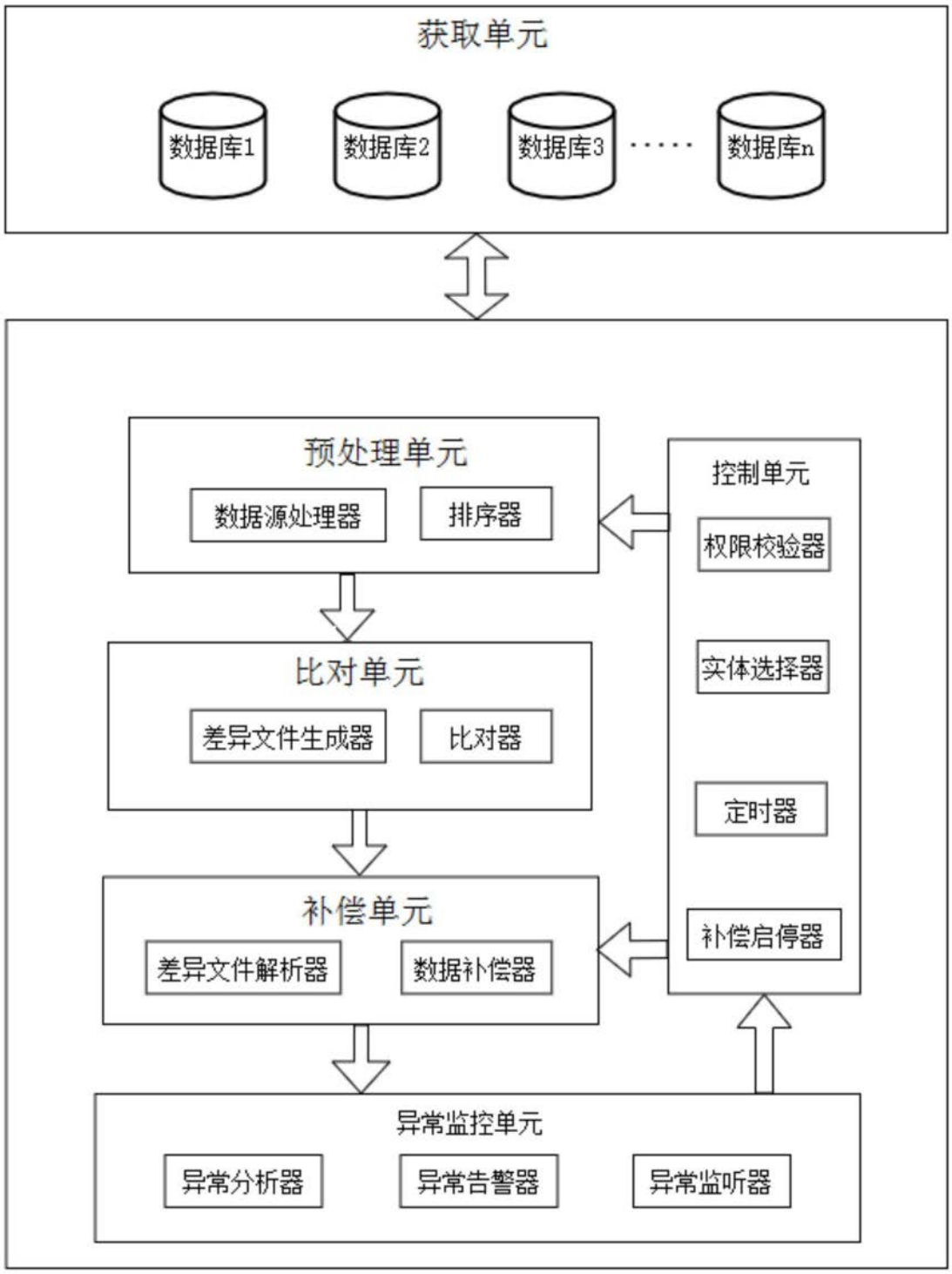

获取单元
数据库1
数据库2
数据库3
.....
数据库n
预处理单元
数据源处理器
排序器
比对单元
差异文件生成器
比对器
补偿单元
差异文件解析器
数据补偿器
控制单元
权限校验器
实体选择器
定时器
补偿启停器
异常监控单元
异常分析器
异常告警器
异常监听器

图4

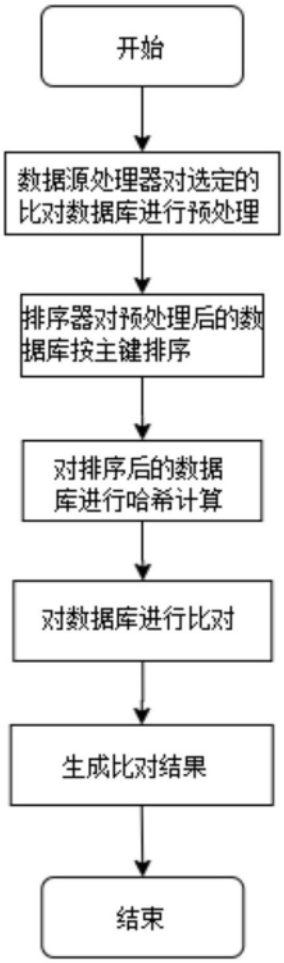

开始
数据源处理器对选定的比对数据库进行预处理
排序器对预处理后的数据库按主键排序
对排序后的数据库进行哈希计算
对数据库进行比对
生成比对结果
结束

图5

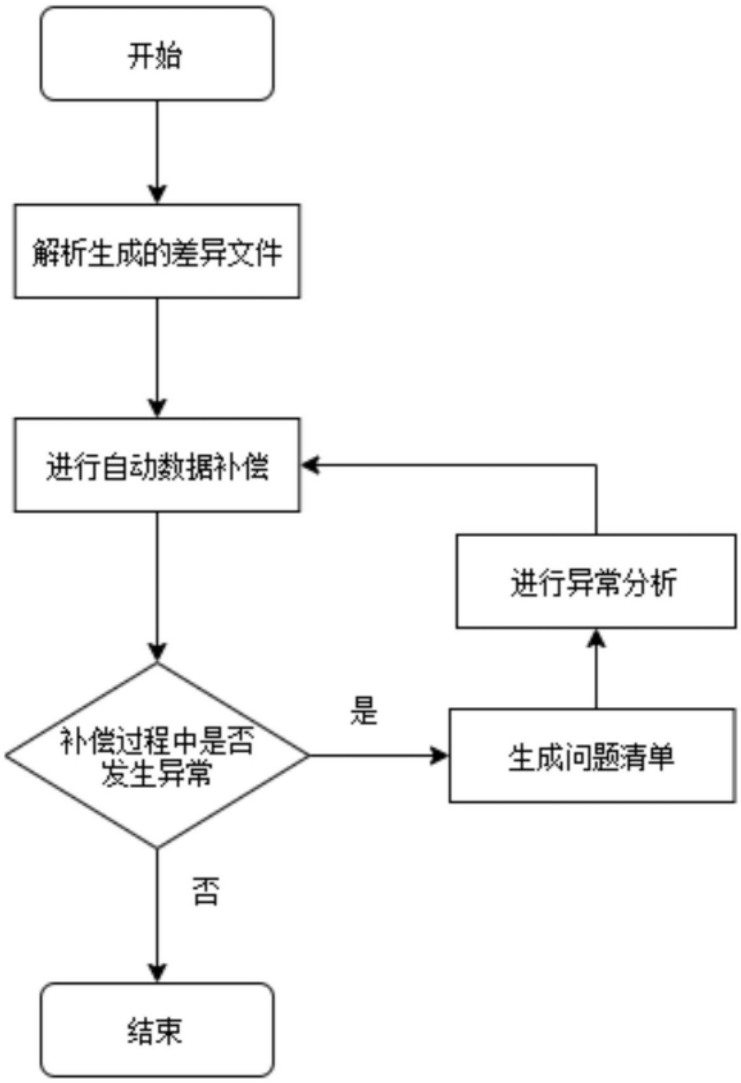


图6

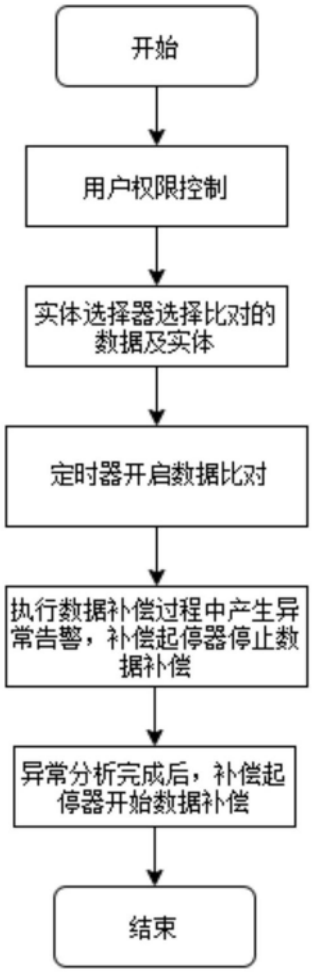

开始
用户权限控制
实体选择器选择比对的数据及实体
定时器开启数据比对
执行数据补偿过程中产生异常告警，补偿起停器停止数据补偿
异常分析完成后，补偿起停器开始数据补偿
结束

图7

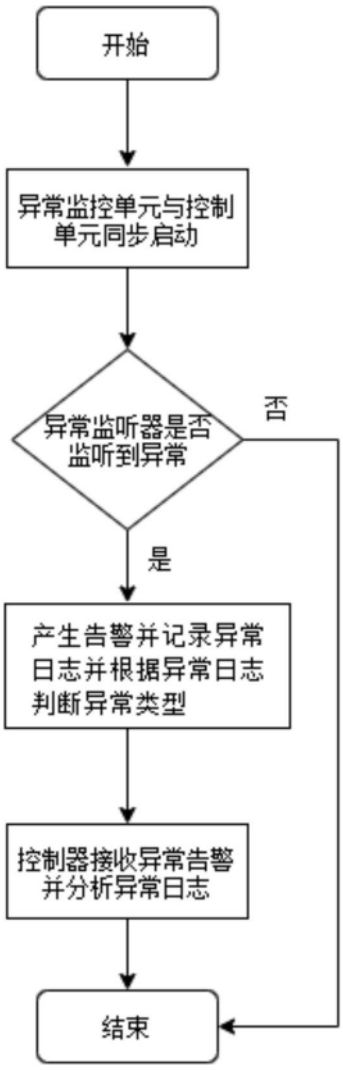

开始
异常监控单元与控制单元同步启动
异常监听器是否监听到异常
否
是
产生告警并记录异常日志并根据异常日志判断异常类型
控制器接收异常告警并分析异常日志
结束

图8

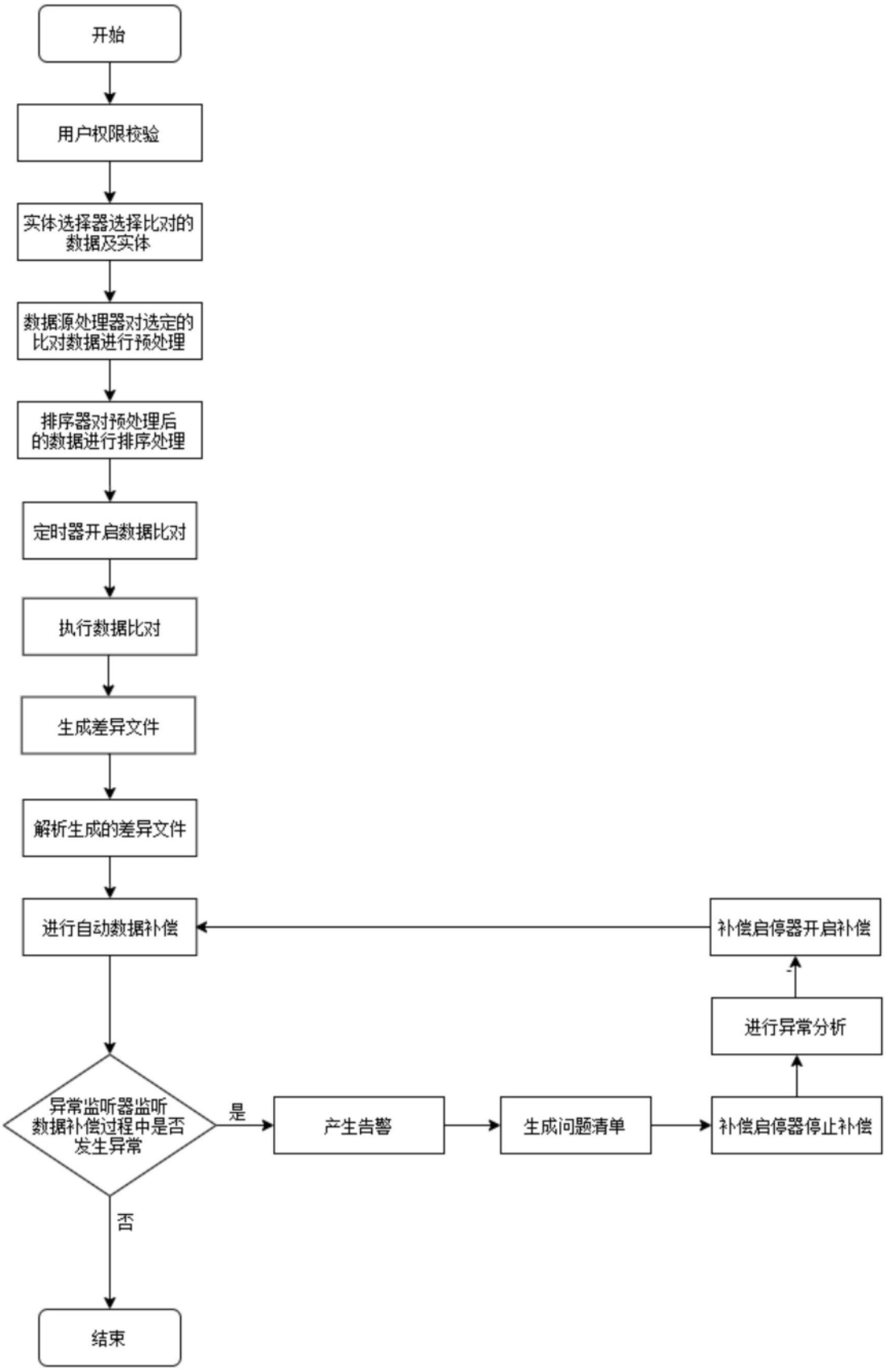

开始
用户权限校验
实体选择器选择比对的数据及实体
数据源处理器对选定的比对数据进行预处理
排序器对预处理后的数据进行排序处理
定时器开启数据比对
执行数据比对
生成差异文件
解析生成的差异文件
进行自动数据补偿
补偿启停器开启补偿
进行异常分析
异常监听器监听数据补偿过程中是否发生异常
是
产生告警
生成问题清单
补偿启停器停止补偿
否
结束

图9