



(12) 发明专利

(10) 授权公告号 CN 112445854 B

(45) 授权公告日 2024. 05. 03

(21) 申请号 202011345654.X

(22) 申请日 2020.11.25

(65) 同一申请的已公布的文献号

申请公布号 CN 112445854 A

(43) 申请公布日 2021.03.05

(73) 专利权人 北京品域互联科技有限公司

地址 100020 北京市朝阳区东三环中路20

号楼20层2009单元

(72) 发明人 杜駟骏

(74) 专利代理机构 深圳市新峰知识产权代理事

务所(普通合伙) 441078

专利代理师 李超

(51) Int. Cl.

G06F 16/25 (2019.01)

G06F 16/248 (2019.01)

(56) 对比文件

CN 109815013 A, 2019.05.28

CN 110704422 A, 2020.01.17

CN 110784419 A, 2020.02.11

CN 111460023 A, 2020.07.28

US 2020293549 A1, 2020.09.17

审查员 邢玉良

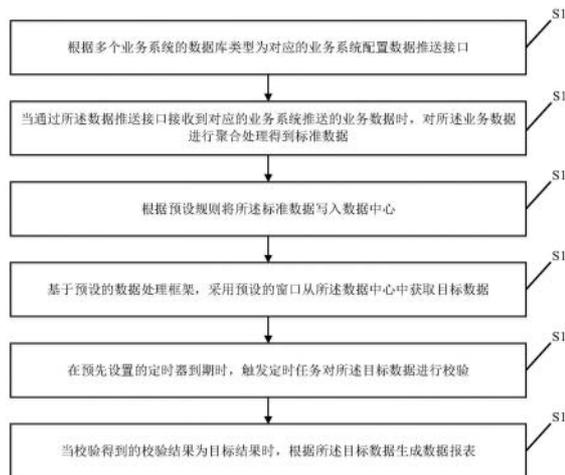
权利要求书2页 说明书14页 附图3页

(54) 发明名称

多源业务数据实时处理方法、装置、终端及存储介质

(57) 摘要

本发明涉及大数据技术领域,提供一种多源业务数据实时处理方法、装置、终端及存储介质,包括:根据多个业务系统的数据库类型为对应的业务系统配置数据推送接口;当通过所述数据推送接口接收到对应的业务系统推送的业务数据时,对所述业务数据进行聚合处理得到标准数据;根据预设规则将所述标准数据写入数据中心;基于预设的数据处理框架,采用预设的窗口从所述数据中心中获取目标数据;在预先设置的定时器到期时,触发定时任务对所述目标数据进行校验;当校验得到的校验结果为目标结果时,根据所述目标数据生成数据报表。本发明能够对多个数据源的业务数据进行实时处理,提高了业务数据的处理效率。



1. 一种多源业务数据实时处理方法,其特征在于,所述方法包括:
 - 根据多个业务系统的数据库类型为对应的业务系统配置数据推送接口;
 - 当通过所述数据推送接口接收到对应的业务系统推送的业务数据时,对所述业务数据进行聚合处理得到标准数据;
 - 根据预设规则将所述标准数据写入数据中心;
 - 基于预设的数据处理框架,采用预设的窗口从所述数据中心中获取目标数据;
 - 在预先设置的定时器到期时,触发定时任务对所述目标数据进行校验;
 - 当校验得到的校验结果为目标结果时,根据所述目标数据生成数据报表;
 - 所述根据预设规则将所述标准数据写入数据中心包括:判断缓存中是否存在与所述标准数据对应的关键字段;当所述缓存中不存在与所述标准数据对应的关键字段时,在所述缓存中创建所述关键字段;当所述缓存中存在与所述标准数据对应的关键字段时,以所述关键字段为主键,将所述关键字段对应的标准数据写入缓存;当所述关键字段对应的所有标准数据均写入缓存时,对写入缓存中的标准数据进行去重处理,并将去重处理后的标准数据写入数据中心;
 - 所述触发定时任务对所述目标数据进行校验包括:确定所述目标数据对应的校验标识符;采用所述预设的窗口从所述数据中心中获取与所述校验标识符对应的校验数据;比对所述目标数据与所述校验数据是否一致;当所述目标数据与所述校验数据一致时,得到第一校验结果;当所述目标数据与所述校验数据不一致时,得到第二校验结果;
 - 所述根据所述目标数据生成数据报表包括:输入所述目标数据至预先定义的多个函数中进行计算,得到多个函数值;根据预设业务模板,基于所述多个函数值生成数据报表。
2. 如权利要求1所述的多源业务数据实时处理方法,其特征在于,所述对所述业务数据进行聚合处理得到标准数据包括:
 - 确定查看所述数据报表的对象的喜好;
 - 根据所述对象的喜好从多个预设的聚合处理方式中选择出目标聚合处理方式;
 - 根据所述目标聚合处理方式,使用预设函数对所述业务数据进行聚合处理得到标准数据。
3. 如权利要求1或2所述的多源业务数据实时处理方法,其特征在于,所述方法还包括:
 - 获取所述目标数据的推送时间;
 - 根据所述推送时间将所述目标数据分成多个分区;
 - 计算每个分区中的目标数据的数据量;
 - 确定大于预设数据量阈值的目标分区;
 - 为所述目标分区中的目标数据建立二叉树索引。
4. 如权利要求3所述的多源业务数据实时处理方法,其特征在于,所述方法还包括:
 - 接收业务数据的查询请求并解析出所述查询请求得到查询字段及查询时间;
 - 根据所述查询时间确定查询分区;
 - 判断所述查询分区是否为目标分区;
 - 当所述查询分区为目标分区时,基于所述二叉树索引查询与所述查询字段对应的业务数据。
5. 一种多源业务数据实时处理装置,其特征在于,所述装置用于实现如权利要求1所述

的方法,所述装置包括:

配置模块,用于根据多个业务系统的数据库类型为对应的业务系统配置数据推送接口;

处理模块,用于当通过所述数据推送接口接收到对应的业务系统推送的业务数据时,对所述业务数据进行聚合处理得到标准数据;

写入模块,用于根据预设规则将所述标准数据写入数据中心;

获取模块,用于基于预设的数据处理框架,采用预设的窗口从所述数据中心的获取目标数据;

校验模块,用于在预先设置的定时器到期时,触发定时任务对所述目标数据进行校验;

生成模块,用于当校验得到的校验结果为目标结果时,根据所述目标数据生成数据报表。

6.一种终端,其特征在于,所述终端包括处理器,所述处理器用于执行存储器中存储的计算机程序时实现如权利要求1至4中任意一项所述多源业务数据实时处理方法。

7.一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,其特征在于,所述计算机程序被处理器执行时实现如权利要求1至4中任意一项所述多源业务数据实时处理方法。

多源业务数据实时处理方法、装置、终端及存储介质

技术领域

[0001] 本发明涉及数据处理技术领域,具体涉及一种多源业务数据实时处理方法、装置、终端及存储介质。

背景技术

[0002] 随着公司业务的发展,业务数据越来越多,业务数据的来源也越来越繁杂,例如,来自于关系型数据库,非关系型数据库或者诸如excel等的文件,为了商业发展,需要实时对这些不同业务系统的业务数据进行汇总分析。

[0003] 然而,发明人在实现本发明的过程中发现,现有的数据汇总方法,是以离线的方式进行业务数据的同步,然后对同步后的业务数据进行汇总分析,汇总分析的时效性通常是T-1日,导致业务数据的分析结果具有滞后性,准确率无法得到保障;此外,仅能对单个业务系统的数据进行实时汇总,由于业务数据落库时存在一定的延迟,因此对多个业务系统的业务数据交叉实时汇总非常困难。

发明内容

[0004] 鉴于以上内容,有必要提出一种多源业务数据实时处理方法、装置、终端及存储介质,能够提高业务数据的处理效率。

[0005] 本发明的第一方面提供一种多源业务数据实时处理方法,所述方法包括:

[0006] 根据多个业务系统的数据库类型为对应的业务系统配置数据推送接口;

[0007] 当通过所述数据推送接口接收到对应的业务系统推送的业务数据时,对所述业务数据进行聚合处理得到标准数据;

[0008] 根据预设规则将所述标准数据写入数据中心;

[0009] 基于预设的数据处理框架,采用预设的窗口从所述数据中心的获取目标数据;

[0010] 在预先设置的定时器到期时,触发定时任务对所述目标数据进行校验;

[0011] 当校验得到的校验结果为目标结果时,根据所述目标数据生成数据报表。

[0012] 根据本发明的一个可选的实施例,所述对所述业务数据进行聚合处理得到标准数据包括:

[0013] 确定查看所述数据报表的对象的喜好;

[0014] 根据所述对象的喜好从多个预设的聚合处理方式中选择出目标聚合处理方式;

[0015] 根据所述目标聚合处理方式,使用预设函数对所述业务数据进行聚合处理得到标准数据。

[0016] 根据本发明的一个可选的实施例,所述根据预设规则将所述标准数据写入数据中心包括:

[0017] 判断缓存中是否存在与所述标准数据对应的关键字段;

[0018] 当所述缓存中不存在与所述标准数据对应的关键字段时,在所述缓存中创建所述关键字段;

- [0019] 当所述缓存中存在与所述标准数据对应的关键字段时,以所述关键字段为主键,将所述关键字段对应的标准数据写入缓存;
- [0020] 当所述关键字段对应的所有标准数据均写入缓存时,对写入缓存中的标准数据进行去重处理,并将去重处理后的标准数据写入数据中心。
- [0021] 根据本发明的一个可选的实施例,所述触发定时任务对所述目标数据进行校验包括:
- [0022] 确定所述目标数据对应的校验标识符;
- [0023] 采用所述预设的窗口从所述数据中心中获取与所述校验标识符对应的校验数据;
- [0024] 比对所述目标数据与所述校验数据是否一致;
- [0025] 当所述目标数据与所述校验数据一致时,得到第一校验结果;
- [0026] 当所述目标数据与所述校验数据不一致时,得到第二校验结果。
- [0027] 根据本发明的一个可选的实施例,所述根据所述目标数据生成数据报表包括:
- [0028] 输入所述目标数据至预先定义的多个函数中进行计算,得到多个函数值;
- [0029] 根据预设业务模板,基于所述多个函数值生成数据报表。
- [0030] 根据本发明的一个可选的实施例,所述方法还包括:
- [0031] 获取所述目标数据的推送时间;
- [0032] 根据所述推送时间将所述目标数据分成多个分区;
- [0033] 计算每个分区中的目标数据的数据量;
- [0034] 确定大于预设数据量阈值的目标分区;
- [0035] 为所述目标分区中的目标数据建立二叉树索引。
- [0036] 根据本发明的一个可选的实施例,所述方法还包括:
- [0037] 接收业务数据的查询请求并解析出所述查询请求得到查询字段及查询时间;
- [0038] 根据所述查询时间确定查询分区;
- [0039] 判断所述查询分区是否为目标分区;
- [0040] 当所述查询分区为目标分区时,基于所述二叉树索引查询与所述查询字段对应的业务数据。
- [0041] 本发明的第二方面提供一种多源业务数据实时处理装置,所述装置包括:
- [0042] 配置模块,用于根据多个业务系统的数据库类型为对应的业务系统配置数据推送接口;
- [0043] 处理模块,用于当通过所述数据推送接口接收到对应的业务系统推送的业务数据时,对所述业务数据进行聚合处理得到标准数据;
- [0044] 写入模块,用于根据预设规则将所述标准数据写入数据中心;
- [0045] 获取模块,用于基于预设的数据处理框架,采用预设的窗口从所述数据中心中获取目标数据;
- [0046] 校验模块,用于在预先设置的定时器到期时,触发定时任务对所述目标数据进行校验;
- [0047] 生成模块,用于当校验得到的校验结果为目标结果时,根据所述目标数据生成数据报表。
- [0048] 本发明的第三方面提供一种终端,所述终端包括处理器,所述处理器用于执行存

存储器中存储的计算机程序时实现所述多源业务数据实时处理方法。

[0049] 本发明的第四方面提供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执行时实现所述多源业务数据实时处理方法。

[0050] 综上所述,本发明所述的多源业务数据实时处理方法、装置、终端及存储介质,根据业务系统的数据库类型配置数据推送接口并根据数据推送接口接收业务系统推送的业务数据,使得接收业务系统的业务数据的效率能够得以提高,在接收到业务数据时,通过对所述业务数据进行聚合处理得到标准数据,并根据预设规则将所述标准数据写入数据中心,再基于预设的数据处理框架,采用预设的窗口从所述数据中心中获取目标数据,最后在预先设置的定时器到期时,触发定时任务对所述目标数据进行校验,不仅能够实时处理多个业务系统的业务数据,还能确保写入数据中心的业务数据是正确的。当校验得到的校验结果为目标结果时,根据所述目标数据生成数据报表,便于直观的展示业务数据。

附图说明

[0051] 图1是本发明实施例一提供的多源业务数据实时处理方法的流程图。

[0052] 图2是本发明实施例二提供的多源业务数据实时处理装置的结构图。

[0053] 图3是本发明实施例三提供的终端的结构示意图。

具体实施方式

[0054] 为了能够更清楚地理解本发明的上述目的、特征和优点,下面结合附图和具体实施例对本发明进行详细描述。需要说明的是,在不冲突的情况下,本发明的实施例及实施例中的特征可以相互组合。

[0055] 除非另有定义,本文所使用的所有的技术和科学术语与属于本发明的技术领域的技术人员通常理解的含义相同。本文中在本发明的说明书中所使用的术语只是为了描述具体的实施例的目的,不是旨在于限制本发明。

[0056] 本发明实施例提供的多源业务数据实时处理方法由终端执行,相应地,多源业务数据实时处理装置运行于终端中。

[0057] 图1是本发明实施例一提供的多源业务数据实时处理方法的流程图。所述多源业务数据实时处理方法具体包括以下步骤,根据不同的需求,该流程图中步骤的顺序可以改变,某些可以省略。

[0058] S11,根据多个业务系统的数据库类型为对应的业务系统配置数据推送接口。

[0059] 终端对应多个业务系统,业务系统是指承接具体业务的系统,业务系统事先需要在终端中完成注册,以注册自身所使用的数据库类型。其中,所述数据库类型可以包括诸如mysql,oracle,postgres等关系型数据库,hbase,hive,kudu,elasticsearch等非关系型数据库,以及excel等文本型数据库。所述关系型数据库是指以关系数学模型来表示的数据表,所述非关系型数据库是指没有固定类型且数据之间无规律的数据表。

[0060] 终端在侦测到业务系统注册成功时,根据业务系统的数据库类型配置终端与业务系统之间进行数据传输的数据推送接口。示例性的,假设多个业务系统包括:电销系统、售前系统及售后系统,所述电销系统对应的数据库类型为MySQL,则终端为电销系统配置MQ数据推送接口,所述售前系统对应的数据库类型为postgres,则终端为售前系统配置kafka

数据推送接口,所述售后系统对应的数据库类型为oracle,则终端为售后系统配置gg数据推送接口。

[0061] S12,当通过所述数据推送接口接收到对应的业务系统推送的业务数据时,对所述业务数据进行聚合处理得到标准数据。

[0062] 不同的业务系统对应不同类型的数据库,不同类型的数据库对数据的同种操作函数表达不同,导致数据管理无法统一,将数据写入数据中心之后,会造成后期的数据请求较为繁琐,因此,终端在接收到业务数据之后,对所述业务数据进行聚合处理得到标准数据。

[0063] 在一个可选的实施例中,所述对所述业务数据进行聚合处理得到标准数据包括:

[0064] 确定查看所述数据报表的对象的喜好;

[0065] 根据所述对象的喜好从多个预设的聚合处理方式中选择出目标聚合处理方式;

[0066] 根据所述目标聚合处理方式,使用预设函数对所述业务数据进行聚合处理得到标准数据。

[0067] 其中,所述多个预设的聚合处理方式可以包括:日期聚合、数字转换、度量列转换和值映转换。

[0068] 所述查看所述数据报表的对象是指基于多个业务系统的业务数据进行分析并生成的数据报表所要显示的对象,例如,企业领导。

[0069] 不同角色的报表展示的对象具有不同的喜好,某些报表展示的对象喜好日期聚合方式,某些报表展示的对象喜好数字转换方式。

[0070] 该可选的实施例中,基于Spark大数据处理框架,根据报表展示的对象喜好确定目标聚合处理方式,并使用目标聚合处理方式及Spark大数据处理框架所支持的SQL函数对业务数据进行聚合处理得到标准数据,能够实现对不同数据源不同类型的数据库中的业务数据的统一处理;将关系型数据库、非关系型数据库及文本型数据库中的业务数据统一转换成spark支持的SQL函数表示的标准数据,能够减少后期的维护,避免因数据源的不同而请求不同的接口,提高了数据处理的效率。

[0071] S13,根据预设规则将所述标准数据写入数据中心。

[0072] 所述预设规则可由终端的管理人员自定义。

[0073] 业务系统推送业务数据到终端时,会有时间上的误差,可能出现售后数据已推送到终端中,售前数据还没有推送到终端中,则需要根据预设规则将所述标准数据写入数据中心,以解决业务系统对应的标准数据写入数据中心不同步的问题。

[0074] 在一个可选的实施例中,所述根据预设规则将所述标准数据写入数据中心包括:

[0075] 判断缓存中是否存在与所述标准数据对应的关键字段;

[0076] 当所述缓存中不存在与所述标准数据对应的关键字段时,在所述缓存中创建所述关键字段;

[0077] 当所述缓存中存在与所述标准数据对应的关键字段时,以所述关键字段为主键,将所述关键字段对应的标准数据写入缓存;

[0078] 当所述关键字段对应的所有标准数据均写入缓存时,对写入缓存中的标准数据进行去重处理,并将去重处理后的标准数据写入数据中心。

[0079] 其中,所述关键字段由管理人员在接收到的业务数据的多个字段中进行选择,或由终端根据多个业务系统推送的业务数据的共同字段确定的。

[0080] 不同的业务系统对应的标准数据具有至少两个相同的字段,将所述至少两个相同的字段中的任意一个字段确定为关键字段。示例性的,所述关键字段为申请号。

[0081] 终端得到标准数据之后,先判断缓存中是否存在所述关键字段,当缓存中存在关键字段时,表明缓存中已存储有其他业务系统对应的标准数据,则以关键字段为主键,将具有同一个关键字段的多个业务系统对应的标准数据写入所述缓存中的同一行。当缓存中不存在所述关键字段时,表明缓存中未存储任何业务系统对应的标准数据,则创建关键字段并以创建的关键字段为主键。

[0082] 当所有的业务系统对应的标准数据均写入缓存中之后,终端对同一行的标准数据进行去重处理,去重处理有助于减少重复的标准数据,节约缓存空间,且由于去重处理之后,减少了标准数据的数据量,因此将去重处理后的标准数据写入数据中心能够加快数据写入的速度,提高数据写入的效率。

[0083] S14,基于预设的数据处理框架,采用预设的窗口从所述数据中心的获取目标数据。

[0084] 其中,所述预设的数据处理框架为基于Spark的大数据处理框架。

[0085] 其中,所述预设的窗口为5秒。采用预设的窗口从所述数据中心加载标准数据,并将加载的标准数据确定为目标数据。

[0086] S15,在预先设置的定时器到期时,触发定时任务对所述目标数据进行校验。

[0087] 终端可以预先设置一个多时定时器,例如,设置在每个整点00分触发定时任务,以对与所述目标数据进行验证,确保所述目标数据的有效性。

[0088] 在一个可选的实施例中,所述触发定时任务对所述目标数据进行校验包括:

[0089] 确定所述目标数据对应的校验标识符;

[0090] 采用所述预设的窗口从所述数据中心的获取与所述校验标识符对应的校验数据;

[0091] 比对所述目标数据与所述校验数据是否一致;

[0092] 当所述目标数据与所述校验数据一致时,得到第一校验结果;

[0093] 当所述目标数据与所述校验数据不一致时,得到第二校验结果。

[0094] 该可选的实施例中,由于不同的业务系统对应的标准数据具有至少两个相同的字段,该至少两个相同的字段中的一个字段被确定为关键字段,则该至少两个相同的字段中的除关键字段外的其他字段均可以被确定为校验标识符,例如,所述校验标识符为身份证号码。

[0095] 终端确定校验标识符之后,基于所述预设的数据处理框架,采用预设的窗口从所述数据中心加载所述校验标识符对应的标准数据,并将所加载的标准数据确定为校验数据。

[0096] 终端将所述目标数据与校验数据中的每一个具有相同字段的数据进行比对,当所述目标数据与校验数据中的所有具有相同字段的数据比对一致时,确定所述目标数据与所述校验数据相同,即表明所述目标数据有效,得到第一校验结果;当所述目标数据与校验数据中的任意一个具有相同字段的数据比对不一致时,确定所述目标数据与所述校验数据不相同,即表明所述目标数据无效,得到第二校验结果。

[0097] S16,当校验得到的校验结果为目标结果时,根据所述目标数据生成数据报表。

[0098] 其中,所述目标结果为校验一致。

[0099] 终端在确定校验结果为目标结果时,生成数据报表,以实现和数据的一个直观展示。进一步的,所述终端还可以具有如下功能:维度切换,排版切换,可删除/替换/追加图表,实现自由灵活的交互。

[0100] 在一个可选的实施例中,所述根据所述目标数据生成数据报表包括:

[0101] 输入所述目标数据至预先定义的多个函数中进行计算,得到多个函数值;

[0102] 根据预设业务模板,基于所述多个函数值生成数据报表。

[0103] 其中,所述多个函数包括申请总和,申请通过率,放款量,放款件,平均通过时间,平均通过次数,同比增长值等。例如,输入所述目标数据至申请总和中进行计算,可计算得到申请总和。

[0104] 该可选的实施例中,根据预设业务模板生成数据报表,实现了不同数据源的业务数据的交叉实时处理,具有较高的时效性;并以数据报表的方式进行直观的展示,用户体验高。

[0105] 进一步的,所述方法还可以包括:比较每一个函数值与对应的告警阈值;当所述函数值大于所述告警阈值时,触发警报指令。

[0106] 在一个可选的实施例中,所述方法还包括:

[0107] 获取所述目标数据的推送时间;

[0108] 根据所述推送时间将所述目标数据分成多个分区;

[0109] 计算每个分区中的目标数据的数据量;

[0110] 确定大于预设数据量阈值的目标分区;

[0111] 为所述目标分区中的目标数据建立二叉树索引。

[0112] 例如,将推送时间为T1(例如,1月份)的目标数据分在第一分区,将推送时间为T2(例如,2月份)的目标数据分在第二分区等等。每个分区的目标数据的数据量可以相同或者不同。

[0113] 当某个分区中的目标数据的数据量大于预设数据量阈值时,表明该分区中的目标数据较多,若需从该分区中查询目标数据时,需要耗费较多的查询时间。当某个分区中的目标数据的数据量小于或者等于预设数据量阈值时,表明该分区中的目标数据较少,若需从该分区中查询目标数据时,不需要耗费较多的查询时间。

[0114] 上述可选的实施例中,通过将目标数据存储在不同的分区中,便于后续查询时,能够从对应的分区中查询目标数据,而不需对所有的目标数据中的每一条目标数据进行逐一比对,减少数据查询时间;而根据每个分区中的目标数据的数据量确定是否建立二叉树索引,通过二叉树索引查询目标数据,能够进一步减少数据查询时间。

[0115] 在一个可选的实施例中,所述为所述目标分区中的目标数据建立二叉树索引包括:

[0116] 从所述目标数据中随机选取两个目标数据作为初始中心点;

[0117] 利用预设算法,基于所述初始中心点进行迭代聚类,在迭代次数大于1的当前迭代的上一次迭代中能得到至少一组候选类别,所述至少一组候选类别中每组候选类别对应第一候选类别及第二候选类别;

[0118] 所述每组候选类别对应的二叉树节点表示每组候选类别中第一候选类别与第二候选类别间的差分特征值;

[0119] 在当前迭代中,分别对所述每组候选类别中第一候选类别及第二候选类别分别进行迭代聚类,得到所述每组候选类别中第一候选类别对应的两个候选类别及所述每组候选类别中第二候选类别对应的两个候选类别;

[0120] 将表示所述每组候选类别中第一候选类别对应的两个候选类别的差分特征的二叉树节点作为所述每组候选类别对应的二叉树节点的左节点或右节点,将表示所述每组候选类别中第二候选类别对应的两个候选类别的差分特征的二叉树节点作为所述每组候选类别对应的二叉树节点的右节点或左节点;

[0121] 执行多次迭代聚类,直至达到迭代终止条件,得到所述目标分区对应的多个类别,其中二叉树的一个叶子节点表示一个类别的目标数据。

[0122] 在本发明的优选实施例中,在对每个区中的目标数据生成二叉树索引时,在所述二叉树索引中,二叉树的一个叶子节点表示一个类别的目标数据,除叶子节点外的每个节点表示两个候选类别的差分特征,除叶子节点及根节点外任意一个节点对应的两个候选类别由对所述任意一个节点的父节点对应的两个候选类别中第一候选类别分类得到的,所述任意一个节点的兄弟节点对应的两个候选类别由对所述任意一个节点的父节点对应的两个候选类别中第二候选类别分类得到的。这样建立二叉树索引后,后续在对数据进行搜索,无需对每一条的数据进行比对,大大提升了搜索速度。

[0123] 在一个可选的实施例中,所述方法还包括:

[0124] 接收业务数据的查询请求并解析出所述查询请求得到查询字段及查询时间;

[0125] 根据所述查询时间确定查询分区;

[0126] 判断所述查询分区是否为目标分区;

[0127] 当所述查询分区为目标分区时,基于所述二叉树索引查询与所述查询字段对应的业务数据。

[0128] 该可选的实施例中,由于查询分区中的目标数据是以推送时间进行存储的,因此根据查询时间即可确定查询分区。

[0129] 所述目标分区是指建立有二叉树索引的分区,每个分区对应分区标识,根据分区标识即可确定所述查询分区是否为目标分区。

[0130] 当所述查询分区为目标分区时,采用优先队列从所述二叉树索引的根节点开始遍历,将遍历的节点插入在所述优先队列中,基于所述查询字段对应的业务数据与遍历的节点表示的业务数据的距离,对所述优先队列进行从大到小排序,将排序在第一位的业务数据推送给用户。

[0131] 基于所述查询字段对应的业务数据与遍历的节点表示的业务数据的距离进行查询,能够避免遍历二叉树的所有节点,提高了查询速度。在没有建立二叉树索引的分区中,基于现有技术中的深度优先搜索算法搜索与所述查询字段对应的业务数据。

[0132] 需要强调的是,为进一步保证上述二叉树索引的私密性和安全性,上述二叉树索引可存储于区块链的节点中。

[0133] 图2是本发明实施例二提供的多源业务数据实时处理装置的结构图。

[0134] 在一些实施例中,所述多源业务数据实时处理装置20可以包括多个由计算机程序段所组成的功能模块。所述多源业务数据实时处理装置20中的各个程序段的计算机程序可以存储于终端的存储器中,并由至少一个处理器所执行,以执行(详见图1描述)多源业务数

据实时处理的功能。

[0135] 本实施例中,所述多源业务数据实时处理装置20根据其所执行的功能,可以被划分为多个功能模块。所述功能模块可以包括:配置模块201、处理模块202、写入模块203、获取模块204、校验模块205、生成模块206、警报模块207、建立模块208及查询模块209。本发明所称的模块是指一种能够被至少一个处理器所执行并且能够完成固定功能的一系列计算机程序段,其存储在存储器中。在本实施例中,关于各模块的功能将在后续的实施例中详述。

[0136] 所述配置模块201,用于根据多个业务系统的数据库类型为对应的业务系统配置数据推送接口。

[0137] 终端对应多个业务系统,业务系统是指承接具体业务的系统,业务系统事先需要在终端中完成注册,以注册自身所使用的数据库类型。其中,所述数据库类型可以包括诸如mysql,oracle,postgres等关系型数据库,hbase,hive,kudu,elasticsearch等非关系型数据库,以及excel等文本型数据库。所述关系型数据库是指以关系数学模型来表示的数据表,所述非关系型数据库是指没有固定类型且数据之间无规律的数据表。

[0138] 终端在侦测到业务系统注册成功时,根据业务系统的数据库类型配置终端与业务系统之间进行数据传输的数据推送接口。示例性的,假设多个业务系统包括:电销系统、售前系统及售后系统,所述电销系统对应的数据库类型为MySQL,则终端为电销系统配置MQ数据推送接口,所述售前系统对应的数据库类型为postgres,则终端为售前系统配置kafka数据推送接口,所述售后系统对应的数据库类型为oracle,则终端为售后系统配置gg数据推送接口。

[0139] 所述处理模块202,用于当通过所述数据推送接口接收到对应的业务系统推送的业务数据时,对所述业务数据进行聚合处理得到标准数据。

[0140] 不同的业务系统对应不同类型的数据库,不同类型的数据库对数据的同种操作函数表达不同,导致数据管理无法统一,将数据写入数据中心之后,会造成后期的数据请求较为繁琐,因此,终端在接收到业务数据之后,对所述业务数据进行聚合处理得到标准数据。

[0141] 在一个可选的实施例中,所述处理模块202对所述业务数据进行聚合处理得到标准数据包括:

[0142] 确定查看所述数据报表的对象的喜好;

[0143] 根据所述对象的喜好从多个预设的聚合处理方式中选择出目标聚合处理方式;

[0144] 根据所述目标聚合处理方式,使用预设函数对所述业务数据进行聚合处理得到标准数据。

[0145] 其中,所述多个预设的聚合处理方式可以包括:日期聚合、数字转换、度量列转换和值映转换。

[0146] 所述查看所述数据报表的对象是指基于多个业务系统的业务数据进行分析并生成的数据报表所要显示的对象,例如,企业领导。

[0147] 不同角色的报表展示的对象具有不同的喜好,某些报表展示的对象喜好日期聚合方式,某些报表展示的对象喜好数字转换方式。

[0148] 该可选的实施例中,基于Spark大数据处理框架,根据报表展示的对象喜好确定目标聚合处理方式,并使用目标聚合处理方式及Spark大数据处理框架所支持的SQL函数对

业务数据进行聚合处理得到标准数据,能够实现对不同数据源不同类型的数据库中的业务数据的统一处理;将关系型数据库、非关系型数据库及文本型数据库中的业务数据统一转换成spark支持的SQL函数表示的标准数据,能够减少后期的维护,避免因数据源的不同而请求不同的接口,提高了数据处理效率。

[0149] 所述写入模块203,用于根据预设规则将所述标准数据写入数据中心。

[0150] 所述预设规则可由终端的管理人员自定义。

[0151] 业务系统推送业务数据到终端时,会有时间上的误差,可能出现售后数据已推送到终端中,售前数据还没有推送到终端中,则需要根据预设规则将所述标准数据写入数据中心,以解决业务系统对应的标准数据写入数据中心不同步的问题。

[0152] 在一个可选的实施例中,所述写入模块203根据预设规则将所述标准数据写入数据中心包括:

[0153] 判断缓存中是否存在与所述标准数据对应的关键字段;

[0154] 当所述缓存中不存在与所述标准数据对应的关键字段时,在所述缓存中创建所述关键字段;

[0155] 当所述缓存中存在与所述标准数据对应的关键字段时,以所述关键字段为主键,将所述关键字段对应的标准数据写入缓存;

[0156] 当所述关键字段对应的所有标准数据均写入缓存时,对写入缓存中的标准数据进行去重处理,并将去重处理后的标准数据写入数据中心。

[0157] 其中,所述关键字段由管理人员在接收到的业务数据的多个字段中进行选择,或由终端根据多个业务系统推送的业务数据的共同字段确定的。

[0158] 不同的业务系统对应的标准数据具有至少两个相同的字段,将所述至少两个相同的字段中的任意一个字段确定为关键字段。示例性的,所述关键字段为申请号。

[0159] 终端得到标准数据之后,先判断缓存中是否存在所述关键字段,当缓存中存在关键字段时,表明缓存中已存储有其他业务系统对应的标准数据,则以关键字段为主键,将具有同一个关键字段的多个业务系统对应的标准数据写入所述缓存中的同一行。当缓存中不存在所述关键字段时,表明缓存中未存储任何业务系统对应的标准数据,则创建关键字段并以创建的关键字段为主键。

[0160] 当所有的业务系统对应的标准数据均写入缓存中之后,终端对同一行的标准数据进行去重处理,去重处理有助于减少重复的标准数据,节约缓存空间,且由于去重处理之后,减少了标准数据的数据量,因此将去重处理后的标准数据写入数据中心能够加快数据写入的速度,提高数据写入的效率。

[0161] 所述获取模块204,用于基于预设的数据处理框架,采用预设的窗口从所述数据中心中获取目标数据。

[0162] 其中,所述预设的数据处理框架为基于Spark的大数据处理框架。

[0163] 其中,所述预设的窗口为5秒。采用预设的窗口从所述数据中心加载标准数据,并将加载的标准数据确定为目标数据。

[0164] 所述校验模块205,用于在预先设置的定时器到期时,触发定时任务对所述目标数据进行校验。

[0165] 终端可以预先设置一个多时定时器,例如,设置在每个整点00分触发定时任务,以

对与所述目标数据进行验证,确保所述目标数据的有效性。

[0166] 在一个可选的实施例中,所述校验模块205触发定时任务对所述目标数据进行校验包括:

[0167] 确定所述目标数据对应的校验标识符;

[0168] 采用所述预设的窗口从所述数据中心中获取与所述校验标识符对应的校验数据;

[0169] 比对所述目标数据与所述校验数据是否一致;

[0170] 当所述目标数据与所述校验数据一致时,得到第一校验结果;

[0171] 当所述目标数据与所述校验数据不一致时,得到第二校验结果。

[0172] 该可选的实施例中,由于不同的业务系统对应的标准数据具有至少两个相同的字段,该至少两个相同的字段中的一个字段被确定为关键字段,则该至少两个相同的字段中的除关键字段外的其他字段均可以被确定为校验标识符,例如,所述校验标识符为身份证号码。

[0173] 终端确定校验标识符之后,基于所述预设的数据处理框架,采用预设的窗口从所述数据中心加载所述校验标识符对应的标准数据,并将所加载的标准数据确定为校验数据。

[0174] 终端将所述目标数据与校验数据中的每一个具有相同字段的数据进行比对,当所述目标数据与校验数据中的所有具有相同字段的数据比对一致时,确定所述目标数据与所述校验数据相同,即表明所述目标数据有效,得到第一校验结果;当所述目标数据与校验数据中的任意一个具有相同字段的数据比对不一致时,确定所述目标数据与所述校验数据不相同,即表明所述目标数据无效,得到第二校验结果。

[0175] 所述生成模块206,用于当校验得到的校验结果为目标结果时,根据所述目标数据生成数据报表。

[0176] 其中,所述目标结果为校验一致。

[0177] 终端在确定校验结果为目标结果时,生成数据报表,以实现数据的一个直观展示。进一步的,所述终端还可以具有如下功能:维度切换,排版切换,可删除/替换/追加图表,实现自由灵活的交互。

[0178] 在一个可选的实施例中,所述生成模块206根据所述目标数据生成数据报表包括:

[0179] 输入所述目标数据至预先定义的多个函数中进行计算,得到多个函数值;

[0180] 根据预设业务模板,基于所述多个函数值生成数据报表。

[0181] 其中,所述多个函数包括申请总和,申请通过率,放款量,放款件,平均通过时间,平均通过次数,同比增长值等。例如,输入所述目标数据至申请总和中进行计算,可计算得到申请总和。

[0182] 该可选的实施例中,根据预设业务模板生成数据报表,实现了不同数据源的业务数据的交叉实时处理,具有较高的时效性;并以数据报表的方式进行直观展示,用户体验高。

[0183] 所述警报模块207,用于比较每一个函数值与对应的告警阈值;当所述函数值大于所述告警阈值时,触发警报指令。

[0184] 所述建立模块208,用于为所述目标数据建立二叉树索引。

[0185] 在一个可选的实施例中,所述建立模块208为所述目标数据建立二叉树索引包括:

- [0186] 获取所述目标数据的推送时间；
- [0187] 根据所述推送时间将所述目标数据分成多个分区；
- [0188] 计算每个分区中的目标数据的数据量；
- [0189] 确定大于预设数据量阈值的目标分区；
- [0190] 为所述目标分区中的目标数据建立二叉树索引。
- [0191] 例如,将推送时间为T1(例如,1月份)的目标数据分在第一分区,将推送时间为T2(例如,2月份)的目标数据分在第二分区等等。每个分区的目标数据的数据量可以相同或者不同。
- [0192] 当某个分区中的目标数据的数据量大于预设数据量阈值时,表明该分区中的目标数据较多,若需从该分区中查询目标数据时,需要耗费较多的查询时间。当某个分区中的目标数据的数据量小于或者等于预设数据量阈值时,表明该分区中的目标数据较少,若需从该分区中查询目标数据时,不需要耗费较多的查询时间。
- [0193] 上述可选的实施例中,通过将目标数据存储在不同的分区中,便于后续查询时,能够从对应的分区中查询目标数据,而不需对所有的目标数据中的每一条目标数据进行逐一比对,减少数据查询时间;而根据每个分区中的目标数据的数据量确定是否建立二叉树索引,通过二叉树索引查询目标数据,能够进一步减少数据查询时间。
- [0194] 在一个可选的实施例中,所述为所述目标分区中的目标数据建立二叉树索引包括:
- [0195] 从所述目标数据中随机选取两个目标数据作为初始中心点;
- [0196] 利用预设算法,基于所述初始中心点进行迭代聚类,在迭代次数大于1的当前迭代的上一次迭代中能得到至少一组候选类别,所述至少一组候选类别中每组候选类别对应第一候选类别及第二候选类别;
- [0197] 所述每组候选类别对应的二叉树节点表示每组候选类别中第一候选类别与第二候选类别间的差分特征值;
- [0198] 在当前迭代中,分别对所述每组候选类别中第一候选类别及第二候选类别分别进行迭代聚类,得到所述每组候选类别中第一候选类别对应的两个候选类别及所述每组候选类别中第二候选类别对应的两个候选类别;
- [0199] 将表示所述每组候选类别中第一候选类别对应的两个候选类别的差分特征的二叉树节点作为所述每组候选类别对应的二叉树节点的左节点或右节点,将表示所述每组候选类别中第二候选类别对应的两个候选类别的差分特征的二叉树节点作为所述每组候选类别对应的二叉树节点的右节点或左节点;
- [0200] 执行多次迭代聚类,直至达到迭代终止条件,得到所述目标分区对应的多个类别,其中二叉树的一个叶子节点表示一个类别的目标数据。
- [0201] 在本发明的优选实施例中,在对每个区中的目标数据生成二叉树索引时,在所述二叉树索引中,二叉树的一个叶子节点表示一个类别的目标数据,除叶子节点外的每个节点表示两个候选类别的差分特征,除叶子节点及根节点外任意一个节点对应的两个候选类别由对所述任意一个节点的父节点对应的两个候选类别中第一候选类别分类得到的,所述任意一个节点的兄弟节点对应的两个候选类别由对所述任意一个节点的父节点对应的两个候选类别中第二候选类别分类得到的。这样建立二叉树索引后,后续在对数据进行搜索,

无需对每一条的数据进行比对,大大提升了搜索速度。

[0202] 在一个可选的实施例中,所述查询模块209,用于接收业务数据的查询请求并解析出所述查询请求得到查询字段及查询时间;根据所述查询时间确定查询分区;判断所述查询分区是否为目标分区;当所述查询分区为目标分区时,基于所述二叉树索引查询与所述查询字段对应的业务数据。

[0203] 该可选的实施例中,由于查询分区中的目标数据是以推送时间进行存储的,因此根据查询时间即可确定查询分区。

[0204] 所述目标分区是指建立有二叉树索引的分区,每个分区对应有分区标识,根据分区标识即可确定所述查询分区是否为目标分区。

[0205] 当所述查询分区为目标分区时,采用优先队列从所述二叉树索引的根节点开始遍历,将遍历的节点插入在所述优先队列中,基于所述查询字段对应的业务数据与遍历的节点表示的业务数据的距离,对所述优先队列进行从大到小排序,将排序在第一的业务数据推送给用户。

[0206] 基于所述查询字段对应的业务数据与遍历的节点表示的业务数据的距离进行查询,能够避免遍历二叉树的所有节点,提高了查询速度。在没有建立二叉树索引的分区中,基于现有技术中的深度优先搜索算法搜索与所述查询字段对应的业务数据。

[0207] 需要强调的是,为进一步保证上述二叉树索引的私密性和安全性,上述二叉树索引可存储于区块链的节点中。

[0208] 参阅图3所示,为本发明实施例三提供的终端的结构示意图。在本发明较佳实施例中,所述终端3包括存储器31、至少一个处理器32、至少一条通信总线33及收发器34。

[0209] 本领域技术人员应该了解,图3示出的终端的结构并不构成本发明实施例的限定,既可以是总线型结构,也可以是星形结构,所述终端3还可以包括比图示更多或更少的其他硬件或者软件,或者不同的部件布置。

[0210] 在一些实施例中,所述终端3是一种能够按照事先设定或存储的指令,自动进行数值计算和/或信息处理的终端,其硬件包括但不限于微处理器、专用集成电路、可编程门阵列、数字处理器及嵌入式设备等。所述终端3还可包括客户设备,所述客户设备包括但不限于任何一种可与客户通过键盘、鼠标、遥控器、触摸板或声控设备等方式进行人机交互的电子产品,例如,个人计算机、平板电脑、智能手机、数码相机等。

[0211] 需要说明的是,所述终端3仅为举例,其他现有的或今后可能出现的电子产品如可适应于本发明,也应包含在本发明的保护范围以内,并以引用方式包含于此。

[0212] 在一些实施例中,所述存储器31中存储有计算机程序,所述计算机程序被所述至少一个处理器32执行时实现如所述的多源业务数据实时处理方法中的全部或者部分步骤。所述存储器31包括只读存储器(Read-Only Memory,ROM)、可编程只读存储器(Programmable Read-Only Memory, PROM)、可擦除可编程只读存储器(Erasable Programmable Read-Only Memory, EPROM)、一次可编程只读存储器(One-time Programmable Read-Only Memory, OTPROM)、电子擦除式可复写只读存储器(Electrically-Erasable Programmable Read-Only Memory, EEPROM)、只读光盘(Compact Disc Read-Only Memory, CD-ROM)或其他光盘存储器、磁盘存储器、磁带存储器、或者能够用于携带或存储数据的计算机可读的任何其他介质。

[0213] 进一步地,所述计算机可读存储介质可主要包括存储程序区和存储数据区,其中,存储程序区可存储操作系统、至少一个功能所需的应用程序等;存储数据区可存储根据区块链节点的使用所创建的数据等。

[0214] 本发明所指区块链是分布式数据存储、点对点传输、共识机制、加密算法等计算机技术的新型应用模式。区块链(Blockchain),本质上是一个去中心化的数据库,是一串使用密码学方法相关联产生的数据块,每一个数据块中包含了一批次网络交易的信息,用于验证其信息的有效性(防伪)和生成下一个区块。区块链可以包括区块链底层平台、平台产品服务层以及应用服务层等。

[0215] 在一些实施例中,所述至少一个处理器32是所述终端3的控制核心(Control Unit),利用各种接口和线路连接整个终端3的各个部件,通过运行或执行存储在所述存储器31内的程序或者模块,以及调用存储在所述存储器31内的数据,以执行终端3的各种功能和处理数据。例如,所述至少一个处理器32执行所述存储器中存储的计算机程序时实现本发明实施例中所述的多源业务数据实时处理方法的全部或者部分步骤;或者实现多源业务数据实时处理装置的全部或者部分功能。所述至少一个处理器32可以由集成电路组成,例如可以由单个封装的集成电路所组成,也可以是由多个相同功能或不同功能封装的集成电路所组成,包括一个或者多个中央处理器(Central Processing unit,CPU)、微处理器、数字处理芯片、图形处理器及各种控制芯片的组合等。

[0216] 在一些实施例中,所述至少一条通信总线33被设置为实现所述存储器31以及所述至少一个处理器32等之间的连接通信。

[0217] 尽管未示出,所述终端3还可以包括给各个部件供电的电源(比如电池),优选的,电源可以通过电源管理装置与所述至少一个处理器32逻辑相连,从而通过电源管理装置实现管理充电、放电、以及功耗管理等功能。电源还可以包括一个或一个以上的直流或交流电源、再充电装置、电源故障检测电路、电源转换器或者逆变器、电源状态指示器等任意组件。所述终端3还可以包括多种传感器、蓝牙模块、Wi-Fi模块等,在此不再赘述。

[0218] 上述以软件功能模块的形式实现的集成的单元,可以存储在一个计算机可读存储介质中。上述软件功能模块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,终端,或者网络设备等)或处理器(processor)执行本发明各个实施例所述方法的部分。

[0219] 在本发明所提供的几个实施例中,应该理解到,所揭露的装置和方法,可以通过其它的方式实现。例如,以上所描述的装置实施例仅仅是示意性的,例如,所述模块的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式。

[0220] 所述作为分离部件说明的模块可以是或者也可以不是物理上分开的,作为模块显示的部件可以是或者也可以不是物理单元,既可以位于一个地方,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。

[0221] 另外,在本发明各个实施例中的各功能模块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。上述集成的单元既可以采用硬件的形式实现,也可以采用硬件加软件功能模块的形式实现。

[0222] 对于本领域技术人员而言,显然本发明不限于上述示范性实施例的细节,而且在

不背离本发明的精神或基本特征的情况下,能够以其他的具体形式实现本发明。因此,无论从哪一点来看,均应将实施例看作是示范性的,而且是非限制性的,本发明的范围由所附权利要求而不是上述说明限定,因此旨在将落在权利要求的等同要件的含义和范围内的所有变化涵括在本发明内。不应将权利要求中的任何附图标记视为限制所涉及的权利要求。此外,显然“包括”一词不排除其他单元或,单数不排除复数。本发明陈述的多个单元或装置也可以由一个单元或装置通过软件或者硬件来实现。第一,第二等词语用来表示名称,而并不表示任何特定的顺序。

[0223] 最后应说明的是,以上实施例仅用以说明本发明的技术方案而非限制,尽管参照较佳实施例对本发明进行了详细说明,本领域的普通技术人员应当理解,可以对本发明的技术方案进行修改或等同替换,而不脱离本发明技术方案的精神和范围。

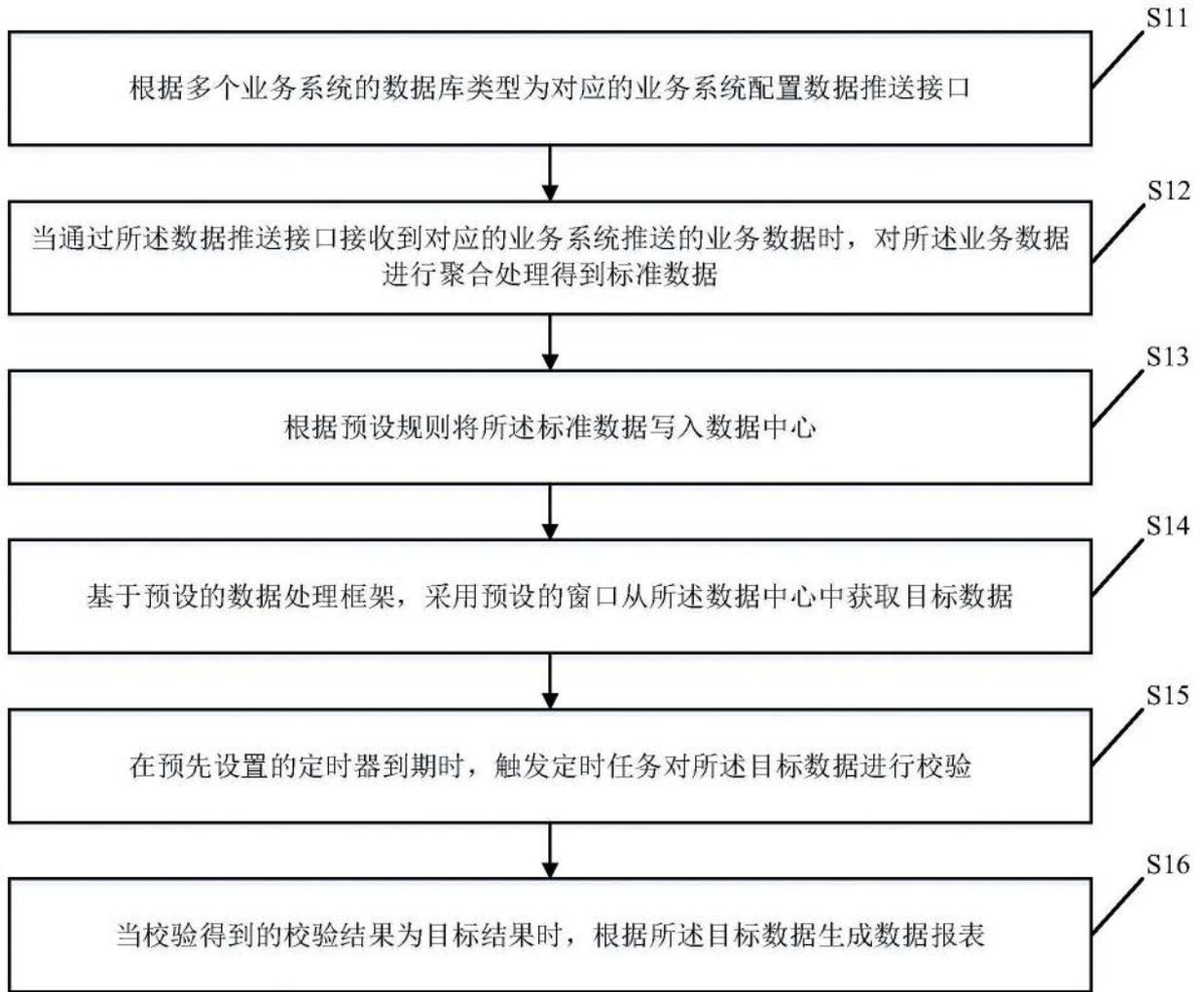


图1

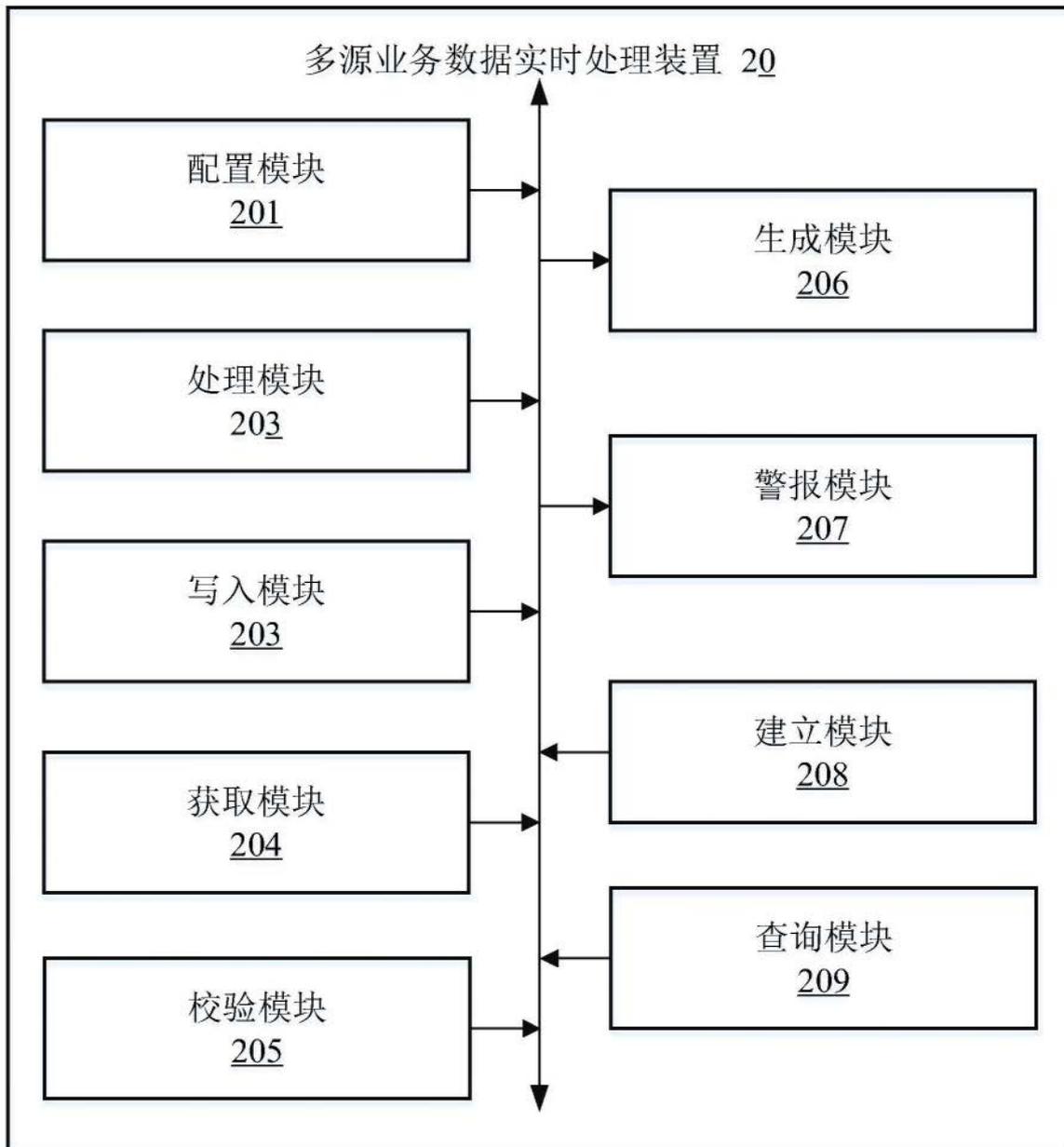


图2

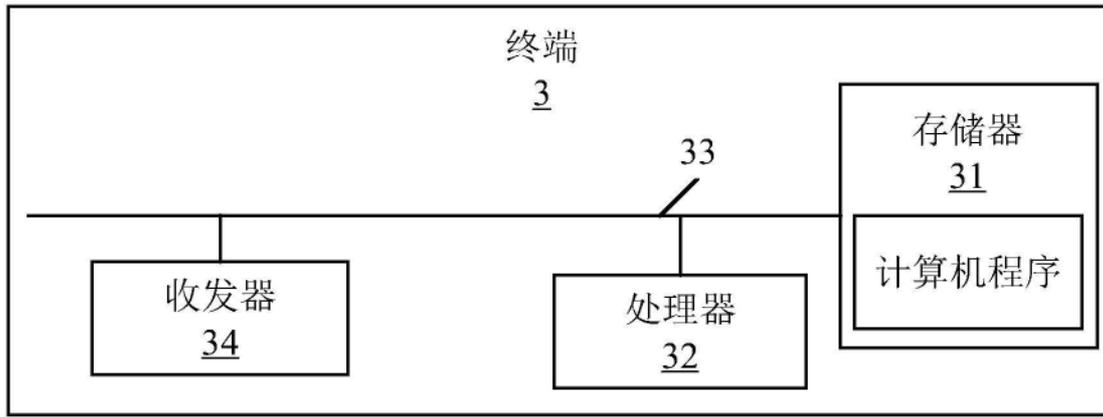


图3