



(12) 发明专利申请

(10) 申请公布号 CN 115168353 A

(43) 申请公布日 2022. 10. 11

(21) 申请号 202210796591.2

(22) 申请日 2022.07.06

(71) 申请人 北京汇通天下物联科技有限公司
地址 100082 北京市海淀区建材城中路27号8幢二层201号

(72) 发明人 蒲海洋

(74) 专利代理机构 北京超凡宏宇专利代理事务
所(特殊普通合伙) 11463
专利代理师 曹瑞敏

(51) Int.Cl.

G06F 16/22 (2019.01)

G06F 16/2455 (2019.01)

G06F 16/27 (2019.01)

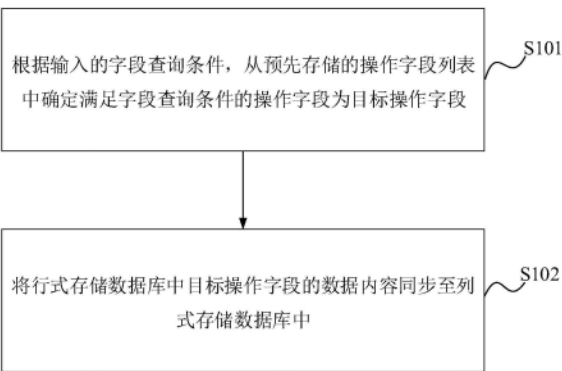
权利要求书2页 说明书10页 附图4页

(54) 发明名称

数据同步方法、装置、设备及存储介质

(57) 摘要

本申请提供了一种数据同步方法、装置、设备及存储介质,其中,该方法包括:根据输入的字段查询条件,从预先存储的操作字段列表中确定满足字段查询条件的操作字段为目标操作字段,其中,操作字段列表中包括:多个操作字段,每个操作字段为预先从针对行式存储数据库的多个数据库操作语句中提取的字段,将行式存储数据库中目标操作字段的数据内容同步至列式存储数据库中。在本申请中,根据输入的字段查询条件从操作字段列表中确定同步至列式存储数据库的目标操作字段,灵活性高,与实际场景相符,并且操作字段列表是基于实际场景从数据库操作语句中提取的字段,为行式存储数据库中的数据同步至列式存储数据库提供了一种可参考依据。



1. 一种数据同步方法,其特征在于,包括:

根据输入的字段查询条件,从预先存储的操作字段列表中确定满足字段查询条件的操作字段为目标操作字段,其中,所述操作字段列表中包括:多个操作字段,每个操作字段为预先从针对行式存储数据库的多个数据库操作语句中提取的字段;

将所述行式存储数据库中所述目标操作字段的数据内容同步至列式存储数据库中。

2. 根据权利要求1所述的方法,其特征在于,所述根据输入的字段查询条件,从预先存储的操作字段列表中确定满足字段查询条件的操作字段为目标操作字段之前,所述方法还包括:

获取预设时间段内针对所述行式存储数据库的多个数据库操作语句;

根据所述行式存储数据库的数据库类型,分别对所述多个数据库操作语句进行解析处理,得到所述操作字段列表。

3. 根据权利要求2所述的方法,其特征在于,所述根据所述行式存储数据库的数据库类型,分别对所述多个数据库操作语句进行解析处理,得到所述操作字段列表,包括:

获取所述数据库类型对应的词法文件和语法规则文件;

根据所述词法文件和所述语法规则文件,生成解析处理代码;

根据所述解析处理代码,分别对所述多个数据库操作语句进行解析处理,得到所述操作字段列表。

4. 根据权利要求1所述的方法,其特征在于,所述字段查询条件包括:查询时间段;

所述根据输入的字段查询条件,从预先存储的操作字段列表中确定满足字段查询条件的操作字段为目标操作字段,包括:

从所述操作字段列表中确定字段操作时间满足所述查询时间段的操作字段为所述目标操作字段,所述字段操作时间为所述操作字段列表中各操作字段对应的数据库操作语句的操作时间。

5. 根据权利要求1所述的方法,其特征在于,所述字段查询条件包括:字段置信度;

所述根据输入的字段查询条件,从预先存储的操作字段列表中确定满足字段查询条件的操作字段为目标操作字段,包括:

从所述操作字段列表中确定满足所述字段置信度的操作字段为所述目标操作字段。

6. 根据权利要求1所述的方法,其特征在于,所述字段查询条件包括:所述行式存储数据库中目标数据表的标识;

所述根据输入的字段查询条件,从预先存储的操作字段列表中确定满足字段查询条件的操作字段为目标操作字段,包括:

从所述操作字段列表中确定所述目标数据表对应的操作字段为所述目标操作字段。

7. 根据权利要求1所述的方法,其特征在于,所述字段查询条件包括:所述行式存储数据库对应的目标应用的标识;

所述根据输入的字段查询条件,从预先存储的操作字段列表中确定满足字段查询条件的操作字段为目标操作字段,包括:

从所述操作字段列表中确定所述目标应用对应的操作字段为所述目标操作字段。

8. 一种数据同步装置,其特征在于,包括:

确定模块,用于根据输入的字段查询条件,从预先存储的操作字段列表中确定满足字

段查询条件的操作字段为目标操作字段,其中,所述操作字段列表包括:多个操作字段,每个操作字段为预先从针对行式存储数据库的数据库操作语句中提取的字段;

同步模块,用于将所述行式存储数据库中所述目标操作字段的数据内容同步至列式存储数据库中。

9.一种电子设备,其特征在于,包括:处理器、存储器和总线,所述存储器存储有所述处理器可执行的机器可读指令,当电子设备运行时,所述处理器与所述存储器之间通过总线通信,所述处理器执行所述机器可读指令,以执行权利要求1至7中任一项所述的数据同步方法。

10.一种计算机可读存储介质,其特征在于,所述计算机可读存储介质上存储有计算机程序,所述计算机程序被处理器运行时执行权利要求1至7中任一项所述的数据同步方法。

数据同步方法、装置、设备及存储介质

技术领域

[0001] 本申请涉及计算机技术领域,具体而言,涉及一种数据同步方法、装置、设备及存储介质。

背景技术

[0002] 在快速迭代的业务系统中,数据通常存放在关系型数据库中,如mysql、oracle、sql server等数据库,随着业务发展以及需求持续变更,数据库的数据表以及数据表中的字段量越来越多,其中,关系型数据库采用行式存储进行数据存储。

[0003] 现有技术中,从行式存储数据库中提取字段的数据内容时,处理速度较慢,特别是在对字段的数据内容进行统计分析时,由于行式存储数据库的自身特性,可能给业务系统的数据库带来压力瓶颈问题,为提高数据统计性能,往往需要将行式存储数据库的部分字段的数据内容同步至列式存储数据库中,以从列式存储数据库中提取数据进行统计分析。

[0004] 然而,在上述字段的数据内容同步过程中,仅仅依靠人工确定同步字段,缺乏一定依据,如果字段选择不合适,还可能造成更多工作开销。

发明内容

[0005] 有鉴于此,本申请实施例提供了一种数据同步方法、装置、设备及存储介质,以解决现有技术中人工确定同步字段,缺乏依据,且字段选择不合适,造成更多工作开销的问题。

[0006] 第一方面,本申请实施例提供了一种数据同步方法,包括:

[0007] 根据输入的字段查询条件,从预先存储的操作字段列表中确定满足字段查询条件的操作字段为目标操作字段,其中,所述操作字段列表中包括:多个操作字段,每个操作字段为预先从针对行式存储数据库的多个数据库操作语句中提取的字段;

[0008] 将所述行式存储数据库中所述目标操作字段的数据内容同步至列式存储数据库中。

[0009] 在一可选的实施方式中,所述根据输入的字段查询条件,从预先存储的操作字段列表中确定满足字段查询条件的操作字段为目标操作字段之前,所述方法还包括:

[0010] 获取预设时间段内针对所述行式存储数据库的多个数据库操作语句;

[0011] 根据所述行式存储数据库的数据库类型,分别对所述多个数据库操作语句进行解析处理,得到所述操作字段列表。

[0012] 在一可选的实施方式中,所述根据所述行式存储数据库的数据库类型,分别对所述多个数据库操作语句进行解析处理,得到所述操作字段列表,包括:

[0013] 获取所述数据库类型对应的词法文件和语法规则文件;

[0014] 根据所述词法文件和所述语法规则文件,生成解析处理代码;

[0015] 根据所述解析处理代码,分别对所述多个数据库操作语句进行解析处理,得到所

述操作字段列表。

[0016] 在一可选的实施方式中,所述字段查询条件包括:查询时间段;

[0017] 所述根据输入的字段查询条件,从预先存储的操作字段列表中确定满足字段查询条件的操作字段为目标操作字段,包括:

[0018] 从所述操作字段列表中确定字段操作时间满足所述查询时间段的操作字段为所述目标操作字段,所述字段操作时间为所述操作字段列表中各操作字段对应的数据库操作语句的操作时间。

[0019] 在一可选的实施方式中,所述字段查询条件包括:字段置信度;

[0020] 所述根据输入的字段查询条件,从预先存储的操作字段列表中确定满足字段查询条件的操作字段为目标操作字段,包括:

[0021] 从所述操作字段列表中确定满足所述字段置信度的操作字段为所述目标操作字段。

[0022] 在一可选的实施方式中,所述字段查询条件包括:所述行式存储数据库中目标数据表的标识;

[0023] 所述根据输入的字段查询条件,从预先存储的操作字段列表中确定满足字段查询条件的操作字段为目标操作字段,包括:

[0024] 从所述操作字段列表中确定所述目标数据表对应的操作字段为所述目标操作字段。

[0025] 在一可选的实施方式中,所述字段查询条件包括:所述行式存储数据库对应的目标应用的标识;

[0026] 所述根据输入的字段查询条件,从预先存储的操作字段列表中确定满足字段查询条件的操作字段为目标操作字段,包括:

[0027] 从所述操作字段列表中确定所述目标应用对应的操作字段为所述目标操作字段。

[0028] 第二方面,本申请实施例还提供了一种数据同步装置,包括:

[0029] 确定模块,用于根据输入的字段查询条件,从预先存储的操作字段列表中确定满足字段查询条件的操作字段为目标操作字段,其中,所述操作字段列表中包括:多个操作字段,每个操作字段为预先从针对行式存储数据库的数据库操作语句中提取的字段;

[0030] 同步模块,用于将所述行式存储数据库中所述目标操作字段的数据内容同步至列式存储数据库中。

[0031] 在一可选的实施方式中,所述装置还包括:

[0032] 获取模块,用于获取预设时间段内针对所述行式存储数据库的多个数据库操作语句;

[0033] 处理模块,用于根据所述行式存储数据库的数据库类型,分别对所述多个数据库操作语句进行解析处理,得到所述操作字段列表。

[0034] 在一可选的实施方式中,所述处理模块具体用于:

[0035] 获取所述数据库类型对应的词法文件和语法规则文件;

[0036] 根据所述词法文件和所述语法规则文件,生成解析处理代码;

[0037] 根据所述解析处理代码,分别对所述多个数据库操作语句进行解析处理,得到所述操作字段列表。

- [0038] 在一可选的实施方式中,所述字段查询条件包括:查询时间段;
- [0039] 所述确定模块,具体用于:
- [0040] 从所述操作字段列表中确定字段操作时间满足所述查询时间段的操作字段为所述目标操作字段,所述字段操作时间为所述操作字段列表中各操作字段对应的数据库操作语句的操作时间。
- [0041] 在一可选的实施方式中,所述字段查询条件包括:字段置信度;
- [0042] 所述确定模块,具体用于:
- [0043] 从所述操作字段列表中确定满足所述字段置信度的操作字段为所述目标操作字段。
- [0044] 在一可选的实施方式中,所述字段查询条件包括:所述行式存储数据库中目标数据表的标识;
- [0045] 所述确定模块,具体用于:
- [0046] 从所述操作字段列表中确定所述目标数据表对应的操作字段为所述目标操作字段。
- [0047] 在一可选的实施方式中,所述字段查询条件包括:所述行式存储数据库对应的目标应用的标识;
- [0048] 所述确定模块,具体用于:
- [0049] 从所述操作字段列表中确定所述目标应用对应的操作字段为所述目标操作字段。
- [0050] 第三方面,本申请实施例还提供了一种电子设备,包括:处理器、存储器和总线,所述存储器存储有所述处理器可执行的机器可读指令,当电子设备运行时,所述处理器与所述存储器之间通过总线通信,所述处理器执行所述机器可读指令,以执行第一方面中任一项所述的数据同步方法。
- [0051] 第四方面,本申请实施例还提供了一种计算机可读存储介质,所述计算机可读存储介质上存储有计算机程序,所述计算机程序被处理器运行时执行第一方面中任一项所述的数据同步方法。
- [0052] 本申请提供了一种数据同步方法、装置、设备及存储介质,其中,该方法包括:根据输入的字段查询条件,从预先存储的操作字段列表中确定满足字段查询条件的操作字段为目标操作字段,其中,操作字段列表中包括:多个操作字段,每个操作字段为预先从针对行式存储数据库的多个数据库操作语句中提取的字段,将行式存储数据库中目标操作字段的数据内容同步至列式存储数据库中。在本申请中,根据输入的字段查询条件从操作字段列表中确定同步至列式存储数据库的目标操作字段,灵活性高,与实际场景相符,并且操作字段列表是基于实际场景从数据库操作语句中提取的字段,为行式存储数据库中的数据同步至列式存储数据库提供了一种可参考依据。
- [0053] 为使本申请的上述目的、特征和优点能更明显易懂,下文特举较佳实施例,并配合所附附图,作详细说明如下。

附图说明

- [0054] 为了更清楚地说明本申请实施例的技术方案,下面将对实施例中所需要使用的附图作简单地介绍,应当理解,以下附图仅示出了本申请的某些实施例,因此不应被看作是对

范围的限定,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他相关的附图。

[0055] 图1为本申请实施例提供的数据同步方法的流程示意图一;

[0056] 图2为本申请实施例提供的数据同步方法的流程示意图二;

[0057] 图3为本申请实施例提供的数据同步方法的流程示意图三;

[0058] 图4为本申请实施例提供的数据同步方法的流程示意图四;

[0059] 图5为本申请实施例提供的数据同步装置的结构示意图;

[0060] 图6为本申请实施例提供的电子设备的结构示意图。

具体实施方式

[0061] 为使本申请实施例的目的、技术方案和优点更加清楚,下面将结合本申请实施例中附图,对本申请实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本申请一部分实施例,而不是全部的实施例。通常在此处附图中描述和示出的本申请实施例的组件可以以各种不同的配置来布置和设计。因此,以下对在附图中提供的本申请的实施例的详细描述并非旨在限制要求保护的本申请的范围,而是仅仅表示本申请的选定实施例。基于本申请的实施例,本领域技术人员在没有做出创造性劳动的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。

[0062] 对于行式存储数据库而言,当行式存储数据库的数据表内的数据量很大时,从行式存储数据库中提取数据,需要指定行的整列数据提取出来,然后从整列数据中提取特定数据,处理速度较慢,还可能给业务系统的数据库带来瓶颈问题,特别是遇到统计分析场景时,处理速度更慢,为提高数据统计性能,往往需要使用列式存储数据库,如elasticsearch、mongodb数据库,这时,需要将行式存储数据库中的数据根据字段的需要,同步至列式存储数据库中,在同步时需要行式存储数据库中的部分字段进行选择性同步,目前,仅仅依靠人工来评估确定同步字段,缺乏一定依据,如果字段选择不合适,还可能造成更多工作开销。

[0063] 基于此,本申请提供了一种数据同步方法,对行式存储数据库的数据库操作语句进行采集,并对采集到的数据库操作语句进行处理,得到操作字段列表,为数据由行式存储数据库同步至列式存储数据库提供一种可参考依据,提高了软件设计的合理性,并缩短了软件开发时间,并且该过程嵌入到应用中,采用异步方式进行数据库操作语句批量提交,该过程和业务系统分离,性能高,不影响业务系统的使用,并且字段查询条件由用户输入,灵活性高,根据输入的字段查询条件从操作字段列表中确定同步至列式存储数据库的目标操作字段,与实际场景相符。

[0064] 下面结合几个具体实施例对本申请提供的的数据同步方法进行详细说明。

[0065] 图1为本申请实施例提供的的数据同步方法的流程示意图,本实施例的执行主体可以为电子设备,如终端设备、后台服务器等。

[0066] 如图1所示,该方法包括:

[0067] S101、根据输入的字段查询条件,从预先存储的操作字段列表中确定满足字段查询条件的操作字段为目标操作字段。

[0068] 其中,操作字段列表中包括:多个操作字段,每个操作字段为预先从针对行式存储

数据库的多个数据库操作语句中提取的字段。

[0069] 数据库操作语句用于对行式存储数据库中的数据表进行操作,例如查询操作、更新操作、删除操作等,数据库操作语句中可以包括行式存储数据库中数据表的部分字段,数据库操作语句可以为结构化查询语言 (Structured Query Language,SQL) 语句。

[0070] 例如,数据库操作语句为SELECT name from table1 where menu,其中,name为操作字段,table1表示数据表1,name表示姓名,where表示条件语句,menu可以为“男”,即查询数据表1中性别为男的姓名。

[0071] 对于实际应用场景而言,在软件应用的使用过程中,可能需要对行式存储数据库中的数据表进行操作,例如,基于数据库操作语句对行式存储数据库中的数据表进行查询,则可以对数据库操作语句进行拦截,并从数据库操作语句中提取操作字段存储至字段操作列表中,其中,该分析过程与业务系统分离,不影响软件应用的使用,即业务系统正常执行数据库操作语句,分析过程是为了提取数据库操作语句中的操作字段。

[0072] 用户可以输入字段查询条件,从预先存储的操作字段列表中确定满足字段查询条件的操作字段为目标操作字段,也即,目标操作字段是根据用户输入的字段查询条件,从操作字段列表中确定的,并且操作字段列表是基于实际场景预先从针对行式存储数据库的多个数据库操作语句中提取到的,这样,为行式存储数据库中的数据同步至列式存储数据库提供了一种可参考依据。

[0073] 其中,行式存储数据库的数量包括但不限于一个,各行式存储数据库中包括至少一个数据表。

[0074] S102、将行式存储数据库中目标操作字段的数据内容同步至列式存储数据库中。

[0075] 确定出目标操作字段之后,即可将行式存储数据库中目标操作字段的数据内容同步至列式存储数据库中,这样,可对列式存储数据库中目标操作字段的数据内容进行统计分析,和对行式存储数据库中目标操作字段的数据内容进行统计分析相比,处理速度快。

[0076] 在本实施例的数据同步方法中,根据输入的字段查询条件,从预先存储的操作字段列表中确定满足字段查询条件的操作字段为目标操作字段,其中,操作字段列表中包括:多个操作字段,每个操作字段为预先从针对行式存储数据库的多个数据库操作语句中提取的字段,将行式存储数据库中目标操作字段的数据内容同步至列式存储数据库中。根据输入的字段查询条件从操作字段列表中确定同步至列式存储数据库的目标操作字段,灵活性强,与实际场景相符,并且操作字段列表是基于实际场景从数据库操作语句中提取的字段,为行式存储数据库中的数据同步至列式存储数据库提供了一种可参考依据。

[0077] 图2为本申请实施例提供的数据库同步方法的流程图二,如图2所示,根据输入的字段查询条件,从预先存储的操作字段列表中确定满足字段查询条件的操作字段为目标操作字段之前,该方法还包括:

[0078] S201、获取预设时间段内针对行式存储数据库的多个数据库操作语句。

[0079] 预设时间段可以为当前时间点之前的预设历史时间段,如当前时间点之前7天,本实施例对预设时间段不做特别限定。

[0080] 针对行式存储数据库的数据库操作语句可以为通过多个应用采集到的针对行式存储数据库中数据表的操作语句,即通过该多个应用均可以对行式存储数据库中数据表进行操作,其中,对数据表的操作例如可以包括查询操作、更新操作、删除操作等。

[0081] 后台服务器可以由客户端(agent端)和服务端(server端)组成,其中,客户端作为数据库操作语句的采集客户端,用于采集应用中的数据库操作语句,服务端作为数据库操作语句解析处理端,用于对应用中的数据库操作语句进行解析处理。

[0082] 所有应用均可以绑定有一个客户端,在软件应用的使用过程中,可能需要对行式存储数据库中的数据表进行操作,即应用要执行数据库操作语句时,可以对数据库操作语句进行拦截,并将该数据库操作语句提交到服务端,为了后续统计需要,还可以提交应用的标识、数据库操作语句的操作时间、以及行式存储数据库的数据库类型,其中,该操作时间可以为数据库操作语句的解析时间。

[0083] S202、根据行式存储数据库的数据库类型,分别对多个数据库操作语句进行解析处理,得到操作字段列表。

[0084] 获取到预设时间段内针对行式存储数据库的多个数据库操作语句之后,还可以根据行式存储数据库的数据库类型,分别对多个数据库操作语句进行解析处理,得到多个数据库操作语句中的操作字段,并根据多个数据库操作语句中的操作字段得到操作字段列表。

[0085] 其中,不同的数据库类型对应有不同的解析规则,因此服务端接收到客户端发送的多个数据库操作语句之后,根据行式存储数据库的数据库类型对应的解析规则,对多个数据库操作语句进行解析处理得到操作字段列表。

[0086] 图3为本申请实施例提供的数据库同步方法的流程示意图三,如图3所示,根据行式存储数据库的数据库类型,分别对多个数据库操作语句进行解析处理,得到操作字段列表,包括:

[0087] S301、获取数据库类型对应的词法文件和语法规则文件。

[0088] S302、根据词法文件和语法规则文件,生成数据库操作语句的解析处理代码。

[0089] 其中,词法文件用于该数据库类型对应的数据库操作语句中词的使用法则,语法规则文件用于指示该数据库类型对应的数据库操作语句中词与词的语法规则。

[0090] 根据行式存储数据库的数据库类型可以确定该数据库类型对应的词法文件和语法规则文件,然后根据词法文件和语法规则文件,采用语法解析器,如antlr语法解析器,生成解析处理代码。

[0091] 需要说明的是,在解析处理之前,可以先对客户端上传的数据库类型对应的词法文件和语法规则文件进行编写,编写时根据不同数据库的语法规则,可从预设数据库中关于数据库操作语句的语法描述文件中进行参考编写,在完成语法规则文件编写之后,根据词法文件和语法规则文件采用语法解析器,生成解析处理代码,并将解析处理代码整合到服务端,在解析数据库操作语句时,直接调用解析处理代码进行解析处理。

[0092] S303、根据解析处理代码,分别对多个数据库操作语句进行解析处理,得到操作字段列表。

[0093] 根据该解析处理代码,对多个数据库操作语句进行解析处理,得到每个数据库操作语句中的操作字段,然后根据多个数据库操作语句中的操作字段得到操作字段列表。

[0094] 需要说明的是,解析处理过程中,可以根据解析处理代码对多个数据库操作语句解析,以将每个数据库操作语句构建为一个抽象语法树,然后对抽象语法树进行进一步解析得到每个数据库操作语句中的操作字段,其中,从抽象语法树中可以很容易地根据各分

支节点的对象信息获取每个数据库操作语句中的字段信息。

[0095] 其中,解析处理过程中,还可以提取出每个数据库操作语句中的数据表的标识和条件字段,其中,条件字段例如可以为where之后的字段,并将操作字段、数据表的标识、条件字段进行保存。例如,数据表为table1,操作字段为field1、field2,存储为tab1.field1、tab1.field2,则操作字段列表中包括两个操作字段以及对应的数据表的标识。

[0096] 在一些实施例中,操作字段列表中可以包括:多个操作字段、各操作字段对应的数据表的标识、各操作字段对应的数据库操作语句的操作时间、各操作字段对应的应用的标识、各操作字段对应的业务标识,可以用sqlOptLog进行表示,以SQL语句SELECT name from table1 where menu为例,操作字段为SELECT的字段name。

[0097] 在一可选的实施方式中,字段查询条件包括:查询时间段,根据输入的字段查询条件,从预先存储的操作字段列表中确定满足字段查询条件的操作字段为目标操作字段,包括:

[0098] 从操作字段列表中确定字段操作时间满足查询时间段的操作字段为目标操作字段。

[0099] 其中,字段操作时间为操作字段列表中各操作字段对应的数据库操作语句的操作时间,数据库操作语句的操作时间可以为数据库操作语句的解析时间,操作字段列表中可以包括:多个操作字段以及各操作字段对应的操作时间,则可以输入查询时间段,将操作字段列表中操作时间在该查询时间段内的操作字段确定为目标操作字段。

[0100] 在一可选的实施方式中,字段查询条件包括:行式存储数据库中目标数据表的标识;根据输入的字段查询条件,从预先存储的操作字段列表中确定满足字段查询条件的操作字段为目标操作字段,包括:

[0101] 从操作字段列表中确定目标数据表对应的操作字段为目标操作字段。

[0102] 其中,操作字段列表中可以包括:多个操作字段以及各操作字段对应的数据表的标识,则可以输入行式存储数据库中目标数据表的标识,将操作字段列表中目标数据表的标识对应的操作字段确定为目标操作字段。

[0103] 在一可选的实施方式中,字段查询条件包括:行式存储数据库对应的目标应用的标识;根据输入的字段查询条件,从预先存储的操作字段列表中确定满足字段查询条件的操作字段为目标操作字段,包括:

[0104] 从操作字段列表中确定目标应用对应的操作字段为目标操作字段。

[0105] 操作字段列表中可以包括:多个操作字段以及各操作字段对应的应用的标识,则可以输入目标应用的标识,将操作字段列表中满足该目标数据表标识的操作字段确定为目标操作字段。

[0106] 在一可选的实施方式中,字段查询条件包括:字段置信度,根据输入的字段查询条件,从预先存储的操作字段列表中确定满足字段查询条件的操作字段为目标操作字段,包括:

[0107] 从操作字段列表中确定满足字段置信度的操作字段为目标操作字段。

[0108] 其中,字段置信度用于指示字段在操作字段列表中出现的频繁程度,计算操作字段列表中每个操作字段的置信度,然后将计算的置信度大于或等于输入的字段置信度的操作字段确定为目标操作字段。

[0109] 图4为本申请实施例提供的高频字段筛选过程的示意图,如图4所示,对5个数据库操作语句分别进行解析得到操作字段列表,该操作字段列表中包括5条操作记录,分别为{table1.filed1,table2.filed2,table1.filed3}、{table2.filed2,table1.filed3}、{table1.filed1,table1.filed3}、{table1.filed1,table2.filed2,table1.filed3,table2.filed4}、{table2.filed4,table2.filed5},其中,table1.filed1表示对数据表1(table1)中的filed1字段进行操作,其它解释类似。

[0110] 计算操作列表中每个操作字段的置信度,{table1.filed1}在操作字段列表中出现的总数为3,操作字段列表中的记录总数为5,则{table1.filed1}的置信度(即table1中filed1的置信度)为 $3/5=60\%$,类似地,{table2.filed2}的置信度为60%,{table1.filed3}的置信度为80%,{table2.filed4}的置信度为40%,{table2.filed5}的置信度为20%。

[0111] 设置输入的字段置信度为40%,将置信度未达到40%的字段过滤,得到{table1.filed1}、{table2.filed2}、{table1.filed3}、{table2.filed4},即目标字段为filed1、filed2、filed3、filed4。

[0112] 在一些实施例中,还可以根据Apriori关联规则算法,对满足字段置信度的字段进行关联组合,即关联组合得到{table1.filed1,table2.filed2}、{table1.filed1,table1.filed3}、{table1.filed1,table2.filed4}、{table2.filed2,table1.filed3}、{table2.filed2,table2.filed4}、{table1.filed3,table2.filed4},再进行下一次迭代,计算字段的置信度,其中,{table1.filed1,table2.filed2}在操作字段列表中出现的总数为2,操作字段列表中的记录总数为5,则{table1.filed1,table2.filed2}的置信度(即table1中filed1和table2中filed2的置信度)为40%,{table1.filed1,table1.filed3}的置信度为60%,{table1.filed1,table2.filed4}的置信度为20%,{table2.filed2,table1.filed3}的置信度为60%,{table2.filed2,table2.filed4}的置信度为20%,{table1.filed3,table2.filed4}的置信度为20%。

[0113] 然后将不满足输入的字段置信度的字段删除,得到{table1.filed1,table2.filed2}、{table1.filed1,table1.filed3}、{table2.filed2,table1.filed3}。

[0114] 接着再进行组合直至得到一条操作记录,最终得到{table1.filed1,table2.filed2,table1.filed3},{table1.filed1,table2.filed2,table1.filed3},对应的置信度为40%,其中,该操作记录中的操作字段filed1、filed2、filed3即为目标操作字段,目标操作字段即为操作字段列表中的高频字段。

[0115] 图5为本申请实施例提供的数据库同步装置的结构示意图,该装置可以集成在电子设备中。如图5所示,该装置包括:

[0116] 确定模块401,用于根据输入的字段查询条件,从预先存储的操作字段列表中确定满足字段查询条件的操作字段为目标操作字段,其中,所述操作字段列表中包括:多个操作字段,每个操作字段为预先从针对行式存储数据库的数据库操作语句中提取的字段;

[0117] 同步模块402,用于将所述行式存储数据库中所述目标操作字段的数据内容同步至列式存储数据库中。

[0118] 在一可选的实施方式中,所述装置还包括:

[0119] 获取模块403,用于获取预设时间段内针对所述行式存储数据库的多个数据库操

作语句；

[0120] 处理模块404,用于根据所述行式存储数据库的数据库类型,分别对所述多个数据库操作语句进行解析处理,得到所述操作字段列表。

[0121] 在一可选的实施方式中,所述处理模块404具体用于:

[0122] 获取所述数据库类型对应的词法文件和语法规则文件;

[0123] 根据所述词法文件和所述语法规则文件,生成解析处理代码;

[0124] 根据所述解析处理代码,分别对所述多个数据库操作语句进行解析处理,得到所述操作字段列表。

[0125] 在一可选的实施方式中,所述字段查询条件包括:查询时间段;

[0126] 所述确定模块401,具体用于:

[0127] 从所述操作字段列表中确定字段操作时间满足所述查询时间段的操作字段为所述目标操作字段,所述字段操作时间为所述操作字段列表中各操作字段对应的数据库操作语句的操作时间。

[0128] 在一可选的实施方式中,所述字段查询条件包括:字段置信度;

[0129] 所述确定模块401,具体用于:

[0130] 从所述操作字段列表中确定满足所述字段置信度的操作字段为所述目标操作字段。

[0131] 在一可选的实施方式中,所述字段查询条件包括:所述行式存储数据库中目标数据表的标识;

[0132] 所述确定模块401,具体用于:

[0133] 从所述操作字段列表中确定所述目标数据表对应的操作字段为所述目标操作字段。

[0134] 在一可选的实施方式中,所述字段查询条件包括:所述行式存储数据库对应的目标应用的标识;

[0135] 所述确定模块401,具体用于:

[0136] 从所述操作字段列表中确定所述目标应用对应的操作字段为所述目标操作字段。

[0137] 在本实施例的数据同步装置中,实现过程和实现原理与上述数据同步方法,在此不再赘述。

[0138] 图6为本申请实施例所提供的电子设备的结构示意图,如图6所示,该设备包括:处理器501、存储器502和总线503,存储器502存储有处理器501可执行的机器可读指令,当电子设备运行时,处理器501与存储器502之间通过总线503通信,处理器501执行机器可读指令,以执行上述方法实施例。

[0139] 本申请实施例还提供了一种计算机可读存储介质,该计算机可读存储介质上存储有计算机程序,该计算机程序被处理器运行时执行,所述处理器执行上述数据同步方法。

[0140] 在本申请实施例中,该计算机程序被处理器运行时还可以执行其它机器可读指令,以执行如实施例其它所述的方法,关于具体执行的方法步骤和原理参见实施例的说明,在此不再详细赘述。

[0141] 在本申请所提供的实施例中,应该理解到,所揭露装置和方法,可以通过其它的方式实现。以上所描述的装置实施例仅仅是示意性的,例如,所述单元的划分,仅仅为一种逻辑

辑功能划分,实际实现时可以有另外的划分方式,又例如,多个单元或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另一点,所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些通信接口,装置或单元的间接耦合或通信连接,可以是电性,机械或其它的形式。

[0142] 所述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,作为单元显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单元来实现本实施例方案的目的。

[0143] 另外,在本申请提供的实施例中的各功能单元可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中。

[0144] 所述功能如果以软件功能单元的形式实现并作为独立的产品销售或使用,可以存储在一个计算机可读取存储介质中。基于这样的理解,本申请的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质中,包括若干指令用以使得一台计算机设备(可以是个人计算机,服务器,或者网络设备等)执行本申请各个实施例所述方法的全部或部分步骤。而前述的存储介质包括:U盘、移动硬盘、只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、磁碟或者光盘等各种可以存储程序代码的介质。

[0145] 应注意到:相似的标号和字母在下面的附图中表示类似项,因此,一旦某一项在一个附图中被定义,则在随后的附图中不需要对其进行进一步定义和解释,此外,术语“第一”、“第二”、“第三”等仅用于区分描述,而不能理解为指示或暗示相对重要性。

[0146] 最后应说明的是:以上所述实施例,仅为本申请的具体实施方式,用以说明本申请的技术方案,而非对其限制,本申请的保护范围并不局限于此,尽管参照前述实施例对本申请进行了详细的说明,本领域的普通技术人员应当理解: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,其依然可以对前述实施例所记载的技术方案进行修改或可轻易想到变化,或者对其中部分技术特征进行等同替换;而这些修改、变化或者替换,并不使相应技术方案的本质脱离本申请实施例技术方案的精神和范围。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以所述权利要求的保护范围为准。

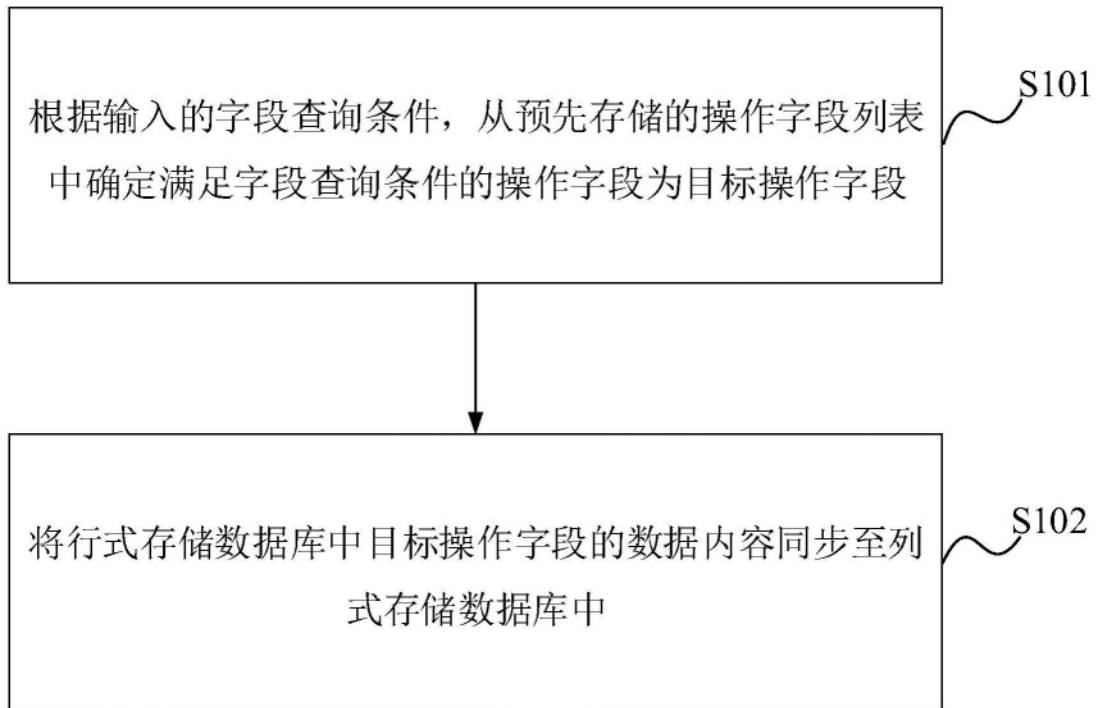


图1

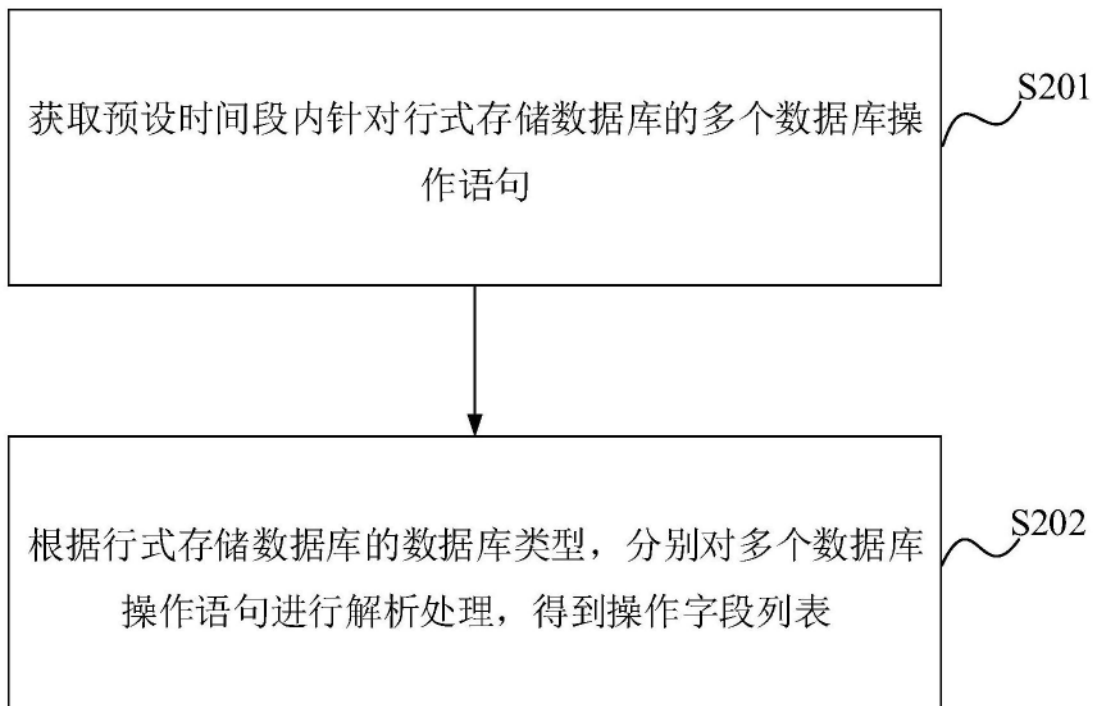


图2

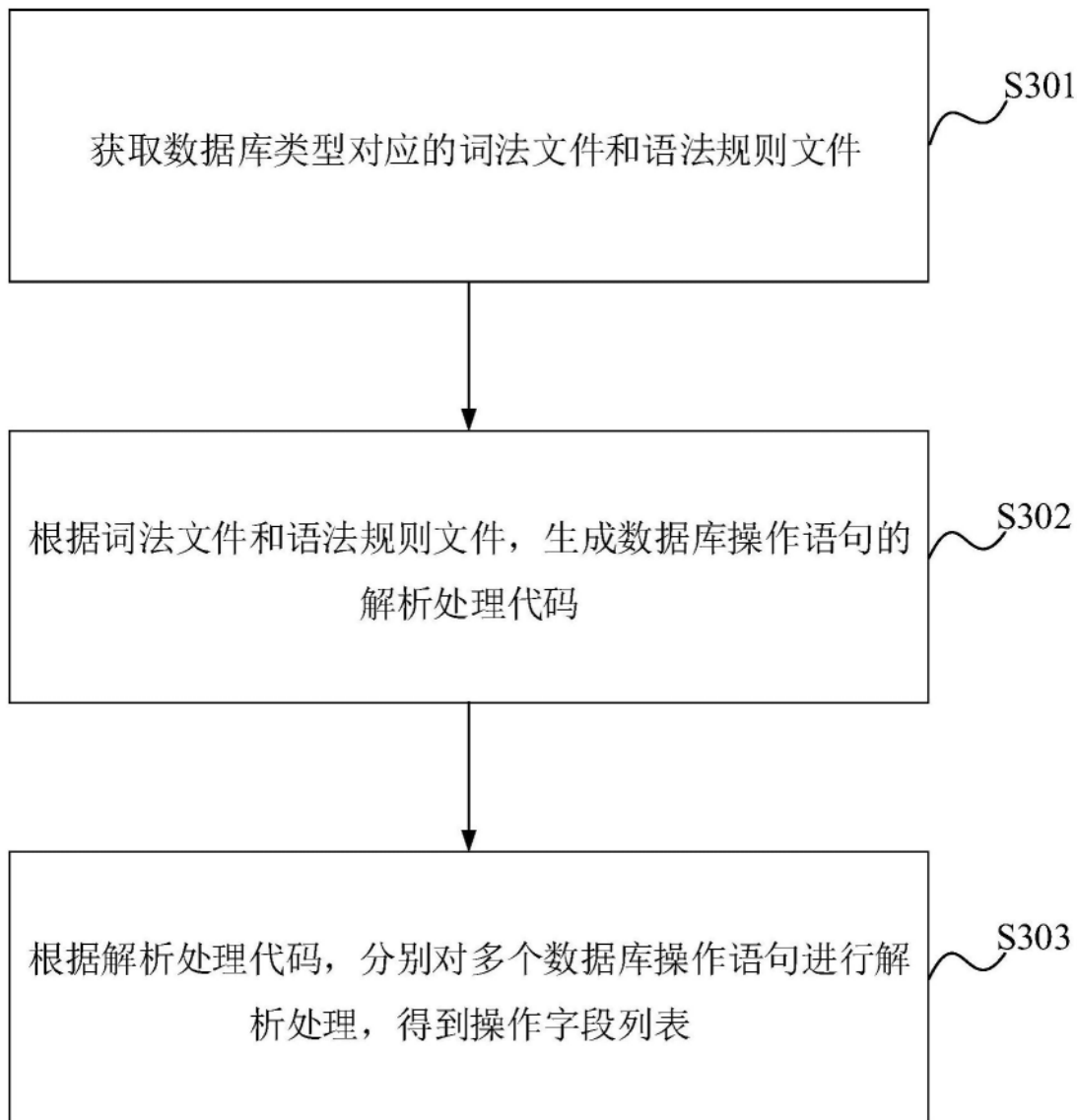


图3

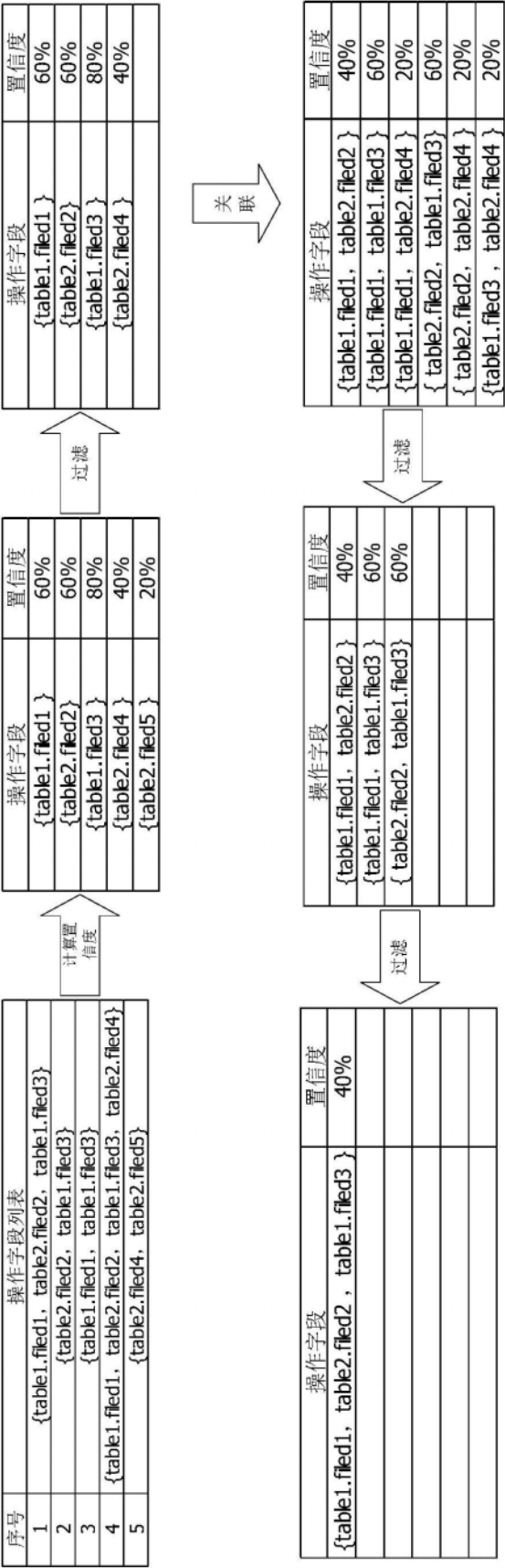


图4

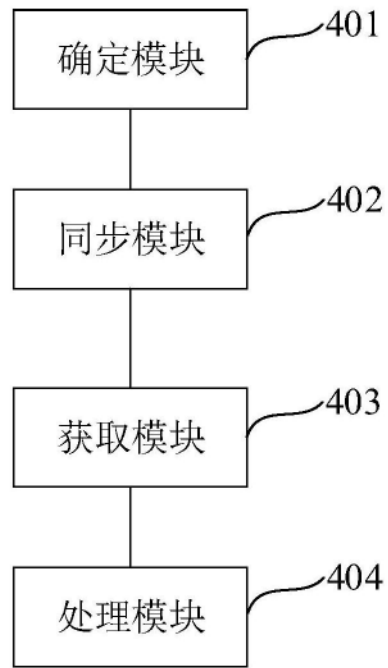


图5

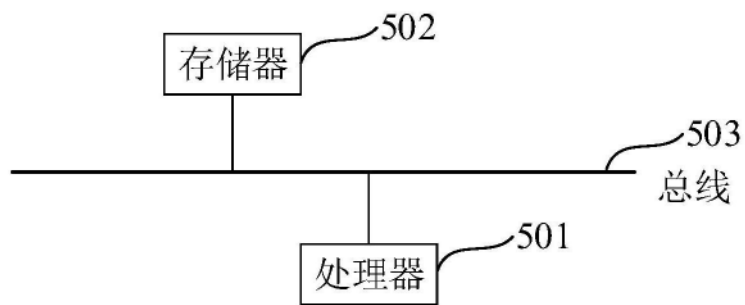


图6