



(12) 发明专利申请

(10) 申请公布号 CN 117370558 A

(43) 申请公布日 2024. 01. 09

(21) 申请号 202311344318.7

(22) 申请日 2023.10.17

(71) 申请人 中国平安财产保险股份有限公司  
地址 518000 广东省深圳市福田区益田路  
5033号平安金融中心12、13、38、39、40  
层

(72) 发明人 卢显锋 吴清苗

(74) 专利代理机构 深圳市世联合知识产权代理  
有限公司 44385  
专利代理师 郝少剑

(51) Int. Cl.

G06F 16/35 (2019.01)

G06F 40/30 (2020.01)

G06F 21/60 (2013.01)

G06Q 40/00 (2023.01)

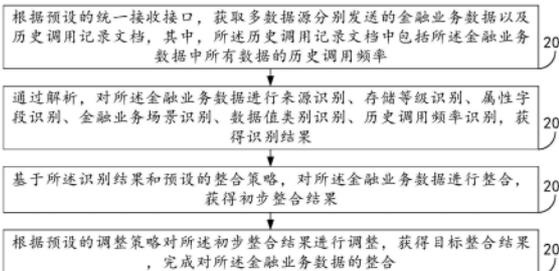
权利要求书4页 说明书12页 附图6页

(54) 发明名称

一种数据整合方法、装置、设备及其存储介质

(57) 摘要

本申请实施例属于金融科技技术领域,应用于金融业务数据多数据源整合场景中,涉及一种数据整合方法、装置、设备及其存储介质,包括根据预设的统一接收接口,获取多数据源分别发送的金融业务数据以及历史调用记录文档;对金融业务数据进行来源、存储等级、属性字段、金融业务场景、数据值类别、历史调用频率识别;对金融业务数据进行初步整合;根据预设的调整策略对初步整合结果进行调整,完成对所述金融业务数据的整合。通过整合策略,保证了多数据源的金融业务数据的合理存储,通过调整策略,保证目标金融业务数据在被调用时,存在多个可调用分区时,预先提供数据量为最小值时的调用分区信息,方便后续业务调用,合理进行多数据源数据整合。



1. 一种数据整合方法,其特征在于,包括下述步骤:

根据预设的统一接收接口,获取多数据源分别发送的金融业务数据以及历史调用记录文档,其中,所述历史调用记录文档中包括所述金融业务数据中所有数据的历史调用频率;

通过解析,对所述金融业务数据进行来源识别、存储等级识别、属性字段识别、金融业务场景识别、数据值类别识别、历史调用频率识别,获得识别结果;

基于所述识别结果和预设的整合策略,对所述金融业务数据进行整合,获得初步整合结果;

根据预设的调整策略对所述初步整合结果进行调整,获得目标整合结果,完成对所述金融业务数据的整合。

2. 根据权利要求1所述的数据整合方法,其特征在于,所述对所述金融业务数据进行来源识别、存储等级识别、属性字段识别、金融业务场景识别、数据值类别识别、历史调用频率识别,获得识别结果的步骤,具体包括:

根据每条金融业务数据的业务渠道标识,进行来源识别,识别出所有业务渠道分别发送的金融业务数据;

根据每条金融业务数据的存储等级标识,进行存储等级识别,识别出每条金融业务数据对应的存储等级;

根据每条金融业务数据分别所包含的数据属性字段,进行属性字段识别,识别出所有数据属性字段分别对应的数据值;

根据预设的金融业务场景标识,对每条金融业务数据进行金融业务场景识别,识别出每条金融业务数据分别所属的金融业务场景;

根据预设的数据值类别,对每条金融业务数据所包含的数据属性字段的数据值进行数据值类别识别,其中,所述预设的数据值类别包括字符数据类别和数值数据类别;

根据所述历史调用记录文档,对每条金融业务数据所包含的数据属性字段的数据值进行调用频率识别,识别出每个数据值的历史调用频率。

3. 根据权利要求2所述的数据整合方法,其特征在于,所述基于所述识别结果和预设的整合策略,对所述金融业务数据进行整合,获得初步整合结果的步骤,具体包括:

根据所述业务渠道标识的不同,对所有业务渠道分别发送的金融业务数据进行第一分区处理,获取第一分区处理后所有分区内的金融业务数据,并将所述业务渠道标识一一设为相应的第一分区标识;

根据所述每条金融业务数据对应的存储等级,对所有金融业务数据进行第二分区处理,获取第二分区处理后所有分区内的金融业务数据,并将所述存储等级标识一一设为相应的第二分区标识,所述存储等级标识根据数据是否需要加密以及加密等级而设定;

根据所述识别结果,识别出每条金融业务数据所包含的数据属性字段,根据所述数据属性字段对所有金融业务数据进行第三分区处理,获取第三分区处理后所有分区内的金融业务数据,并将所述数据属性字段一一设为相应的第三分区标识;

根据每条金融业务数据分别所属的金融业务场景,对所有金融业务数据进行第四分区处理,获得第四分区处理后所有分区内的金融业务数据,并将金融业务场景标识一一设为相应的第四分区标识;

根据每条金融业务数据所包含的数据属性字段的数据值类别,对所有金融业务数据进

行第五分区处理,获得第五分区处理后所有分区内的金融业务数据,其中,所述第五分区处理后所有分区包括字符数据类别分区、数值数据类别分区和杂合性数据类别分区,所述字符数据类别分区内每条金融业务数据所包含的数据属性字段的数据值都为字符数据,所述数值数据类别分区内每条金融业务数据所包含的数据属性字段的数据值都为数值数据,所述杂合性数据类别分区内每条金融业务数据所包含的数据属性字段的数据值即存在字符数据也存在数值数据;

根据每条金融业务数据所包含的数据属性字段的数据值的历史调用频率,以及预设的不同调用频率区别,对所有金融业务数据进行第六分区处理,获得第六分区处理后所有分区内的金融业务数据,其中,第六分区处理后不同分区分别对应预设的不同调用频率区间;

获得第一分区处理结果、第二分区处理结果、第三分区处理结果、第四分区处理结果、第五分区处理结果以及第六分区处理结果,作为所述初步整合结果。

4. 根据权利要求3所述的数据整合方法,其特征在于,在执行所述根据所述业务渠道标识的不同,对所有业务渠道分别发送的金融业务数据进行第一分区处理,获取第一分区处理后所有分区内的金融业务数据,并将所述业务渠道标识一一设为相应的第一分区标识的步骤之后,所述方法还包括:

根据所述统一接收接口,获取目标数据源最新发送的金融业务数据以及历史调用记录文档;

以目标数据源对应的业务渠道标识作为筛选字段,进行第一分区筛选;

若筛选结果为空值,则在目标数据仓库内新增第一分区,并以所述目标数据源对应的业务渠道标识作为所述新增第一分区的分区标识;

若筛选结果非为空值,则将所述目标数据源最新发送的金融业务数据添加到所述筛选字段对应的第一分区内。

5. 根据权利要求3所述的数据整合方法,其特征在于,在执行所述根据所述识别结果,识别出每条金融业务数据所包含的数据属性字段,根据所述数据属性字段对所有金融业务数据进行第三分区处理,获取第三分区处理后所有分区内的金融业务数据,并将所述数据属性字段一一设为相应的第三分区标识的步骤之后,所述方法还包括:

通过语义识别方式,筛选出语义相同的数据属性字段,构建属性字段集合;

获取所述语义相同的数据属性字段分别对应的不同分区,进行合并,获得合并后的分区;

获取所述合并后的分区内的金融业务数据作为所述属性字段集合内所有数据属性字段所对应的金融业务数据。

6. 根据权利要求3所述的数据整合方法,其特征在于,所述根据每条金融业务数据所包含的数据属性字段的数据值的历史调用频率,以及预设的不同调用频率区间,对所有金融业务数据进行第六分区处理,获得第六分区处理后所有分区内的金融业务数据的步骤,具体包括:

获取目标金融业务数据所包含的所有数据属性字段,并识别出所述数据属性字段分别所对应的数据值;

根据所述数据属性字段分别所对应的数据值,确定所述目标金融业务数据所包含的所有数据值;

根据所述历史调用记录文档,识别所述所有数据值分别对应的历史调用频率;  
通过比较,从所述所有数据值分别对应的历史调用频率中筛选出历史调用频率最大值;

以所述历史调用频率最大值作为所述目标金融业务数据的最大调用频率;

根据所述最大调用频率确定所述目标金融业务数据所对应的调用频率区间;

基于所述目标金融业务数据所对应的调用频率区间,确定所述目标金融业务数据所对应的第六分区处理后的分区。

7.根据权利要求1或3所述的数据整合方法,其特征在于,所述根据预设的调整策略对所述初步整合结果进行调整,获得目标整合结果,完成对所述金融业务数据的整合的步骤,具体包括:

步骤601,根据所述初步整合结果,识别出目标金融业务数据经第一分区处理后所处的分区,计算所述分区内的数据量,记为第一数据量;

步骤602,根据所述初步整合结果,识别出所述目标金融业务数据经第二分区处理后所处的分区,计算所述分区内的数据量,记为第二数据量;

步骤603,根据所述初步整合结果,识别出目标金融业务数据经第三分区处理后所处的分区,计算所述分区内的数据量,记为第三数据量;

步骤604,根据所述初步整合结果,识别出所述目标金融业务数据经第四分区处理后所处的分区,计算所述分区内的数据量,记为第四数据量;

步骤605,根据所述初步整合结果,识别出目标金融业务数据经第五分区处理后所处的分区,计算所述分区内的数据量,记为第五数据量;

步骤606,根据所述初步整合结果,识别出所述目标金融业务数据经第六分区处理后所处的分区,计算所述分区内的数据量,记为第六数据量;

步骤607,比较所述第一数据量、第二数据量、第三数据量、第四数据量、第五数据量和第六数据量之间的大小关系,并从中筛选出最小值;

步骤608,识别所述最小值对应的分区,并将所述分区的标识信息作为所述目标金融业务数据的调用分区信息;

步骤609,将每条金融业务数据依次作为所述目标金融业务数据,重复执行步骤601至步骤608,获取每条金融业务数据的调用分区信息,作为所述目标整合结果。

8.一种数据整合装置,其特征在于,包括:

数据获取模块,用于根据预设的统一接收接口,获取多数据源分别发送的金融业务数据以及历史调用记录文档,其中,所述历史调用记录文档中包括所述金融业务数据中所有数据的历史调用频率;

数据解析识别模块,用于通过解析,对所述金融业务数据进行来源识别、存储等级识别、属性字段识别、金融业务场景识别、数据值类别识别、历史调用频率识别,获得识别结果;

初步整合模块,用于基于所述识别结果和预设的整合策略,对所述金融业务数据进行整合,获得初步整合结果;

初步整合结果调整模块,用于根据预设的调整策略对所述初步整合结果进行调整,获得目标整合结果,完成对所述金融业务数据的整合。

9. 一种计算机设备,其特征在于,包括存储器和处理器,所述存储器中存储有计算机可读指令,所述处理器执行所述计算机可读指令时实现如权利要求1至7中任一项所述的数据整合方法的步骤。

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可读指令,所述计算机可读指令被处理器执行时实现如权利要求1至7中任一项所述的数据整合方法的步骤。

## 一种数据整合方法、装置、设备及其存储介质

### 技术领域

[0001] 本申请涉及金融科技技术领域,应用于金融业务数据多数据源整合场景中,尤其涉及一种数据整合方法、装置、设备及其存储介质。

### 背景技术

[0002] 随着互联网的飞速发展,各行各业都在依托互联网寻求行业突破点,近年来,金融行业也正在围绕着互联网进行线上业务拓展。目前,随着人们金融业务意识的提高,例如,越来越多的企业、家庭和个人在购买机动车辆后,会为车辆投保,购买车险。传统的车险销售主要是由业务员和车主沟通,向车主推荐各类车险。

[0003] 目前金融行业的销售模式主要有三种:电话/网上销售、代理人销售、合作商代售,不同的销售模式之间由于各种原因等导致客户数据无法统一,具体体现在客户信息互不相通、信息内容内容不统一、客户数据管理不统一;例如,保险业务中电话销售主要是通过管理电话销售坐席,通过销售坐席对客户进行销售沟通;而代理销售则主要是通过扩张/维护代理,通过代理人对客户进行销售保险服务;车商销售主要是通过和车商进行合作的模式。上述方式虽然拓宽了业务渠道,但是也造成了无法对多种数据源的金融业务数据合理进行整合的问题。

### 发明内容

[0004] 本申请实施例的目的在于提出一种数据整合方法、装置、设备及其存储介质,以解决现有金融业务在多业务渠道下无法对多种数据源的金融业务数据合理进行整合的问题。

[0005] 为了解决上述技术问题,本申请实施例提供数据整合方法,采用了如下所述的技术方案:

[0006] 一种数据整合方法,包括下述步骤:

[0007] 根据预设的统一接收接口,获取多数据源分别发送的金融业务数据以及历史调用记录文档,其中,所述历史调用记录文档中包括所述金融业务数据中所有数据的历史调用频率;

[0008] 通过解析,对所述金融业务数据进行来源识别、存储等级识别、属性字段识别、金融业务场景识别、数据值类别识别、历史调用频率识别,获得识别结果;

[0009] 基于所述识别结果和预设的整合策略,对所述金融业务数据进行整合,获得初步整合结果;

[0010] 根据预设的调整策略对所述初步整合结果进行调整,获得目标整合结果,完成对所述金融业务数据的整合。

[0011] 进一步的,所述对所述金融业务数据进行来源识别、存储等级识别、属性字段识别、金融业务场景识别、数据值类别识别、历史调用频率识别,获得识别结果的步骤,具体包括:

[0012] 根据每条金融业务数据的业务渠道标识,进行来源识别,识别出所有业务渠道分

别发送的金融业务数据；

[0013] 根据每条金融业务数据的存储等级标识,进行存储等级识别,识别出每条金融业务数据对应的存储等级；

[0014] 根据每条金融业务数据分别所包含的数据属性字段,进行属性字段识别,识别出所有数据属性字段分别对应的数据值；

[0015] 根据预设的金融业务场景标识,对每条金融业务数据进行金融业务场景识别,识别出每条金融业务数据分别所属的金融业务场景；

[0016] 根据预设的数据值类别,对每条金融业务数据所包含的数据属性字段的数据值进行数据值类别识别,其中,所述预设的数据值类别包括字符数据类别和数值数据类别；

[0017] 根据所述历史调用记录文档,对每条金融业务数据所包含的数据属性字段的数据值进行调用频率识别,识别出每个数据值的历史调用频率。

[0018] 进一步的,所述基于所述识别结果和预设的整合策略,对所述金融业务数据进行整合,获得初步整合结果的步骤,具体包括：

[0019] 根据所述业务渠道标识的不同,对所有业务渠道分别发送的金融业务数据进行第一分区处理,获取第一分区处理后所有分区内的金融业务数据,并将所述业务渠道标识一一设为相应的第一分区标识；

[0020] 根据所述每条金融业务数据对应的存储等级,对所有金融业务数据进行第二分区处理,获取第二分区处理后所有分区内的金融业务数据,并将所述存储等级标识一一设为相应的第二分区标识,所述存储等级标识根据数据是否需要加密以及加密等级而设定；

[0021] 根据所述识别结果,识别出每条金融业务数据所包含的数据属性字段,根据所述数据属性字段对所有金融业务数据进行第三分区处理,获取第三分区处理后所有分区内的金融业务数据,并将所述数据属性字段一一设为相应的第三分区标识；

[0022] 根据每条金融业务数据分别所属的金融业务场景,对所有金融业务数据进行第四分区处理,获得第四分区处理后所有分区内的金融业务数据,并将金融业务场景标识一一设为相应的第四分区标识；

[0023] 根据每条金融业务数据所包含的数据属性字段的数据值类别,对所有金融业务数据进行第五分区处理,获得第五分区处理后所有分区内的金融业务数据,其中,所述第五分区处理后所有分区包括字符数据类别分区、数值数据类别分区和杂合性数据类别分区,所述字符数据类别分区内每条金融业务数据所包含的数据属性字段的数据值都为字符数据,所述数值数据类别分区内每条金融业务数据所包含的数据属性字段的数据值都为数值数据,所述杂合性数据类别分区内每条金融业务数据所包含的数据属性字段的数据值即存在字符数据也存在数值数据；

[0024] 根据每条金融业务数据所包含的数据属性字段的数据值的历史调用频率,以及预设的不同调用频率区别,对所有金融业务数据进行第六分区处理,获得第六分区处理后所有分区内的金融业务数据,其中,第六分区处理后不同分区分别对应预设的不同调用频率区间；

[0025] 获得第一分区处理结果、第二分区处理结果、第三分区处理结果、第四分区处理结果、第五分区处理结果以及第六分区处理结果,作为所述初步整合结果。

[0026] 进一步的,在执行所述根据所述业务渠道标识的不同,对所有业务渠道分别发送

的金融业务数据进行第一分区处理,获取第一分区处理后所有分区内的金融业务数据,并将所述业务渠道标识一一设为相应的第一分区标识的步骤之后,所述方法还包括:

[0027] 根据所述统一接收接口,获取目标数据源最新发送的金融业务数据以及历史调用记录文档;

[0028] 以目标数据源对应的业务渠道标识作为筛选字段,进行第一分区筛选;

[0029] 若筛选结果为空值,则在目标数据仓库内新增第一分区,并以所述目标数据源对应的业务渠道标识作为所述新增第一分区的分区标识;

[0030] 若筛选结果非为空值,则将所述目标数据源最新发送的金融业务数据添加到所述筛选字段对应的第一分区内。

[0031] 进一步的,在执行所述根据所述识别结果,识别出每条金融业务数据所包含的数据属性字段,根据所述数据属性字段对所有金融业务数据进行第三分区处理,获取第三分区处理后所有分区内的金融业务数据,并将所述数据属性字段一一设为相应的第三分区标识的步骤之后,所述方法还包括:

[0032] 通过语义识别方式,筛选出语义相同的数据属性字段,构建属性字段集合;

[0033] 获取所述语义相同的数据属性字段分别对应的不同分区,进行合并,获得合并后的分区;

[0034] 获取所述合并后的分区内的金融业务数据作为所述属性字段集合内所有数据属性字段所对应的金融业务数据。

[0035] 进一步的,所述根据每条金融业务数据所包含的数据属性字段的数据值的历史调用频率,以及预设的不同调用频率区间,对所有金融业务数据进行第六分区处理,获得第六分区处理后所有分区内的金融业务数据的步骤,具体包括:

[0036] 获取目标金融业务数据所包含的所有数据属性字段,并识别出所述数据属性字段分别所对应的数据值;

[0037] 根据所述数据属性字段分别所对应的数据值,确定所述目标金融业务数据所包含的所有数据值;

[0038] 根据所述历史调用记录文档,识别所述所有数据值分别对应的历史调用频率;

[0039] 通过比较,从所述所有数据值分别对应的历史调用频率中筛选出历史调用频率最大值;

[0040] 以所述历史调用频率最大值作为所述目标金融业务数据的最大调用频率;

[0041] 根据所述最大调用频率确定所述目标金融业务数据所对应的调用频率区间;

[0042] 基于所述目标金融业务数据所对应的调用频率区间,确定所述目标金融业务数据所对应的第六分区处理后的分区。

[0043] 进一步的,所述根据预设的调整策略对所述初步整合结果进行调整,获得目标整合结果,完成对所述金融业务数据的整合的步骤,具体包括:

[0044] 步骤601,根据所述初步整合结果,识别出目标金融业务数据经第一分区处理后所处的分区,计算所述分区内的数据量,记为第一数据量;

[0045] 步骤602,根据所述初步整合结果,识别出所述目标金融业务数据经第二分区处理后所处的分区,计算所述分区内的数据量,记为第二数据量;

[0046] 步骤603,根据所述初步整合结果,识别出目标金融业务数据经第三分区处理后所

处的分区,计算所述分区内的数据量,记为第三数据量;

[0047] 步骤604,根据所述初步整合结果,识别出所述目标金融业务数据经第四分区处理后所处的分区,计算所述分区内的数据量,记为第四数据量;

[0048] 步骤605,根据所述初步整合结果,识别出目标金融业务数据经第五分区处理后所处的分区,计算所述分区内的数据量,记为第五数据量;

[0049] 步骤606,根据所述初步整合结果,识别出所述目标金融业务数据经第六分区处理后所处的分区,计算所述分区内的数据量,记为第六数据量;

[0050] 步骤607,比较所述第一数据量、第二数据量、第三数据量、第四数据量、第五数据量和第六数据量之间的大小关系,并从中筛选出最小值;

[0051] 步骤608,识别所述最小值对应的分区,并将所述分区的标识信息作为所述目标金融业务数据的调用分区信息;

[0052] 步骤609,将每条金融业务数据依次作为所述目标金融业务数据,重复执行步骤601至步骤608,获取每条金融业务数据的调用分区信息,作为所述目标整合结果。

[0053] 为了解决上述技术问题,本申请实施例还提供数据整合装置,采用了如下所述的技术方案:

[0054] 一种数据整合装置,包括:

[0055] 数据获取模块,用于根据预设的统一接收接口,获取多数据源分别发送的金融业务数据以及历史调用记录文档,其中,所述历史调用记录文档中包括所述金融业务数据中所有数据的历史调用频率;

[0056] 数据解析识别模块,用于通过解析,对所述金融业务数据进行来源识别、存储等级识别、属性字段识别、金融业务场景识别、数据值类别识别、历史调用频率识别,获得识别结果;

[0057] 初步整合模块,用于基于所述识别结果和预设的整合策略,对所述金融业务数据进行整合,获得初步整合结果;

[0058] 初步整合结果调整模块,用于根据预设的调整策略对所述初步整合结果进行调整,获得目标整合结果,完成对所述金融业务数据的整合。

[0059] 为了解决上述技术问题,本申请实施例还提供一种计算机设备,采用了如下所述的技术方案:

[0060] 一种计算机设备,包括存储器和处理器,所述存储器中存储有计算机可读指令,所述处理器执行所述计算机可读指令时实现上述所述的数据整合方法的步骤。

[0061] 为了解决上述技术问题,本申请实施例还提供一种计算机可读存储介质,采用了如下所述的技术方案:

[0062] 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可读指令,所述计算机可读指令被处理器执行时实现如上述所述的数据整合方法的步骤。

[0063] 与现有技术相比,本申请实施例主要有以下有益效果:

[0064] 本申请实施例所述数据整合方法,通过根据预设的统一接收接口,获取多数据源分别发送的金融业务数据以及历史调用记录文档,其中,所述历史调用记录文档中包括所述金融业务数据中所有数据的历史调用频率;通过解析,对所述金融业务数据进行来源识别、存储等级识别、属性字段识别、金融业务场景识别、数据值类别识别、历史调用频率识

别,获得识别结果;基于所述识别结果和预设的整合策略,对所述金融业务数据进行整合,获得初步整合结果;根据预设的调整策略对所述初步整合结果进行调整,获得目标整合结果,完成对所述金融业务数据的整合。通过所述整合策略,保证了多数据源的金融业务数据的合理存储,通过所述预设的调整策略,保证了目标金融业务数据在被调用时,存在多个对应的可调用分区时,为了避免调用到数据量较大的分区,直接预先提供数据量为最小值时每条金融业务数据分别对应的调用分区信息,方便后续业务调用,合理进行多数据源数据整合。

### 附图说明

[0065] 为了更清楚地说明本申请中的方案,下面将对本申请实施例描述中所需要使用的附图作一个简单介绍,显而易见地,下面描述中的附图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的前提下,还可以根据这些附图获得其他的附图。

[0066] 图1是本申请可以应用于其中的示例性系统架构图;

[0067] 图2是根据本申请的数据整合方法的一个实施例的流程图;

[0068] 图3是图2所示步骤202的一个具体实施例的流程图;

[0069] 图4是图2所示步骤203的一个具体实施例的流程图;

[0070] 图5是图4所示步骤406的一个具体实施例的流程图;

[0071] 图6是图2所示步骤204的一个具体实施例的流程图;

[0072] 图7是根据本申请的数据整合装置的一个实施例的结构示意图;

[0073] 图8是根据本申请的计算机设备的一个实施例的结构示意图。

### 具体实施方式

[0074] 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申请的技术领域的技术人员通常理解的含义相同;本文中在申请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,不是旨在于限制本申请;本申请的说明书和权利要求书及上述附图说明中的术语“包括”和“具有”以及它们的任何变形,意图在于覆盖不排他的包含。本申请的说明书和权利要求书或上述附图中的术语“第一”、“第二”等是用于区别不同对象,而不是用于描述特定顺序。

[0075] 在本文中提及“实施例”意味着,结合实施例描述的特定特征、结构或特性可以包含在本申请的至少一个实施例中。在说明书中的各个位置出现该短语并不一定均是指相同的实施例,也不是与其它实施例互斥的独立的或备选的实施例。本领域技术人员显式地和隐式地理解的是,本文所描述的实施例可以与其它实施例相结合。

[0076] 为了使本技术领域的人员更好地理解本申请方案,下面将结合附图,对本申请实施例中的技术方案进行清楚、完整地描述。

[0077] 如图1所示,系统架构100可以包括终端设备101、102、103,网络104和服务器105。网络104用以在终端设备101、102、103和服务器105之间提供通信链路的介质。网络104可以包括各种连接类型,例如有线、无线通信链路或者光纤电缆等等。

[0078] 用户可以使用终端设备101、102、103通过网络104与服务器105交互,以接收或发送消息等。终端设备101、102、103上可以安装有各种通讯客户端应用,例如网页浏览器应

用、购物类应用、搜索类应用、即时通信工具、邮箱客户端、社交平台软件等。

[0079] 终端设备101、102、103可以是具有显示屏并且支持网页浏览的各种电子设备,包括但不限于智能手机、平板电脑、电子书阅读器、MP3播放器(Moving Picture Experts Group Audio Layer III,动态影像专家压缩标准音频层面3)、MP4(Moving Picture Experts Group Audio Layer IV,动态影像专家压缩标准音频层面4)播放器、膝上型便携计算机和台式计算机等等。

[0080] 服务器105可以是提供各种服务的服务器,例如对终端设备101、102、103上显示的页面提供支持的后台服务器。

[0081] 需要说明的是,本申请实施例所提供的数据整合方法一般由服务器/终端设备执行,相应地,数据整合装置一般设置于服务器/终端设备中。

[0082] 应该理解,图1中的终端设备、网络和服务器的数目仅仅是示意性的。根据实现需要,可以具有任意数目的终端设备、网络和服务器。

[0083] 继续参考图2,示出了根据本申请的数据整合方法的一个实施例的流程图。所述的数据整合方法,包括以下步骤:

[0084] 步骤201,根据预设的统一接收接口,获取多数据源分别发送的金融业务数据以及历史调用记录文档,其中,所述历史调用记录文档中包括所述金融业务数据中所有数据的历史调用频率。

[0085] 本实施例中,所述多数据源包括不同业务渠道的金融业务数据存储库。

[0086] 步骤202,通过解析,对所述金融业务数据进行来源识别、存储等级识别、属性字段识别、金融业务场景识别、数据值类别识别、历史调用频率识别,获得识别结果。

[0087] 继续参考图3,图3是图2所示步骤202的一个具体实施例的流程图,包括:

[0088] 步骤301,根据每条金融业务数据的业务渠道标识,进行来源识别,识别出所有业务渠道分别发送的金融业务数据;

[0089] 步骤302,根据每条金融业务数据的存储等级标识,进行存储等级识别,识别出每条金融业务数据对应的存储等级;

[0090] 步骤303,根据每条金融业务数据分别所包含的数据属性字段,进行属性字段识别,识别出所有数据属性字段分别对应的数据值;

[0091] 步骤304,根据预设的金融业务场景标识,对每条金融业务数据进行金融业务场景识别,识别出每条金融业务数据分别所属的金融业务场景;

[0092] 步骤305,根据预设的数据值类别,对每条金融业务数据所包含的数据属性字段的数据值进行数据值类别识别,其中,所述预设的数据值类别包括字符数据类别和数值数据类别;

[0093] 步骤306,根据所述历史调用记录文档,对每条金融业务数据所包含的数据属性字段的数据值进行调用频率识别,识别出每个数据值的历史调用频率。

[0094] 通过对每条金融业务数据进行来源识别、存储等级识别、属性字段识别、金融业务场景识别、数据值类别识别、历史调用频率识别,便于对所述多数据源的金融业务数据进行整合。

[0095] 步骤203,基于所述识别结果和预设的整合策略,对所述金融业务数据进行整合,获得初步整合结果。

[0096] 继续参考图4,图4是图2所示步骤203的一个具体实施例的流程图,包括:

[0097] 步骤401,根据所述业务渠道标识的不同,对所有业务渠道分别发送的金融业务数据进行第一分区处理,获取第一分区处理后所有分区内的金融业务数据,并将所述业务渠道标识一一设为相应的第一分区标识;

[0098] 本实施例中,在执行所述根据所述业务渠道标识的不同,对所有业务渠道分别发送的金融业务数据进行第一分区处理,获取第一分区处理后所有分区内的金融业务数据,并将所述业务渠道标识一一设为相应的第一分区标识的步骤之后,所述方法还包括:根据所述统一接收接口,获取目标数据源最新发送的金融业务数据以及历史调用记录文档;以目标数据源对应的业务渠道标识作为筛选字段,进行第一分区筛选;若筛选结果为空值,则在目标数据仓库内新增第一分区,并以所述目标数据源对应的业务渠道标识作为所述新增第一分区的分区标识;若筛选结果非为空值,则将所述目标数据源最新发送的金融业务数据添加到所述筛选字段对应的第一分区内。

[0099] 通过进行第一分区标识筛选,保证了存在新业务渠道或者新的数据来源接入时,自动创建新的分区。

[0100] 步骤402,根据所述每条金融业务数据对应的存储等级,对所有金融业务数据进行第二分区处理,获取第二分区处理后所有分区内的金融业务数据,并将所述存储等级标识一一设为相应的第二分区标识,所述存储等级标识根据数据是否需要加密以及加密等级而设定;

[0101] 步骤403,根据所述识别结果,识别出每条金融业务数据所包含的数据属性字段,根据所述数据属性字段对所有金融业务数据进行第三分区处理,获取第三分区处理后所有分区内的金融业务数据,并将所述数据属性字段一一设为相应的第三分区标识;

[0102] 本实施例中,在执行所述根据所述识别结果,识别出每条金融业务数据所包含的数据属性字段,根据所述数据属性字段对所有金融业务数据进行第三分区处理,获取第三分区处理后所有分区内的金融业务数据,并将所述数据属性字段一一设为相应的第三分区标识的步骤之后,所述方法还包括:通过语义识别方式,筛选出语义相同的数据属性字段,构建属性字段集合;获取所述语义相同的数据属性字段分别对应的不同分区,进行合并,获得合并后的分区;获取所述合并后的分区内的金融业务数据作为所述属性字段集合内所有数据属性字段所对应的金融业务数据。

[0103] 由于多数据源可能涉及到在对数据记录时相同语义的属性字段,存在属性字段命名不一致的问题,因此,通过语义识别方式,筛选出语义相同的数据属性字段,构建属性字段集合,获取所述语义相同的数据属性字段分别对应的不同分区,进行合并,获得合并后的分区,使得相同语义的属性字段对应一个统一的分区,使得数据整合更加合理。

[0104] 步骤404,根据每条金融业务数据分别所属的金融业务场景,对所有金融业务数据进行第四分区处理,获得第四分区处理后所有分区内的金融业务数据,并将金融业务场景标识一一设为相应的第四分区标识;

[0105] 步骤405,根据每条金融业务数据所包含的数据属性字段的数据值类别,对所有金融业务数据进行第五分区处理,获得第五分区处理后所有分区内的金融业务数据;

[0106] 其中,所述第五分区处理后所有分区包括字符数据类别分区、数值数据类别分区和杂合性数据类别分区,所述字符数据类别分区内每条金融业务数据所包含的数据属性字

段的数据值都为字符数据,所述数值数据类别分区内每条金融业务数据所包含的数据属性字段的数据值都为数值数据,所述杂合性数据类别分区内每条金融业务数据所包含的数据属性字段的数据值即存在字符数据也存在数值数据;

[0107] 步骤406,根据每条金融业务数据所包含的数据属性字段的数据值的历史调用频率,以及预设的不同调用频率区别,对所有金融业务数据进行第六分区处理,获得第六分区处理后所有分区内的金融业务数据,其中,第六分区处理后不同分区分别对应预设的不同调用频率区间;

[0108] 继续参考图5,图5是图4所示步骤406的一个具体实施例的流程图,包括:

[0109] 步骤501,获取目标金融业务数据所包含的所有数据属性字段,并识别出所述数据属性字段分别所对应的数据值;

[0110] 步骤502,根据所述数据属性字段分别所对应的数据值,确定所述目标金融业务数据所包含的所有数据值;

[0111] 步骤503,根据所述历史调用记录文档,识别所述所有数据值分别对应的历史调用频率;

[0112] 步骤504,通过比较,从所述所有数据值分别对应的历史调用频率中筛选出历史调用频率最大值;

[0113] 步骤505,以所述历史调用频率最大值作为所述目标金融业务数据的最大调用频率;

[0114] 由于每条金融业务数据,无论调用了其包含的数据属性字段所对应的任何数据值,都是对整体金融业务数据进行了调用,因此,通过比较,从所述所有数据值分别对应的历史调用频率中筛选出历史调用频率最大值,以所述历史调用频率最大值作为所述目标金融业务数据的最大调用频率。便于结合历史调用频率对金融业务数据进行数据整合。

[0115] 步骤506,根据所述最大调用频率确定所述目标金融业务数据所对应的调用频率区间;

[0116] 步骤507,基于所述目标金融业务数据所对应的调用频率区间,确定所述目标金融业务数据所对应的第六分区处理后的分区。

[0117] 步骤407,获得第一分区处理结果、第二分区处理结果、第三分区处理结果、第四分区处理结果、第五分区处理结果以及第六分区处理结果,作为所述初步整合结果。

[0118] 通过所述整合策略,根据金融业务数据的来源、存储等级、属性字段、金融业务场景、数据值类别、历史调用频率进行数据整合,保证了多数据源的金融业务数据的合理存储。

[0119] 步骤204,根据预设的调整策略对所述初步整合结果进行调整,获得目标整合结果,完成对所述金融业务数据的整合。

[0120] 继续参考图6,图6是图2所示步骤204的一个具体实施例的流程图,包括:

[0121] 步骤601,根据所述初步整合结果,识别出目标金融业务数据经第一分区处理后所处的分区,计算所述分区内的数据量,记为第一数据量;

[0122] 步骤602,根据所述初步整合结果,识别出所述目标金融业务数据经第二分区处理后所处的分区,计算所述分区内的数据量,记为第二数据量;

[0123] 步骤603,根据所述初步整合结果,识别出目标金融业务数据经第三分区处理后所

处的分区,计算所述分区内的数据量,记为第三数据量;

[0124] 步骤604,根据所述初步整合结果,识别出所述目标金融业务数据经第四分区处理后所处的分区,计算所述分区内的数据量,记为第四数据量;

[0125] 步骤605,根据所述初步整合结果,识别出目标金融业务数据经第五分区处理后所处的分区,计算所述分区内的数据量,记为第五数据量;

[0126] 步骤606,根据所述初步整合结果,识别出所述目标金融业务数据经第六分区处理后所处的分区,计算所述分区内的数据量,记为第六数据量;

[0127] 步骤607,比较所述第一数据量、第二数据量、第三数据量、第四数据量、第五数据量和第六数据量之间的大小关系,并从中筛选出最小值;

[0128] 步骤608,识别所述最小值对应的分区,并将所述分区的标识信息作为所述目标金融业务数据的调用分区信息;

[0129] 步骤609,将每条金融业务数据依次作为所述目标金融业务数据,重复执行步骤601至步骤608,获取每条金融业务数据的调用分区信息,作为所述目标整合结果。

[0130] 通过所述预设的调整策略,即通过识别出每条金融业务数据分别所对应的分区,以及计算所对应分区的数据量,选择数据量为最小值时每条金融业务数据分别对应的分区,作为所述目标金融业务数据的调用分区信息。目的在于,保证了目标金融业务数据在被调用时,存在多个对应的可调用分区,为了避免调用到数据量较大的分区,直接预先提供数据量为最小值时每条金融业务数据分别对应的调用分区信息,方便后续业务调用,合理进行多数据源数据整合。

[0131] 本申请通过根据预设的统一接收接口,获取多数据源分别发送的金融业务数据以及历史调用记录文档,其中,所述历史调用记录文档中包括所述金融业务数据中所有数据的历史调用频率;通过解析,对所述金融业务数据进行来源识别、存储等级识别、属性字段识别、金融业务场景识别、数据值类别识别、历史调用频率识别,获得识别结果;基于所述识别结果和预设的整合策略,对所述金融业务数据进行整合,获得初步整合结果;根据预设的调整策略对所述初步整合结果进行调整,获得目标整合结果,完成对所述金融业务数据的整合。通过所述整合策略,根据金融业务数据的来源、存储等级、属性字段、金融业务场景、数据值类别、历史调用频率进行数据整合,保证了多数据源的金融业务数据的合理存储,通过所述预设的调整策略,保证了目标金融业务数据在被调用时,存在多个对应的可调用分区时,为了避免调用到数据量较大的分区,直接预先提供数据量为最小值时每条金融业务数据分别对应的调用分区信息,方便后续业务调用,合理进行多数据源数据整合。

[0132] 本申请实施例可以基于人工智能技术对相关的数据进行获取和处理。其中,人工智能(Artificial Intelligence, AI)是利用数字计算机或者数字计算机控制的机器模拟、延伸和扩展人的智能,感知环境、获取知识并使用知识获得最佳结果的理论、方法、技术及应用系统。

[0133] 人工智能基础技术一般包括如传感器、专用人工智能芯片、云计算、分布式存储、大数据整合技术、操作/交互系统、机电一体化等技术。人工智能软件技术主要包括计算机视觉技术、机器人技术、生物识别技术、语音处理技术、自然语言处理技术以及机器学习/深度学习等几大方向。

[0134] 本申请实施例中,通过根据预设的统一接收接口,获取多数据源分别发送的金融

业务数据以及历史调用记录文档,其中,所述历史调用记录文档中包括所述金融业务数据中所有数据的历史调用频率;通过解析,对所述金融业务数据进行来源识别、存储等级识别、属性字段识别、金融业务场景识别、数据值类别识别、历史调用频率识别,获得识别结果;基于所述识别结果和预设的整合策略,对所述金融业务数据进行整合,获得初步整合结果;根据预设的调整策略对所述初步整合结果进行调整,获得目标整合结果,完成对所述金融业务数据的整合。通过所述整合策略,保证了多数据源的金融业务数据的合理存储,通过所述预设的调整策略,保证了目标金融业务数据在被调用时,存在多个对应的可调用分区时,为了避免调用到数据量较大的分区,直接预先提供数据量为最小值时每条金融业务数据分别对应的调用分区信息,方便后续业务调用,合理进行多数据源数据整合。

[0135] 进一步参考图7,作为对上述图2所示方法的实现,本申请提供了数据整合装置的一个实施例,该装置实施例与图2所示的方法实施例相对应,该装置具体可以应用于各种电子设备中。

[0136] 如图7所示,本实施例所述的数据整合装置700包括:数据获取模块701、数据解析识别模块702、初步整合模块703和初步整合结果调整模块704。其中:

[0137] 数据获取模块701,用于根据预设的统一接收接口,获取多数据源分别发送的金融业务数据以及历史调用记录文档,其中,所述历史调用记录文档中包括所述金融业务数据中所有数据的历史调用频率;

[0138] 数据解析识别模块702,用于通过解析,对所述金融业务数据进行来源识别、存储等级识别、属性字段识别、金融业务场景识别、数据值类别识别、历史调用频率识别,获得识别结果;

[0139] 初步整合模块703,用于基于所述识别结果和预设的整合策略,对所述金融业务数据进行整合,获得初步整合结果;

[0140] 初步整合结果调整模块704,用于根据预设的调整策略对所述初步整合结果进行调整,获得目标整合结果,完成对所述金融业务数据的整合。

[0141] 本申请通过根据预设的统一接收接口,获取多数据源分别发送的金融业务数据以及历史调用记录文档,其中,所述历史调用记录文档中包括所述金融业务数据中所有数据的历史调用频率;通过解析,对所述金融业务数据进行来源识别、存储等级识别、属性字段识别、金融业务场景识别、数据值类别识别、历史调用频率识别,获得识别结果;基于所述识别结果和预设的整合策略,对所述金融业务数据进行整合,获得初步整合结果;根据预设的调整策略对所述初步整合结果进行调整,获得目标整合结果,完成对所述金融业务数据的整合。通过所述整合策略,根据金融业务数据的来源、存储等级、属性字段、金融业务场景、数据值类别、历史调用频率进行数据整合,保证了多数据源的金融业务数据的合理存储,通过所述预设的调整策略,保证了目标金融业务数据在被调用时,存在多个对应的可调用分区时,为了避免调用到数据量较大的分区,直接预先提供数据量为最小值时每条金融业务数据分别对应的调用分区信息,方便后续业务调用,合理进行多数据源数据整合。

[0142] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程,是可以通过计算机可读指令来指令相关的硬件来完成,该计算机可读指令可存储于计算机可读存储介质中,该程序在执行时,可包括如上述各方法的实施例的流程。其中,前述的存储介质可为磁碟、光盘、只读存储记忆体(Read-Only Memory, ROM)等非易失性存储介质,或随机

存储记忆体(Random Access Memory, RAM)等。

[0143] 应该理解的是,虽然附图的流程图中的各个步骤按照箭头的指示依次显示,但是这些步骤并不是必然按照箭头指示的顺序依次执行。除非本文中有明确的说明,这些步骤的执行并没有严格的顺序限制,其可以以其他的顺序执行。而且,附图的流程图中的至少一部分步骤可以包括多个子步骤或者多个阶段,这些子步骤或者阶段并不必然是在同一时刻执行完成,而是可以在不同的时刻执行,其执行顺序也不必然是依次进行,而是可以与其他步骤或者其他步骤的子步骤或者阶段的至少一部分轮流或者交替地执行。

[0144] 为解决上述技术问题,本申请实施例还提供计算机设备。具体请参阅图8,图8为本实施例计算机设备基本结构框图。

[0145] 所述计算机设备8包括通过系统总线相互通信连接存储器8a、处理器8b、网络接口8c。需要指出的是,图中仅示出了具有组件8a-8c的计算机设备8,但是应理解的是,并不要求实施所有示出的组件,可以替代的实施更多或者更少的组件。其中,本技术领域技术人员可以理解,这里的计算机设备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/或信息处理的设备,其硬件包括但不限于微处理器、专用集成电路(Application Specific Integrated Circuit, ASIC)、可编程门阵列(Field-Programmable Gate Array, FPGA)、数字处理器(Digital Signal Processor, DSP)、嵌入式设备等。

[0146] 所述计算机设备可以是桌上型计算机、笔记本、掌上电脑及云端服务器等计算设备。所述计算机设备可以与用户通过键盘、鼠标、遥控器、触摸板或声控设备等方式进行人机交互。

[0147] 所述存储器8a至少包括一种类型的可读存储介质,所述可读存储介质包括闪存、硬盘、多媒体卡、卡型存储器(例如,SD或DX存储器等)、随机访问存储器(RAM)、静态随机访问存储器(SRAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、可编程只读存储器(PROM)、磁性存储器、磁盘、光盘等。在一些实施例中,所述存储器8a可以是所述计算机设备8的内部存储单元,例如该计算机设备8的硬盘或内存。在另一些实施例中,所述存储器8a也可以是所述计算机设备8的外部存储设备,例如该计算机设备8上配备的插接式硬盘,智能存储卡(Smart Media Card, SMC),安全数字(Secure Digital, SD)卡,闪存卡(Flash Card)等。当然,所述存储器8a还可以既包括所述计算机设备8的内部存储单元也包括其外部存储设备。本实施例中,所述存储器8a通常用于存储安装于所述计算机设备8的操作系统和各类应用软件,例如一种数据整合方法的计算机可读指令等。此外,所述存储器8a还可以用于暂时地存储已经输出或者将要输出的各类数据。

[0148] 所述处理器8b在一些实施例中可以是中央处理器(Central Processing Unit, CPU)、控制器、微控制器、微处理器、或其他数据整合芯片。该处理器8b通常用于控制所述计算机设备8的总体操作。本实施例中,所述处理器8b用于运行所述存储器8a中存储的计算机可读指令或者处理数据,例如运行所述数据整合方法的计算机可读指令。

[0149] 所述网络接口8c可包括无线网络接口或有线网络接口,该网络接口8c通常用于在所述计算机设备8与其他电子设备之间建立通信连接。

[0150] 本实施例提出的计算机设备,属于金融科技技术领域,应用于金融业务数据多数据源整合场景中。本申请通过根据预设的统一接收接口,获取多数据源分别发送的金融业务数据以及历史调用记录文档,其中,所述历史调用记录文档中包括所述金融业务数据中

所有数据的历史调用频率;通过解析,对所述金融业务数据进行来源识别、存储等级识别、属性字段识别、金融业务场景识别、数据值类别识别、历史调用频率识别,获得识别结果;基于所述识别结果和预设的整合策略,对所述金融业务数据进行整合,获得初步整合结果;根据预设的调整策略对所述初步整合结果进行调整,获得目标整合结果,完成对所述金融业务数据的整合。通过所述整合策略,根据金融业务数据的来源、存储等级、属性字段、金融业务场景、数据值类别、历史调用频率进行数据整合,保证了多数据源的金融业务数据的合理存储,通过所述预设的调整策略,保证了目标金融业务数据在被调用时,存在多个对应的可调用分区时,为了避免调用到数据量较大的分区,直接预先提供数据量为最小值时每条金融业务数据分别对应的调用分区信息,方便后续业务调用,合理进行多数据源数据整合。

[0151] 本申请还提供了另一种实施方式,即提供一种计算机可读存储介质,所述计算机可读存储介质存储有计算机可读指令,所述计算机可读指令可被处理器执行,以使所述处理器执行如上述的数据整合方法的步骤。

[0152] 本实施例提出的计算机可读存储介质,属于金融科技技术领域,应用于金融业务数据多数据源整合场景中。本申请通过根据预设的统一接收接口,获取多数据源分别发送的金融业务数据以及历史调用记录文档,其中,所述历史调用记录文档中包括所述金融业务数据中所有数据的历史调用频率;通过解析,对所述金融业务数据进行来源识别、存储等级识别、属性字段识别、金融业务场景识别、数据值类别识别、历史调用频率识别,获得识别结果;基于所述识别结果和预设的整合策略,对所述金融业务数据进行整合,获得初步整合结果;根据预设的调整策略对所述初步整合结果进行调整,获得目标整合结果,完成对所述金融业务数据的整合。通过所述整合策略,根据金融业务数据的来源、存储等级、属性字段、金融业务场景、数据值类别、历史调用频率进行数据整合,保证了多数据源的金融业务数据的合理存储,通过所述预设的调整策略,保证了目标金融业务数据在被调用时,存在多个对应的可调用分区时,为了避免调用到数据量较大的分区,直接预先提供数据量为最小值时每条金融业务数据分别对应的调用分区信息,方便后续业务调用,合理进行多数据源数据整合。

[0153] 通过以上的实施方式的描述,本领域的技术人员可以清楚地了解到上述实施例方法可借助软件加必需的通用硬件平台的方式来实现,当然也可以通过硬件,但很多情况下前者是更佳的实施方式。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(如ROM/RAM、磁碟、光盘)中,包括若干指令用以使得一台终端设备(可以是手机,计算机,服务器,空调器,或者网络设备等)执行本申请各个实施例所述的方法。

[0154] 显然,以上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,附图中给出了本申请的较佳实施例,但并不限制本申请的专利范围。本申请可以以许多不同的形式来实现,相反地,提供这些实施例的目的是使对本申请的公开内容的理解更加透彻全面。尽管参照前述实施例对本申请进行了详细的说明,对于本领域的技术人员来而言,其依然可以对前述各具体实施方式所记载的技术方案进行修改,或者对其中部分技术特征进行等效替换。凡是利用本申请说明书及附图内容所做的等效结构,直接或间接运用在其他相关的技术领域,均同理在本申请专利保护范围之内。

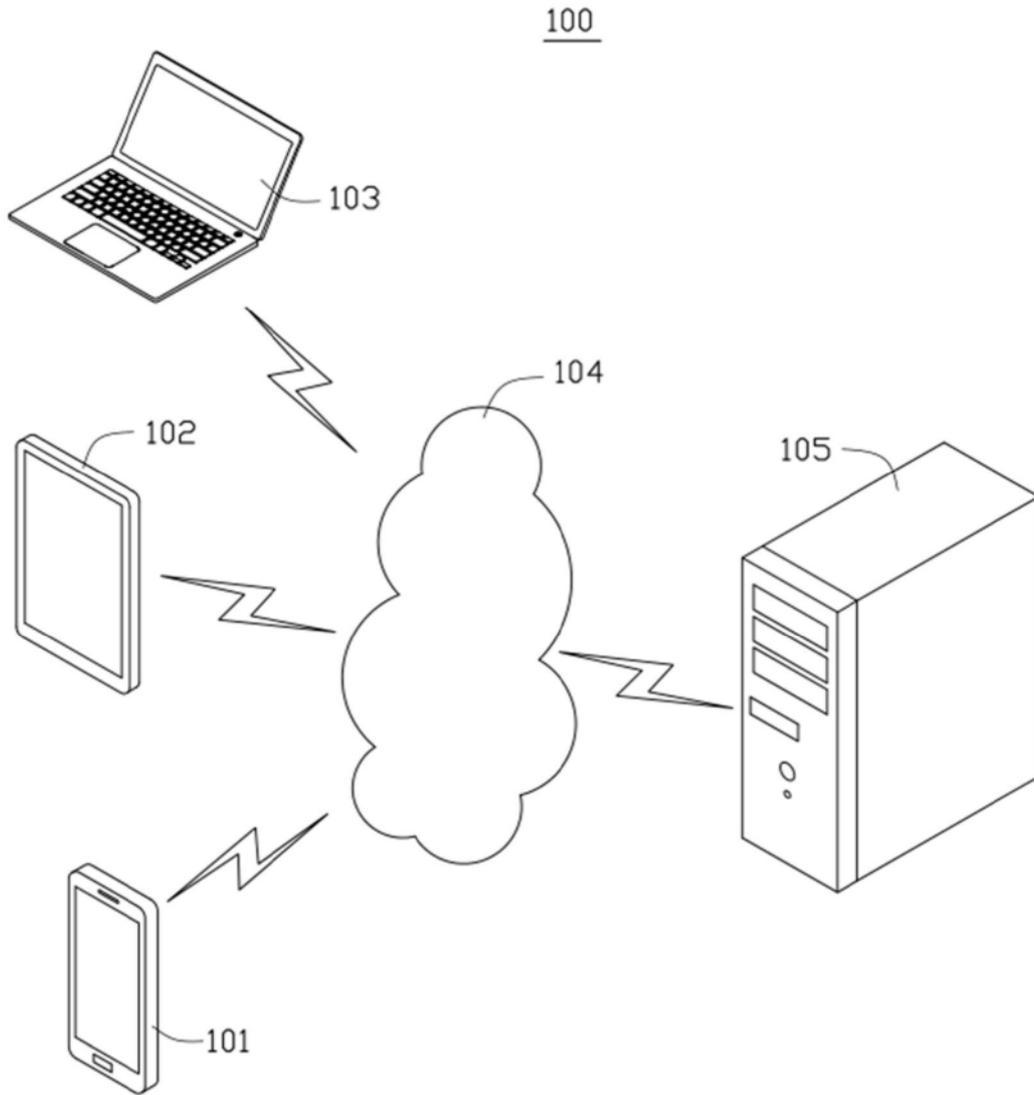


图1

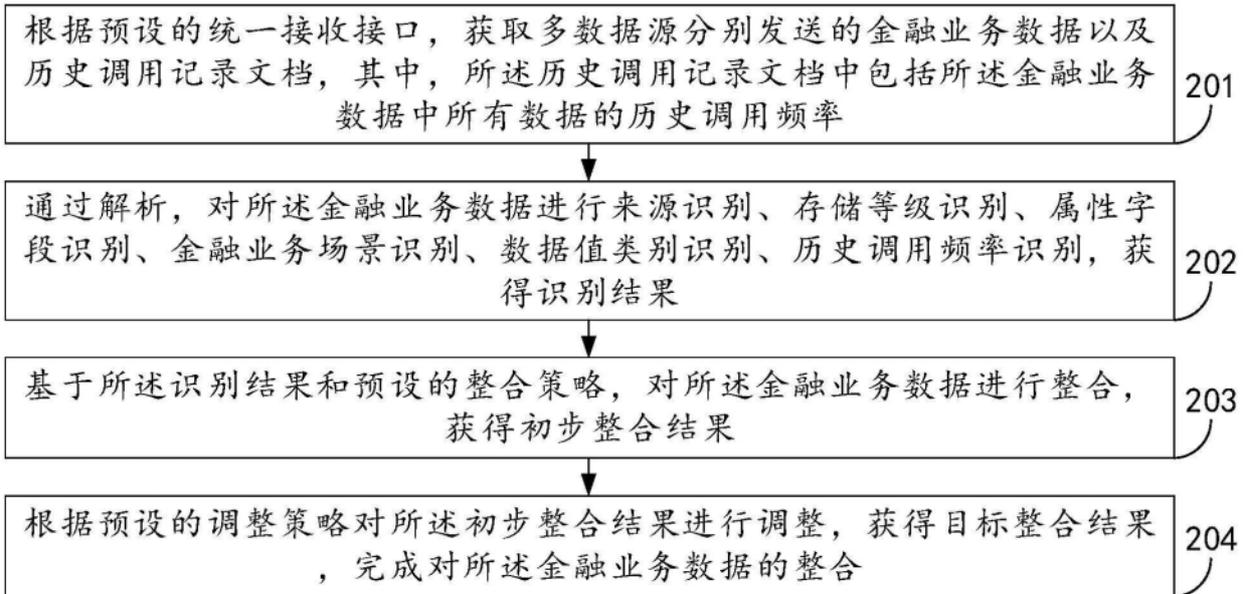


图2

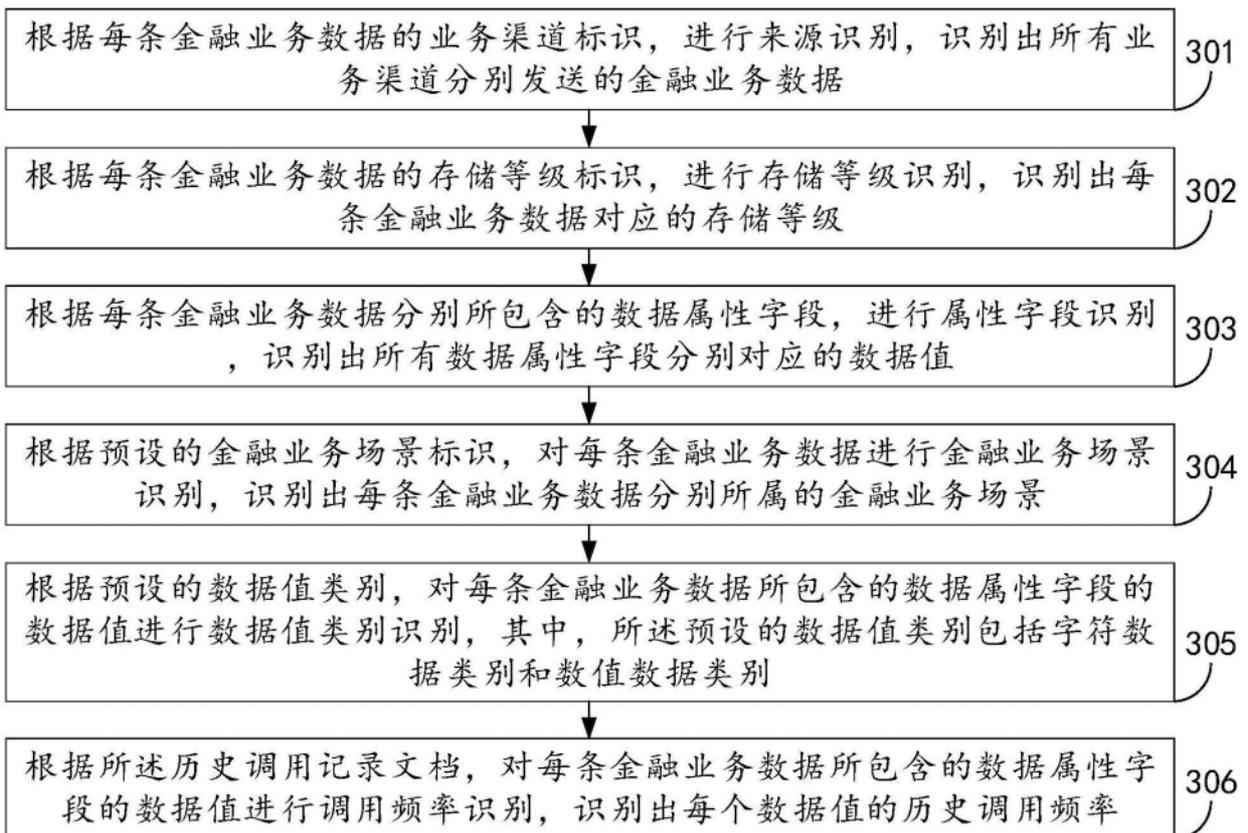


图3

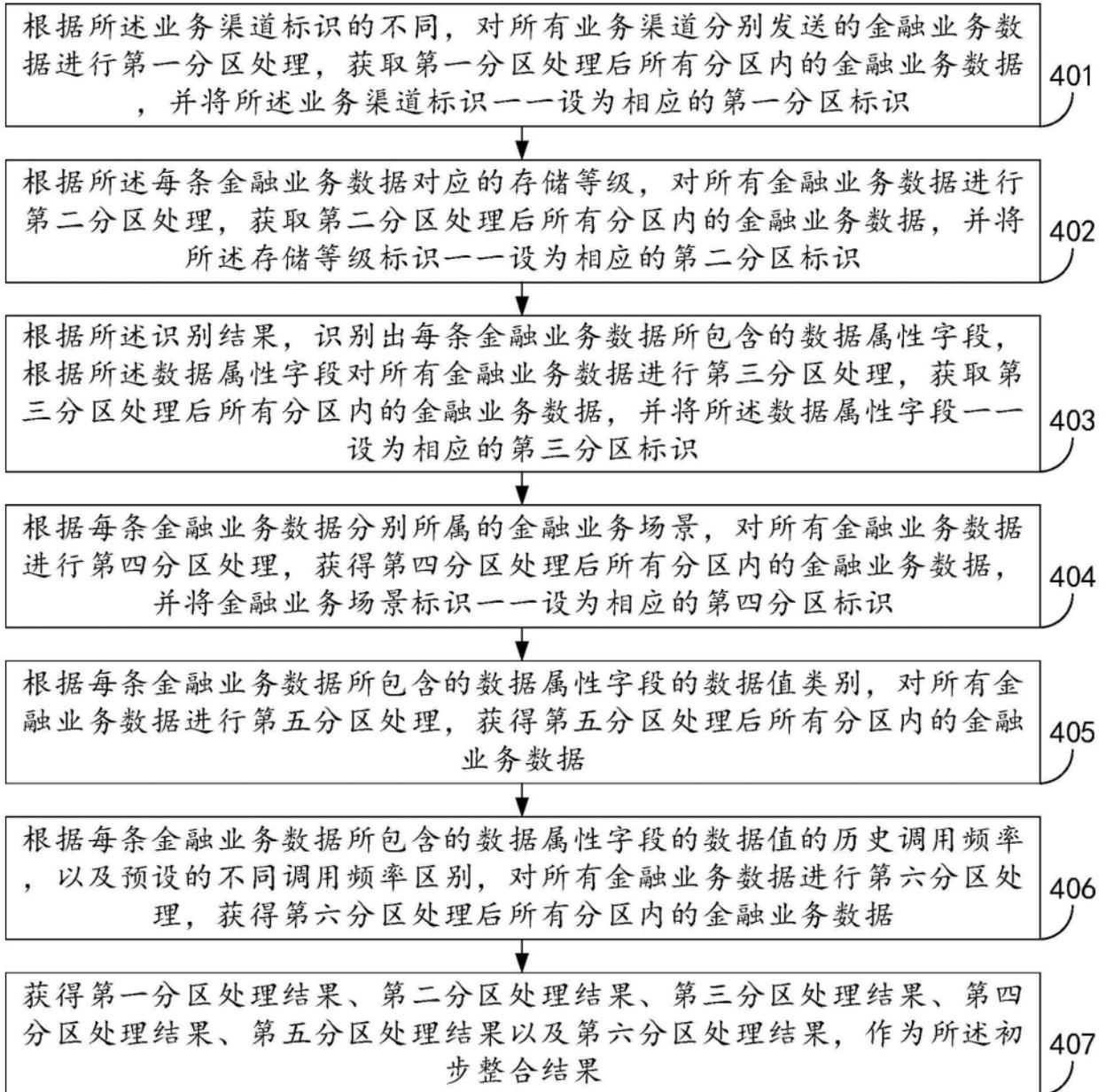


图4

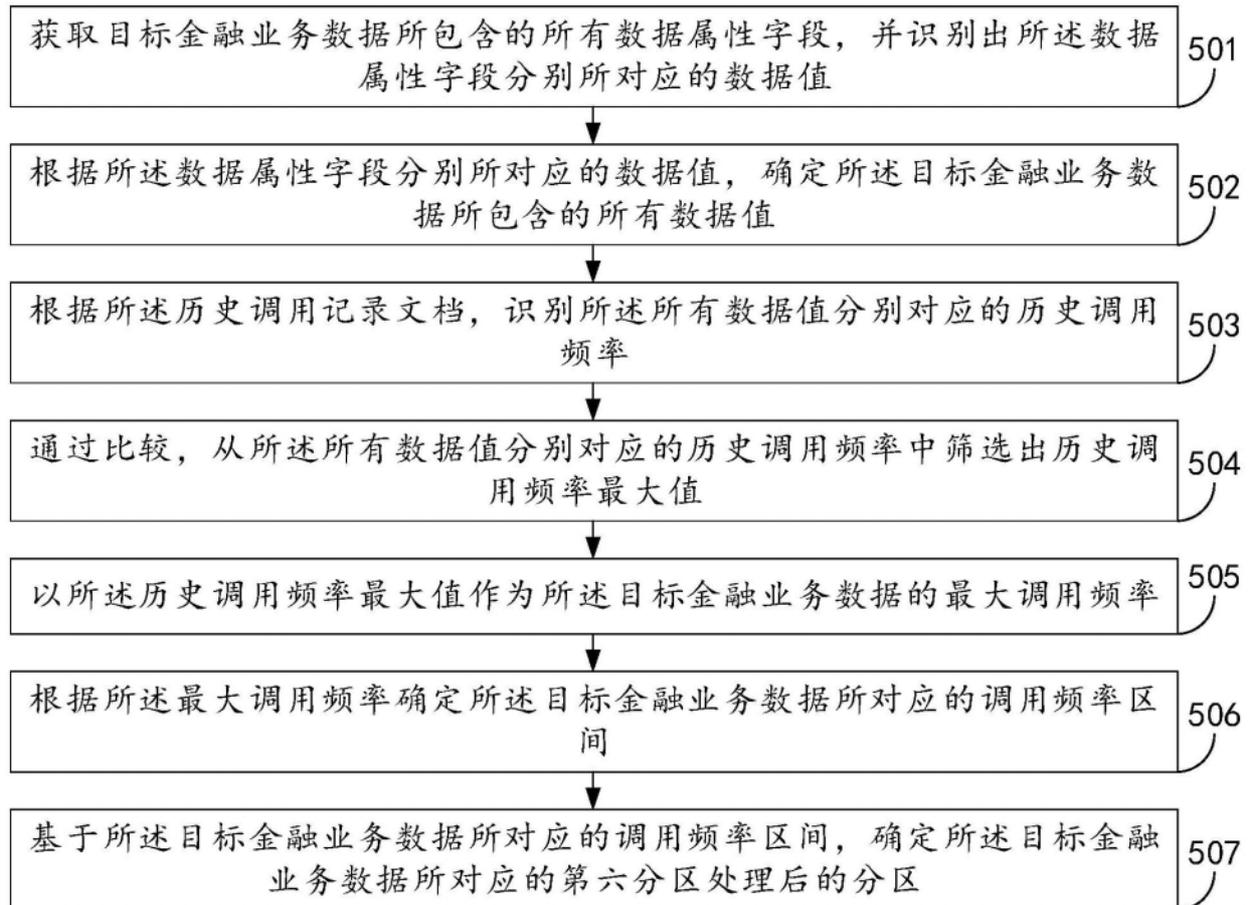


图5

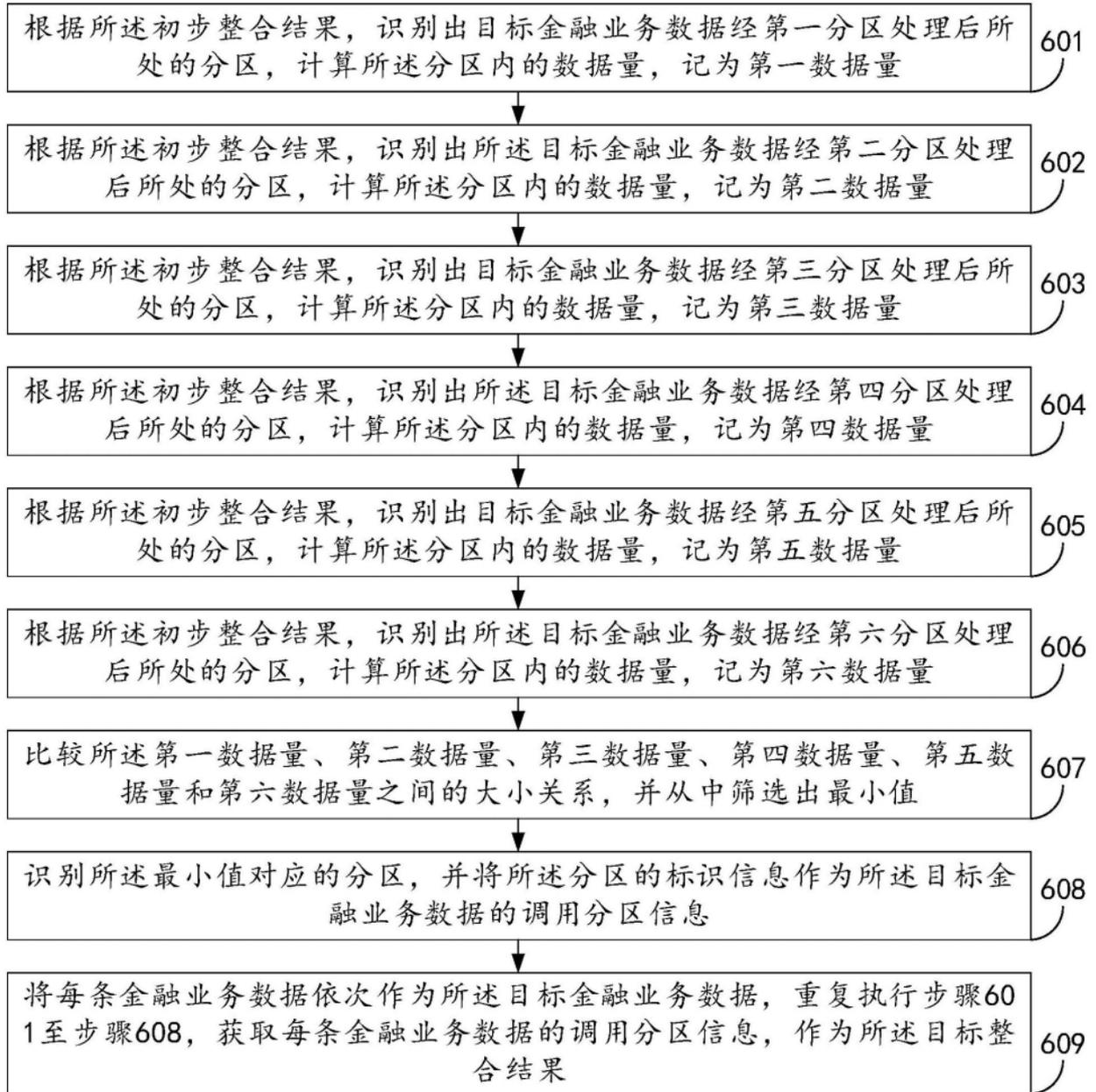


图6

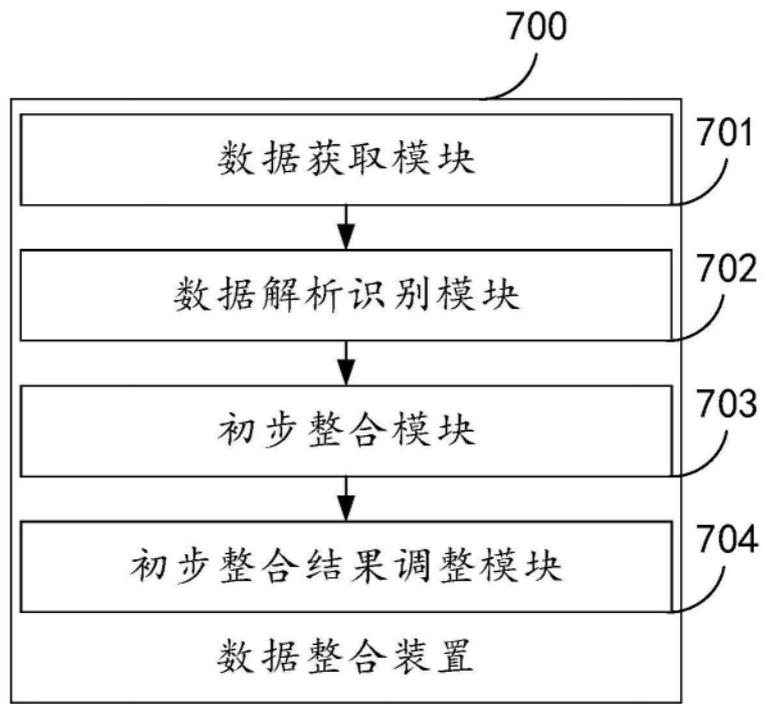


图7

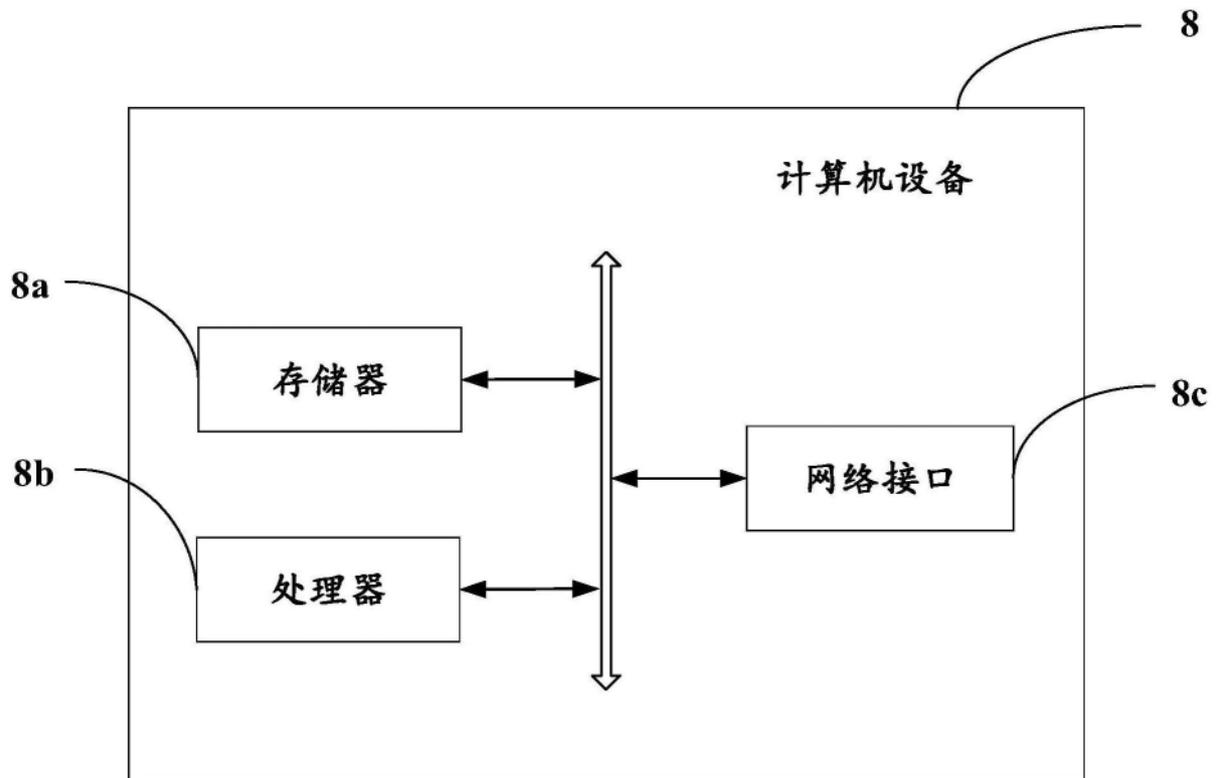


图8