(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 110990390 B
(45)授权公告日 2024.03.08

(21)申请号 201911213587.3

(22)申请日 2019.12.02

(65)同一申请的已公布的文献号
申请公布号 CN 110990390 A

(43)申请公布日 2020.04.10

(73)专利权人 东莞中国科学院云计算产业技术创新与育成中心
地址 523000 广东省东莞市松山湖高新技术产业开发区科汇路1号

(72)发明人 韩超 陈前华 谭思敏

(74)专利代理机构 华进联合专利商标代理有限公司 44224
专利代理师 杨欢

(51)Int.Cl.
G06F 16/215(2019.01)
G06F 16/23(2019.01)
G06F 16/28(2019.01)
G06N 5/025(2023.01)

(56)对比文件
CN 101197876 A,2008.06.11
CN 107729448 A,2018.02.23
US 2012130987 A1,2012.05.24
US 9158827 B1,2015.10.13
US 9910903 B1,2018.03.06
CN 108846076 A,2018.11.20
CN 110489475 A,2019.11.22
CN 108829731 A,2018.11.16
CN 107704590 A,2018.02.16
US 2007239769 A1,2007.10.11
Alexander Albrecht 等.Schema Decryption for Large Extract-Transform-Load Systems. Conceptual Modeling. Proceedings 31st International Conference, ER 2012.2012,第116-25页. (续)

审查员 罗湘

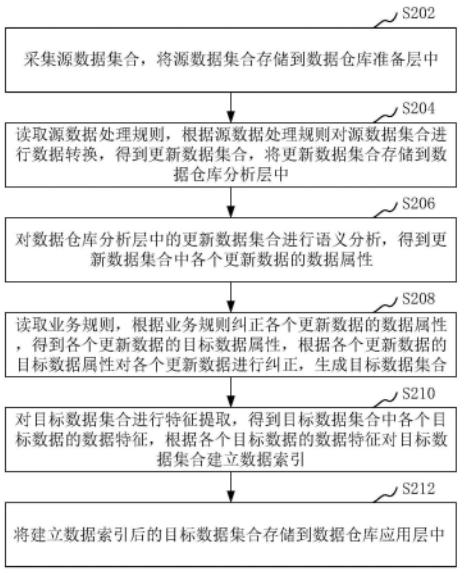
权利要求书2页 说明书13页 附图5页

(54)发明名称

数据协同处理方法、装置、计算机设备和存储介质

(57)摘要

本申请涉及一种数据协同处理方法、装置、计算机设备和存储介质,通过将采集得到的源数据集合存储到数据仓库准备层,根据源数据处理规则对源数据集合进行数据转换后得到更新数据集合,将其存储到数据仓库分析层;进一步地,根据业务规则纠正各个更新数据,得到目标数据集合;对目标数据集合进行特征提取,根据各个目标数据的数据特征对目标数据集合建立数据索引,将建立索引后的目标数据集合存储到数据仓库应用层中。本方法通过数据仓库准备层、数据仓库分析层和数据仓库应用层的三层架构数据仓库体系,对多源异构数据进行处理分析,根据源数据处理规则和业务规则来约束数据,使多源异构数据统一规范,提高数据的利用率。



CN 110990390 B

[转续页]

[接上页]

(56)对比文件

党芳芳.电网企业业务数据质量管控技术的研究.中国优秀硕士学位论文全文数据库信息科技辑.2015,(第第02期期),第I138-577页.

1.一种数据协同处理方法,所述方法包括:

采集源数据集合,将所述源数据集合存储到数据仓库准备层中;其中,所述源数据集合是通过数据采集规则和数据源对应的映射对象采集的,所述数据采集规则是从缓冲区读取的,且所述映射对象通过所述数据源扫描所得;

读取源数据处理规则,根据所述源数据处理规则对所述源数据集合进行数据转换,得到更新数据集合,将所述更新数据集合存储到数据仓库分析层中;

对所述数据仓库分析层中的更新数据集合进行语义分析,得到所述更新数据集合中各个更新数据的数据属性;

读取业务规则,根据所述业务规则纠正所述各个更新数据的数据属性,得到各个更新数据的目标数据属性,根据所述各个更新数据的目标数据属性对所述各个更新数据进行纠正,生成目标数据集合;

对所述目标数据集合进行特征提取,得到所述目标数据集合中各个目标数据的数据特征,根据所述各个目标数据的数据特征对所述目标数据集合建立数据索引;

将建立数据索引后的目标数据集合存储到数据仓库应用层中。

2.根据权利要求1所述的方法,其特征在于,所述采集源数据集合,将所述源数据集合存储到数据仓库准备层中包括:

扫描数据源,获取所述数据源中各个数据的数据信息;

读取缓冲区中的数据采集规则,所述数据采集规则用于确定源数据集合中各个源数据的数据信息;

将所述源数据集合中各个源数据的数据信息和所述数据源中各个数据的数据信息匹配,根据匹配结果在所述数据源中提取源数据集合,将所述源数据集合存储到数据仓库准备层中。

3.根据权利要求2所述的方法,其特征在于,所述方法还包括:

当所述缓冲区中不存在所述数据采集规则时,读取规则配置文件,从所述规则配置文件中获取所述数据采集规则。

4.根据权利要求1所述的方法,其特征在于,所述读取源数据处理规则,根据所述源数据处理规则对所述源数据集合进行数据转换,得到更新数据集合,将所述更新数据集合存储到数据仓库分析层中,包括:

当所述源数据集合中存在异常数据时,根据所述源数据处理规则,对所述异常数据进行数据清洗,得到清洗后的源数据集合;所述清洗后的源数据集合中包括不同数据源中相同类型的数据;

当所述不同数据源中相同类型的数据的编码不同时,根据所述源数据处理规则,对所述不同数据源中相同类型的数据重新编码,得到更新数据集合,所述更新数据集合中不同数据源中相同类型的数据的编码相同;

将所述更新数据集合存储到数据仓库分析层中。

5.根据权利要求1所述的方法,其特征在于,所述业务规则包括数据属性纠正规则和业务决策规则,所述读取业务规则,根据所述业务规则纠正所述各个更新数据的数据属性,得到各个更新数据的目标数据属性,根据所述各个更新数据的目标数据属性对所述各个更新数据进行纠正,生成目标数据集合包括:

当所述各个更新数据的数据属性存在异常时,根据所述数据属性纠正规则,对异常的数据属性进行纠正,得到各个更新数据的目标数据属性;根据所述各个更新数据的目标数据属性对所述各个更新数据进行纠正;

根据所述业务决策规则对纠正后的各个更新数据进行决策判断,当纠正后的各个更新数据满足决策条件时,生成纠正后的各个更新数据对应的决策数据;

根据所述纠正后的各个更新数据和所述决策数据生成目标数据集合。

6.根据权利要求1所述的方法,其特征在于,所述数据仓库应用层包括数据查询接口,在所述将建立数据索引后的目标数据集合存储到数据仓库应用层中之后,还包括:

当所述数据查询接口接收到数据查询请求时,所述数据查询接口根据所述数据查询请求输出对应的目标数据。

7.一种数据协同处理装置,其特征在于,所述装置包括:

源数据集合采集模块,用于采集源数据集合,将所述源数据集合存储到数据仓库准备层中;其中,所述源数据集合是通过数据采集规则和数据源对应的映射对象采集的,所述数据采集规则是从缓冲区读取的,且所述映射对象通过所述数据源扫描所得;

源数据集合转换模块,用于读取源数据处理规则,根据所述源数据处理规则对所述源数据集合进行数据转换,得到更新数据集合,将所述更新数据集合存储到数据仓库分析层中;

数据属性分析模块,用于对所述数据仓库分析层中的更新数据集合进行语义分析,得到所述更新数据集合中各个更新数据的数据属性;

数据属性纠正模块,用于读取业务规则,根据所述业务规则纠正所述各个更新数据的数据属性,得到各个更新数据的目标数据属性,根据所述各个更新数据的目标数据属性对所述各个更新数据进行纠正,生成目标数据集合;

数据索引建立模块,用于对所述目标数据集合进行特征提取,得到所述目标数据集合中各个目标数据的数据特征,根据所述各个目标数据的数据特征对所述目标数据集合建立数据索引;

目标数据集合存储模块,用于将建立数据索引后的目标数据集合存储到数据仓库应用层中。

8.根据权利要求7所述的装置,其特征在于,所述源数据集合采集模块还用于扫描数据源,获取所述数据源中各个数据的数据信息;读取缓存的数据采集规则,所述数据采集规则用于确定源数据集合中各个源数据的数据信息;将所述源数据集合中各个源数据的数据信息和所述数据源中各个数据的数据信息匹配,根据匹配结果在所述数据源中提取源数据集合,将所述源数据集合存储到数据仓库准备层中。

9.一种计算机设备,包括存储器、处理器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时实现权利要求1至6中任一项所述方法的步骤。

10.一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,所述计算机程序被处理器执行时实现权利要求1至6中任一项所述的方法的步骤。

数据协同处理方法、装置、计算机设备和存储介质

技术领域

[0001] 本申请涉及数据处理技术领域,特别是涉及一种数据协同处理方法、装置、计算机设备和存储介质。

背景技术

[0002] 在大数据环境下,由于数据的结构差异大、数据来源广、价值密度较低、更新实时等特点,“数据孤岛”现象日益显现。不同数据源之间的数据差异性大,数据间无法流动,导致信息无法共享,对信息系统建设造成很大的影响。因此,需要提高数据的流动性和利用率。

[0003] 传统方案中,通过RMI远程方法调用来实现数据协同,然而RMI远程方法调用无法跨语言运行,具有一定的局限性,不能很好地实现异构数据间的流动,导致数据利用率较低。

发明内容

[0004] 基于此,有必要针对上述技术问题,提供一种能够提高数据利用率的数据协同处理方法、装置、计算机设备和存储介质。

[0005] 一种数据协同处理方法,所述方法包括:

[0006] 采集源数据集合,将所述源数据集合存储到数据仓库准备层中;

[0007] 读取源数据处理规则,根据所述源数据处理规则对所述源数据集合进行数据转换,得到更新数据集合,将所述更新数据集合存储到数据仓库分析层中;

[0008] 对所述数据仓库分析层中的更新数据集合进行语义分析,得到所述更新数据集合中各个更新数据的数据属性;

[0009] 读取业务规则,根据所述业务规则纠正所述各个更新数据的数据属性,得到各个更新数据的目标数据属性,根据所述各个更新数据的目标数据属性对所述各个更新数据进行纠正,生成目标数据集合;

[0010] 对所述目标数据集合进行特征提取,得到所述目标数据集合中各个目标数据的数据特征,根据所述各个目标数据的数据特征对所述目标数据集合建立数据索引;

[0011] 将建立数据索引后的目标数据集合存储到数据仓库应用层中。

[0012] 在其中一个实施例中,所述采集源数据集合,将所述源数据集合存储到数据仓库准备层中包括:

[0013] 扫描数据源,获取所述数据源中各个数据的数据信息;

[0014] 在读取缓冲区中的数据采集规则,所述数据采集规则用于确定源数据集合中各个源数据的数据信息;

[0015] 将所述源数据集合中各个源数据的数据信息和所述数据源中各个数据的数据信息匹配,根据匹配结果在所述数据源中提取源数据集合,将所述源数据集合存储到数据仓库准备层中。

[0016] 在其中一个实施例中,所述方法还包括:

[0017] 当所述缓冲区中不存在所述数据采集规则时,读取规则配置文件,从所述规则配置文件中获取所述数据采集规则。

[0018] 在其中一个实施例中,所述读取源数据处理规则,根据所述源数据处理规则对所述源数据集合进行数据转换,得到更新数据集合,将所述更新数据集合存储到数据仓库分析层中,包括:

[0019] 当所述源数据集合中存在异常数据时,根据所述源数据处理规则,对所述异常数据进行数据清洗,得到清洗后的源数据集合;所述清洗后的源数据集合中包括不同数据源中相同类型的数据;

[0020] 当所述不同数据源中相同类型的数据的编码不同时,根据所述源数据处理规则,对所述不同数据源中相同类型的数据重新编码,得到更新数据集合,所述更新数据集合中不同数据源中相同类型的数据的编码相同;

[0021] 将所述更新数据集合存储到数据仓库分析层中。

[0022] 在其中一个实施例中,所述业务规则包括数据属性纠正规则和业务决策规则,所述读取业务规则,根据所述业务规则纠正所述各个更新数据的数据属性,得到各个更新数据的目标数据属性,根据所述各个更新数据的目标数据属性对所述各个更新数据进行纠正,生成目标数据集合包括:

[0023] 当所述各个更新数据的数据属性存在异常时,根据所述数据属性纠正规则,对异常的数据属性进行纠正,得到各个更新数据的目标数据属性;根据所述各个更新数据的目标数据属性对所述各个更新数据进行纠正;

[0024] 根据所述业务决策规则对纠正后的各个更新数据进行决策判断,当纠正后的各个更新数据满足决策条件时,生成纠正后的各个更新数据对应的决策数据;

[0025] 根据所述纠正后的各个更新数据和所述决策数据生成目标数据集合。

[0026] 在其中一个实施例中,所述数据仓库应用层包括数据查询接口,在所述将建立数据索引后的目标数据集合存储到数据仓库应用层中之后,还包括:

[0027] 当所述数据查询接口接收到数据查询请求时,所述数据查询接口根据所述数据查询请求输出对应的目标数据。

[0028] 一种数据协同处理装置,所述装置包括:

[0029] 源数据集合采集模块,用于采集源数据集合,将所述源数据集合存储到数据仓库准备层中;

[0030] 源数据集合转换模块,用于读取源数据处理规则,根据所述源数据处理规则对所述源数据集合进行数据转换,得到更新数据集合,将所述更新数据集合存储到数据仓库分析层中;

[0031] 数据属性分析模块,用于对所述数据仓库分析层中的更新数据集合进行语义分析,得到所述更新数据集合中各个更新数据的数据属性;

[0032] 数据属性纠正模块,用于读取业务规则,根据所述业务规则纠正所述各个更新数据的数据属性,得到各个更新数据的目标数据属性,根据所述各个更新数据的目标数据属性对所述各个更新数据进行纠正,生成目标数据集合;

[0033] 数据索引建立模块,用于对所述目标数据集合进行特征提取,得到所述目标数据

集合中各个目标数据的数据特征,根据所述各个目标数据的数据特征对所述目标数据集合建立数据索引;

[0034] 目标数据集合存储模块,用于将建立数据索引后的目标数据集合存储到数据仓库应用层中。

[0035] 一种计算机设备,包括存储器、处理器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时实现以下步骤:

[0036] 采集源数据集合,将所述源数据集合存储到数据仓库准备层中;

[0037] 读取源数据处理规则,根据所述源数据处理规则对所述源数据集合进行数据转换,得到更新数据集合,将所述更新数据集合存储到数据仓库分析层中;

[0038] 对所述数据仓库分析层中的更新数据集合进行语义分析,得到所述更新数据集合中各个更新数据的数据属性;

[0039] 读取业务规则,根据所述业务规则纠正所述各个更新数据的数据属性,得到各个更新数据的目标数据属性,根据所述各个更新数据的目标数据属性对所述各个更新数据进行纠正,生成目标数据集合;

[0040] 对所述目标数据集合进行特征提取,得到所述目标数据集合中各个目标数据的数据特征,根据所述各个目标数据的数据特征对所述目标数据集合建立数据索引;

[0041] 将建立数据索引后的目标数据集合存储到数据仓库应用层中。

[0042] 一种计算机可读存储介质,其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以下步骤:

[0043] 采集源数据集合,将所述源数据集合存储到数据仓库准备层中;

[0044] 读取源数据处理规则,根据所述源数据处理规则对所述源数据集合进行数据转换,得到更新数据集合,将所述更新数据集合存储到数据仓库分析层中;

[0045] 对所述数据仓库分析层中的更新数据集合进行语义分析,得到所述更新数据集合中各个更新数据的数据属性;

[0046] 读取业务规则,根据所述业务规则纠正所述各个更新数据的数据属性,得到各个更新数据的目标数据属性,根据所述各个更新数据的目标数据属性对所述各个更新数据进行纠正,生成目标数据集合;

[0047] 对所述目标数据集合进行特征提取,得到所述目标数据集合中各个目标数据的数据特征,根据所述各个目标数据的数据特征对所述目标数据集合建立数据索引;

[0048] 将建立数据索引后的目标数据集合存储到数据仓库应用层中。

[0049] 上述数据协同处理方法、装置、计算机设备和存储介质,通过将采集得到的多源异构数据存储到数据仓库准备层,读取源数据处理规则,根据源数据处理规则对数据仓库准备层中的源数据集合进行数据转换后得到更新数据集合,将其存储到数据仓库分析层;进一步地,读取业务规则,根据业务规则纠正数据仓库分析层中的各个更新数据的数据属性,从而纠正各个更新数据生成目标数据集合;对目标数据集合进行特征提取,得到目标数据集合中各个目标数据的数据特征,根据各个目标数据的数据特征对目标数据集合建立数据索引,将建立索引后的目标数据集合存储到数据仓库应用层中。与传统方法不同,本方法通过数据仓库准备层、数据仓库分析层和数据仓库应用层的三层架构数据仓库体系,对多源异构数据进行处理分析,根据源数据处理规则和业务规则来约束数据,使多源异构数据统

一规范,提高数据的利用率。

附图说明

[0050] 图1a为一个实施例中数据协同处理方法的应用场景图;

[0051] 图1b为一个实施例中数据仓库服务器的架构示意图;

[0052] 图2为一个实施例中数据协同处理方法的流程示意图;

[0053] 图3为一个实施例中源数据集合采集方法的流程示意图;

[0054] 图4为另一个实施例中数据协同处理方法的流程示意图;

[0055] 图5为一个实施例中数据协同处理装置的结构框图;

[0056] 图6为一个实施例中计算机设备的内部结构图。

具体实施方式

[0057] 为了使本申请的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本申请进行进一步详细说明。应当理解,此处描述的具体实施例仅仅用以解释本申请,并不用于限定本申请。

[0058] 本申请提供的数据协同处理方法,可以应用于如图1a所示的应用环境中。其中,数据库服务器102通过网络与数据仓库服务器104通过网络进行通信,数据仓库服务器104通过网络与终端106通过网络进行通信。数据仓库服务器104的架构如图1b所示,包括ODS层(Operational Data Store,操作性数据存储层),也叫数据仓库准备层;DW层(Data Warehouse,数据仓库),也叫数据仓库分析层;DM层(Data Mart,数据集市),也叫数据仓库应用层。数据库服务器102存储了大量的格式不一的多源异构数据。数据仓库服务器104通过网络连接,对数据库服务器102中存储的多源异构数据进行采集,得到源数据集合,并将其存储到数据仓库准备层中。数据仓库服务器104读取源数据处理规则,根据源数据处理规则对存储在数据仓库准备层中的源数据集合进行数据转换,得到更新数据集合,将更新数据集合存储到数据仓库分析层中。进一步地,数据仓库服务器104读取业务规则,根据业务规则对存储在数据仓库分析层中的更新数据集合进行纠正,得到统一规范的目标数据集合。数据仓库服务器104对目标数据集合进行特征提取,得到目标数据集合中各个目标数据的数据特征。数据仓库服务器104根据各个目标数据的数据特征对目标数据集合建立数据索引,将建立索引后的目标数据集合存储到数据仓库应用层中。终端106可以发起请求,访问数据仓库服务器104的数据仓库应用层并从中获取数据。可以理解的是,不止终端可以访问数据仓库应用层获取数据,服务器也可以。其中,终端106可以但不限于是各种个人计算机、笔记本电脑、智能手机、平板电脑和便携式可穿戴设备,数据库服务器102和数据仓库服务器104可以用独立的服务器或者是多个服务器组成的服务器集群来实现。

[0059] 在一个实施例中,如图2所示,提供了一种数据协同处理方法,以该方法应用于图1中的数据仓库服务器为例进行说明,包括以下步骤:

[0060] 步骤202,采集源数据集合,将源数据集合存储到数据仓库准备层中。

[0061] 其中,源数据集合是数据库服务器中存储的部分数据,可以是数据库服务器中多个数据源的、结构不同的数据。

[0062] 具体地,数据仓库服务器通常采用ETL(Extract-Transform-Load,数据仓库技术)

来处理数据库服务器中的数据。ETL中包括抽取(extract)、转换(transform)和加载(load)。其中,抽取就是从数据库服务器的数据源中采集源数据集合。而数据库服务器中存在着各种各样的数据,所以可以根据数据仓库所需要的数据信息跟数据库服务器中的数据信息进行匹配,再从中抽取数据信息匹配成功的数据,完成源数据集合的采集。

[0063] 进一步地,由于源数据集合来自于不同的数据源,所以采集到的源数据的数据结构不同。数据仓库服务器通过数据仓库技术从数据库服务器中采集得到源数据集合,再将源数据集合存储到数据仓库准备层中。作为数据库到数据仓库的一个过渡,数据仓库准备层中的各个源数据的数据结构保持不变,与原来数据源中的对应数据的数据结构相同。

[0064] 步骤204,读取源数据处理规则,根据源数据处理规则对源数据集合进行数据转换,得到更新数据集合,将更新数据集合存储到数据仓库分析层中。

[0065] 其中,源数据处理规则包括源数据集合的数据转换规则。数据转换规则包括数据结构转换规则、数据粒度转换规则和数据指标转换规则等。

[0066] 具体地,由于步骤202采集到数据仓库准备层中的各个源数据的数据结构与原来数据源中对应的数据结构相同,即数据仓库准备层存储的各个源数据的数据结构互不相同。因此,需要对数据仓库准备层中的源数据集合进行数据转换,使源数据集合的数据结构统一。数据仓库服务器读取源数据处理规则,通过数据结构转换规则将源数据集合中各个源数据的数据结构统一。

[0067] 进一步地,由于数据仓库分析层是用于对数据进行分析的,一般情况下会根据时间粒度来对数据进行统计。因此,数据仓库服务器通过数据粒度转换规则,根据各个源数据的入库时间,对各个源数据进行不同时间粒度上的聚合,实现统计分析。而不同的数据源中数据指标可能不同,对于数据指标不同的数据不能直接累加统计,所以,在进行数据粒度转换时还需要进行数据指标转换,才能实现统计分析。比如,对于相同的试卷,学生A所在的学校将满分设为10分,A在考试中得了10分;学生B所在的学校将满分设为100分,A在考试中得了80分;两个学校对试卷的设置满分指标不同,不能直接得出学生A成绩比学生B差的结论,应该先转换满分指标,可以将学生A所在的学校的满分指标设为100分,那么相对的学生A的考试得分为100,从而得出学生A的成绩比学生B好的结论。

[0068] 步骤206,对数据仓库分析层中的更新数据集合进行语义分析,得到更新数据集合中各个更新数据的数据属性。

[0069] 其中,数据属性包括数据的类型、默认值、取值范围和内容等。

[0070] 具体地,根据更新数据集合中各个更新数据的语义,对各个更新数据进行分析,提取得到各个更新数据的数据属性内容。例如,数据仓库服务器对某个html网页进行数据内容采集,而html网页中包括<html>、<body>和等标签,采集得到的源数据可能是<html><body>你好</body></html>,此时,需要对该源数据进行语义分析,去掉相关的标签,得到该数据的数据属性内容,即得到“内容:你好、类型:字符串、长度:4字节、默认值:空”。

[0071] 步骤208,读取业务规则,根据业务规则纠正各个更新数据的数据属性,得到各个更新数据的目标数据属性,根据各个更新数据的目标数据属性对各个更新数据进行纠正,生成目标数据集合。

[0072] 其中,业务规则包括数据属性纠正规则,可以对错误的数据属性进行纠正,从而使

数据变得更加规范。

[0073] 具体地,数据仓库服务器读取业务规则,对各个更新数据的数据属性进行检测判断,当某个更新数据的数据属性错误时,将其纠正,从而实现更新数据的纠正,生成目标数据集合。例如学生成绩的数据mark,数据属性纠正规则规定其数据类型是double,数值范围是[0,100];当数据mark有一个值为110时,即超过数据属性纠正规则定义的[0,100]的数据范围,那么根据数据属性纠正规则,将其纠正为100。

[0074] 进一步地,业务规则还可以对目标数据集合进行分类,得到多个不同类型的目标数据表。

[0075] 步骤210,对目标数据集合进行特征提取,得到目标数据集合中各个目标数据的数据特征,根据各个目标数据的数据特征对目标数据集合建立数据索引。

[0076] 其中,索引是对数据表中一列或多列的值进行排序的一种结构,使用索引可快速访问数据表中的特定信息。

[0077] 具体地,目标数据集合包括不同类型的目标数据表。数据仓库分析层对各个目标数据表进行特征提取,得到各个目标数据表的数据特征。例如,有一个人员名单的目标数据表,包括姓名、性别、年龄、手机号和身份证号,在建立数据索引时,需要考虑数值的唯一性,因此,可以根据手机号或者身份证号建立索引,即建立主键索引。通过手机号或者身份证号都可以在该人员名单中找到唯一对应的人员。

[0078] 步骤212,将建立数据索引后的目标数据集合存储到数据仓库应用层中。

[0079] 具体地,数据仓库应用层允许多个终端或服务器访问和调用,将建立数据索引后的目标数据集合存储到数据仓库应用层中,可以将规范化处理后的数据投入使用,提高数据的利用率。

[0080] 上述数据协同处理方法中,通过将采集得到的多源异构数据存储到数据仓库准备层,读取源数据处理规则,根据源数据处理规则对数据仓库准备层中的源数据集合进行数据转换后得到更新数据集合,将其存储到数据仓库分析层;进一步地,读取业务规则,根据业务规则纠正数据仓库分析层中的各个更新数据的数据属性,从而纠正各个更新数据生成目标数据集合;对目标数据集合进行特征提取,得到目标数据集合中各个目标数据的数据特征,根据各个目标数据的数据特征对目标数据集合建立数据索引,将建立索引后的目标数据集合存储到数据仓库应用层中。本方法通过数据仓库准备层、数据仓库分析层和数据仓库应用层的三层架构数据仓库体系,对多源异构数据进行处理分析,根据源数据处理规则和业务规则来约束数据,使多源异构数据统一规范,提高数据的利用率。

[0081] 在一个实施例中,步骤202包括:扫描数据源,获取数据源中各个数据的数据信息;读取缓冲区中的数据采集规则,该数据采集规则用于确定源数据集合中各个源数据的数据信息;将源数据集合中各个源数据的数据信息和数据源中各个数据的数据信息匹配,根据匹配结果在数据源中提取源数据集合,将源数据集合存储到数据仓库准备层中。

[0082] 其中,数据仓库服务器具有缓冲区,用于缓存数据,例如数据采集规则。

[0083] 具体地,当数据量较大或者有新增数据时,数据仓库服务器需要反复读取数据库服务器中的数据源来获取源数据集合,导致数据采集效率低。因此,可以先扫描数据源,获取数据源中各个数据的数据信息。再根据各个数据的数据信息生成映射对象并对其赋值,将其初始化,由此建立了数据与映射对象间的对应关系。

[0084] 进一步地,数据仓库服务器中的缓冲区存储有数据采集规则,可以快速地读取到数据采集规则。数据采集规则约束了数据源中各个数据的映射对象,只有满足数据采集规则的映射对象,即数据信息匹配,数据仓库服务器才会从数据源中采集对应的数据。当某一类数据信息中有新增数据时,数据仓库服务器也可以感应到,并对满足数据采集规则的映射对象对应的数据进行采集。

[0085] 在本实施例中,通过扫描数据源,获取到数据源对应的映射对象,读取缓冲区中的数据采集规则,通过数据采集规则和映射对象来采集源数据集合,可以提高数据采集的效率。

[0086] 在一个实施例中,当缓冲区中不存在数据采集规则时,读取规则配置文件,从规则配置文件中获取数据采集规则。

[0087] 其中,规则配置文件包括多种类型的数据处理规则,例如数据采集规则。

[0088] 具体地,源数据集合的采集步骤如图3所示,在此不再赘述。

[0089] 在本实施例中,当缓冲区中不存在数据采集规则时,从规则配置文件中获取数据采集规则,保证数据采集的合理性。

[0090] 在一个实施例中,步骤204包括:当源数据集合中存在异常数据时,根据源数据处理规则,对异常数据进行数据清洗,得到清洗后的源数据集合;清洗后的源数据集合中包括不同数据源中相同类型的数据;当不同数据源中相同类型的数据的编码不同时,根据源数据处理规则,对不同数据源中相同类型的数据重新编码,得到更新数据集合,该更新数据集合中不同数据源中相同类型的数据的编码相同;将更新数据集合存储到数据仓库分析层中。

[0091] 其中,源数据处理规则对数据仓库准备层中的源数据集合进行初步的数据处理,数据处理后得到的更新数据集合相比于源数据集合更加规范。异常数据包括部分信息缺失的数据和重复的数据等类型。源数据处理规则对异常数据进行数据清洗可以是直接删除数据或者对数据进行补全、修正。

[0092] 具体地,当源数据集合中存在部分信息缺失的数据,可以补全该数据的数据信息,使其变得完整。例如,有一个学生成绩表,表中包括各个学生的学号、姓名、学科和成绩,即一个完整的数据应该是“学号-姓名-学科-成绩”,而某个学生对应的数据只有“学号-姓名-成绩”,那么可以根据其他数据推算出该数据中的学科信息,将学科信息按规定的位置填入,即可使该数据变得完整。当源数据集合中存在重复的数据时,可以将重复的数据进行删除,只保留其中的一个。

[0093] 进一步地,因为清洗后的源数据集合中包括不同数据源中相同类型的数据,因此需要将统一数据标准,才能提高数据的利用率。当不同数据源中相同类型的数据的编码不同时,根据源数据处理规则,对不同数据源中相同类型的数据重新编码,得到更新数据集合,该更新数据集合中不同数据源中相同类型的数据的编码相同;将更新数据集合存储到数据仓库分析层中。例如,同一个供应商在结算系统的编码是XX0001,而在CRM(Customer Relationship Management,客户关系管理系统)中编码是YY0001,可以将其统一转换为XX0001,这样才能把同一个供应商在结算系统和CRM中的数据整合到一起并使用这些数据。

[0094] 在本实施例中,通过源数据处理规则对源数据集合进行数据清洗和数据转换,使得源数据集合变得更加规范,提高数据的利用率。

[0095] 在一个实施例中,业务规则包括数据属性纠正规则和业务决策规则,步骤208包括:当各个更新数据的数据属性存在异常时,根据数据属性纠正规则,对异常的数据属性进行纠正,得到各个更新数据的目标数据属性;根据各个更新数据的目标数据属性对各个更新数据进行纠正;根据业务决策规则对纠正后的各个更新数据进行决策判断,当纠正后的各个更新数据满足决策条件时,生成纠正后的各个更新数据对应的决策数据;根据纠正后的各个更新数据和决策数据生成目标数据集合。

[0096] 其中,数据属性纠正规则用于对错误的数据属性进行纠正;业务决策规则用于对更新数据进行决策判断,生成对应的决策数据。

[0097] 具体地,当检测到更新数据的数据属性存在异常时,例如数值与规定的数值类型或数值范围不符时,根据数据属性纠正规则,将异常的数据属性进行纠正,使得纠正后的更新数据变得规范。进一步地,根据业务决策规则对纠正后的各个更新数据进行决策判断,当纠正后的各个更新数据满足决策条件时,生成纠正后的各个更新数据对应的决策数据,根据纠正后的各个更新数据和决策数据生成目标数据集合。例如,某个外卖应用规定了:当一笔订单满30元时减5元,那么某订单原来应该付31元,现在只要付26元。

[0098] 在本实施例中,通过业务规则纠正更新数据,使更新数据更加规范;再对纠正后的更新数据进行决策判断,生成对应的决策数据,进而生成目标数据集合,使得目标数据集合中更加符合实际应用,提高数据的灵活性和合理性。

[0099] 在一个实施例中,数据仓库应用层包括数据查询接口,在步骤212之后,方法还包括:当数据查询接口接收到数据查询请求时,数据查询接口根据数据查询请求输出对应的目标数据。

[0100] 具体地,数据仓库应用层包括数据查询接口。数据查询接口可以用于接收数据查询请求。数据查询请求可以是数据对应的主键。数据查询接口在接收到数据查询请求之后,根据数据对应的主键,利用索引,在目标数据集合中快速查找到对应的数据,将其输出。

[0101] 在一个实施例中,数据仓库应用层还包括缓冲区,数据查询率高的数据存储到缓冲区中,提高数据查询和输出的效率。

[0102] 在本实施例中,通过数据查询接口可以接收数据查询请求,输出对应的目标数据,提高数据协同处理的合理性。

[0103] 在另一个实施例中,如图4所示,提供了一种数据协同处理方法,以该方法应用于图1中的数据仓库服务器为例进行说明,包括以下步骤:

[0104] 步骤402,扫描数据源,获取数据源中各个数据的数据信息;

[0105] 步骤404,读取缓冲区中的数据采集规则;当缓冲区中不存在数据采集规则时,读取规则配置文件,从规则配置文件中获取数据采集规则;该数据采集规则用于确定源数据集合中各个源数据的数据信息;

[0106] 步骤406,将源数据集合中各个源数据的数据信息和数据源中各个数据的数据信息匹配,根据匹配结果在数据源中提取源数据集合,将源数据集合存储到数据仓库准备层中;

[0107] 步骤408,读取源数据集合处理规则,当源数据集合中存在异常数据时,根据源数据处理规则,对异常数据进行数据清洗,得到清洗后的源数据集合;清洗后的源数据集合中包括不同数据源中相同类型的数据;

[0108] 步骤410,当不同数据源中相同类型的数据的编码不同时,根据源数据处理规则,对不同数据源中相同类型的数据重新编码,得到更新数据集合,将更新数据集合存储到数据仓库分析层中;该更新数据集合中不同数据源中相同类型的数据的编码相同;

[0109] 步骤412,对数据仓库分析层中的更新数据集合进行语义分析,得到更新数据集合中各个更新数据的数据属性;

[0110] 步骤414,读取业务规则,业务规则包括数据属性纠正规则和业务决策规则;

[0111] 步骤416,当各个更新数据的数据属性存在异常时,根据数据属性纠正规则,对异常的数据属性进行纠正,得到各个更新数据的目标数据属性;根据各个更新数据的目标数据属性对各个更新数据进行纠正;

[0112] 步骤418,根据业务决策规则对纠正后的各个更新数据进行决策判断,当纠正后的各个更新数据满足决策条件时,生成纠正后的各个更新数据对应的决策数据;根据纠正后的各个更新数据和决策数据生成目标数据集合;

[0113] 步骤420,对目标数据集合进行特征提取,得到目标数据集合中各个目标数据的数据特征,根据各个目标数据的数据特征对目标数据集合建立数据索引;

[0114] 步骤422,将建立数据索引后的目标数据集合存储到数据仓库应用层中,该数据仓库应用层包括数据查询接口;当数据查询接口接收到数据查询请求时,数据查询接口根据数据查询请求输出对应的目标数据。

[0115] 在本实施例中,通过扫描数据源,获取数据源中各个数据的数据信息,通过数据采集规则得到与源数据集合匹配的数据信息,从数据源中采集匹配的数据信息对应的源数据,将各个源数据存储到数据仓库准备层中;通过源数据集合处理规则对各个源数据进行初步的数据处理,得到更加规范的更新数据集合,将更新数据集合存储到数据仓库分析层中;对更新数据集合进行语义分析,得到各个更新数据的数据属性;再通过业务规则的约束,对更新数据进一步规范处理并生成对应的决策数据,由规范的更新数据和对应的决策数据生成目标数据集合。进一步地,提取目标数据集合中各个目标数据的数据特征,根据各个目标数据的数据特征对目标数据集合建立数据索引,将建立数据索引后的目标数据集合存储到数据仓库应用层,使数据查询更加便捷,提高数据查询的效率,提高数据的利用率。

[0116] 应该理解的是,虽然图2-4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,这些步骤可以以其它的顺序执行。而且,图2-4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,这些子步骤或者阶段的执行顺序也不必然是依次进行,而是可以与其它步骤或者其它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。

[0117] 在一个实施例中,如图5所示,提供了一种数据协同处理装置500,包括:源数据集合采集模块501、源数据集合转换模块502、数据属性分析模块503、数据属性纠正模块504、数据索引建立模块505和目标数据集合存储模块506,其中:

[0118] 源数据集合采集模块501,用于采集源数据集合,将源数据集合存储到数据仓库准备层中;

[0119] 源数据集合转换模块502,用于读取源数据处理规则,根据源数据处理规则对源数

据集合进行数据转换,得到更新数据集合,将更新数据集合存储到数据仓库分析层中;

[0120] 数据属性分析模块503,用于对数据仓库分析层中的更新数据集合进行语义分析,得到更新数据集合中各个更新数据的数据属性;

[0121] 数据属性纠正模块504,用于读取业务规则,根据业务规则纠正各个更新数据的数据属性,得到各个更新数据的目标数据属性,根据各个更新数据的目标数据属性对各个更新数据进行纠正,生成目标数据集合;

[0122] 数据索引建立模块505,用于对目标数据集合进行特征提取,得到目标数据集合中各个目标数据的数据特征,根据各个目标数据的数据特征对目标数据集合建立数据索引;

[0123] 目标数据集合存储模块506,用于将建立数据索引后的目标数据集合存储到数据仓库应用层中。

[0124] 在一个实施例中,源数据集合采集模块501还用于扫描数据源,获取数据源中各个数据的数据信息;读取缓冲区中的数据采集规则,该数据采集规则用于确定源数据集合中各个源数据的数据信息;将源数据集合中各个源数据的数据信息和数据源中各个数据的数据信息匹配,根据匹配结果在数据源中提取源数据集合,将源数据集合存储到数据仓库准备层中。

[0125] 在一个实施例中,源数据集合采集模块501还用于当缓冲区中不存在数据采集规则时,读取规则配置文件,从规则配置文件中获取数据采集规则。

[0126] 在一个实施例中,源数据集合转换模块502还用于当源数据集合中存在异常数据时,根据源数据处理规则,对异常数据进行数据清洗,得到清洗后的源数据集合;清洗后的源数据集合中包括不同数据源中相同类型的数据;当不同数据源中相同类型的数据的编码不同时,根据源数据处理规则,对不同数据源中相同类型的数据重新编码,得到更新数据集合,该更新数据集合中不同数据源中相同类型的数据的编码相同;将更新数据集合存储到数据仓库分析层中。

[0127] 在一个实施例中,业务规则包括数据属性纠正规则和业务决策规则,数据属性纠正模块504还用于当各个更新数据的数据属性存在异常时,根据数据属性纠正规则,对异常的数据属性进行纠正,得到各个更新数据的目标数据属性;根据各个更新数据的目标数据属性对各个更新数据进行纠正;根据业务决策规则对纠正后的各个更新数据进行决策判断,当纠正后的各个更新数据满足决策条件时,生成纠正后的各个更新数据对应的决策数据;根据纠正后的各个更新数据和决策数据生成目标数据集合。

[0128] 在一个实施例中,数据仓库应用层包括数据查询接口,数据协同处理装置500还包括目标数据输出模块507,用于当数据查询接口接收到数据查询请求时,数据查询接口根据数据查询请求输出对应的目标数据。

[0129] 关于数据协同处理装置的具体限定可以参见上文中对于数据协同处理方法的限定,在此不再赘述。上述数据协同处理装置中的各个模块可全部或部分通过软件、硬件及其组合来实现。上述各模块可以硬件形式内嵌于或独立于计算机设备中的处理器中,也可以以软件形式存储于计算机设备中的存储器中,以便于处理器调用执行以上各个模块对应的操作。

[0130] 在一个实施例中,提供了一种计算机设备,该计算机设备可以是服务器,其内部结构图可以如图6所示。该计算机设备包括通过系统总线连接的处理器、存储器、网络接口和

数据库。其中,该计算机设备的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统、计算机程序和数据库。该内存储器为非易失性存储介质中的操作系统和计算机程序的运行提供环境。该计算机设备的数据库用于存储多源异构数据。该计算机设备的网络接口用于与外部的终端通过网络连接通信。该计算机程序被处理器执行时以实现一种数据协同处理方法。

[0131] 本领域技术人员可以理解,图6中示出的结构,仅仅是与本申请方案相关的部分结构的框图,并不构成对本申请方案所应用于其上的计算机设备的限定,具体的计算机设备可以包括比图中所示更多或更少的部件,或者组合某些部件,或者具有不同的部件布置。

[0132] 在一个实施例中,提供了一种计算机设备,包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序,处理器执行计算机程序时实现以下步骤:采集源数据集合,将源数据集合存储到数据仓库准备层中;读取源数据处理规则,根据源数据处理规则对源数据集合进行数据转换,得到更新数据集合,将更新数据集合存储到数据仓库分析层中;对数据仓库分析层中的更新数据集合进行语义分析,得到更新数据集合中各个更新数据的数据属性;读取业务规则,根据业务规则纠正各个更新数据的数据属性,得到各个更新数据的目标数据属性,根据各个更新数据的目标数据属性对各个更新数据进行纠正,生成目标数据集合;对目标数据集合进行特征提取,得到目标数据集合中各个目标数据的数据特征,根据各个目标数据的数据特征对目标数据集合建立数据索引;将建立数据索引后的目标数据集合存储到数据仓库应用层中。

[0133] 在一个实施例中,处理器执行计算机程序时还实现以下步骤:扫描数据源,获取数据源中各个数据的数据信息;读取缓冲区中的数据采集规则,该数据采集规则用于确定源数据集合中各个源数据的数据信息;将源数据集合中各个源数据的数据信息和数据源中各个数据的数据信息匹配,根据匹配结果在数据源中提取源数据集合,将源数据集合存储到数据仓库准备层中。

[0134] 在一个实施例中,处理器执行计算机程序时还实现以下步骤:当缓冲区中不存在数据采集规则时,读取规则配置文件,从规则配置文件中获取数据采集规则。

[0135] 在一个实施例中,处理器执行计算机程序时还实现以下步骤:当源数据集合中存在异常数据时,根据源数据处理规则,对异常数据进行数据清洗,得到清洗后的源数据集合;清洗后的源数据集合中包括不同数据源中相同类型的数据;当不同数据源中相同类型的数据的编码不同时,根据源数据处理规则,对不同数据源中相同类型的数据重新编码,得到更新数据集合,该更新数据集合中不同数据源中相同类型的数据的编码相同;将更新数据集合存储到数据仓库分析层中。

[0136] 在一个实施例中,处理器执行计算机程序时还实现以下步骤:业务规则包括数据属性纠正规则和业务决策规则;当各个更新数据的数据属性存在异常时,根据数据属性纠正规则,对异常的数据属性进行纠正,得到各个更新数据的目标数据属性;根据各个更新数据的目标数据属性对各个更新数据进行纠正;根据业务决策规则对纠正后的各个更新数据进行决策判断,当纠正后的各个更新数据满足决策条件时,生成纠正后的各个更新数据对应的决策数据;根据纠正后的各个更新数据和决策数据生成目标数据集合。

[0137] 在一个实施例中,处理器执行计算机程序时还实现以下步骤:数据仓库应用层包括数据查询接口,当数据查询接口接收到数据查询请求时,数据查询接口根据数据查询请

求输出对应的目标数据。

[0138] 在一个实施例中,提供了一种计算机可读存储介质,其上存储有计算机程序,计算机程序被处理器执行时实现以下步骤:采集源数据集合,将源数据集合存储到数据仓库准备层中;读取源数据处理规则,根据源数据处理规则对源数据集合进行数据转换,得到更新数据集合,将更新数据集合存储到数据仓库分析层中;对数据仓库分析层中的更新数据集合进行语义分析,得到更新数据集合中各个更新数据的数据属性;读取业务规则,根据业务规则纠正各个更新数据的数据属性,得到各个更新数据的目标数据属性,根据各个更新数据的目标数据属性对各个更新数据进行纠正,生成目标数据集合;对目标数据集合进行特征提取,得到目标数据集合中各个目标数据的数据特征,根据各个目标数据的数据特征对目标数据集合建立数据索引;将建立数据索引后的目标数据集合存储到数据仓库应用层中。

[0139] 在一个实施例中,处理器执行计算机程序时还实现以下步骤:扫描数据源,获取数据源中各个数据的数据信息;读取缓冲区中的数据采集规则,该数据采集规则用于确定源数据集合中各个源数据的数据信息;将源数据集合中各个源数据的数据信息和数据源中各个数据的数据信息匹配,根据匹配结果在数据源中提取源数据集合,将源数据集合存储到数据仓库准备层中。

[0140] 在一个实施例中,处理器执行计算机程序时还实现以下步骤:当缓冲区中不存在数据采集规则时,读取规则配置文件,从规则配置文件中获取数据采集规则。

[0141] 在一个实施例中,处理器执行计算机程序时还实现以下步骤:当源数据集合中存在异常数据时,根据源数据处理规则,对异常数据进行数据清洗,得到清洗后的源数据集合;清洗后的源数据集合中包括不同数据源中相同类型的数据;当不同数据源中相同类型的数据的编码不同时,根据源数据处理规则,对不同数据源中相同类型的数据重新编码,得到更新数据集合,该更新数据集合中不同数据源中相同类型的数据的编码相同;将更新数据集合存储到数据仓库分析层中。

[0142] 在一个实施例中,处理器执行计算机程序时还实现以下步骤:业务规则包括数据属性纠正规则和业务决策规则;当各个更新数据的数据属性存在异常时,根据数据属性纠正规则,对异常的数据属性进行纠正,得到各个更新数据的目标数据属性;根据各个更新数据的目标数据属性对各个更新数据进行纠正;根据业务决策规则对纠正后的各个更新数据进行决策判断,当纠正后的各个更新数据满足决策条件时,生成纠正后的各个更新数据对应的决策数据;根据纠正后的各个更新数据和决策数据生成目标数据集合。

[0143] 在一个实施例中,处理器执行计算机程序时还实现以下步骤:数据仓库应用层包括数据查询接口,当数据查询接口接收到数据查询请求时,数据查询接口根据数据查询请求输出对应的目标数据。

[0144] 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,是可以通过计算机程序来指令相关的硬件来完成,所述的计算机程序可存储于一非易失性计算机可读取存储介质中,该计算机程序在执行时,可包括如上述各方法的实施例的流程。其中,本申请所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用,均可包括非易失性和/或易失性存储器。非易失性存储器可包括只读存储器(ROM)、可编程ROM(PROM)、电可编程ROM(EPROM)、电可擦除可编程ROM(EEPROM)或闪存。易失性存储器可包括

随机存取存储器(RAM)或者外部高速缓冲存储器。作为说明而非局限,RAM以多种形式可得,诸如静态RAM(SRAM)、动态RAM(DRAM)、同步DRAM(SDRAM)、双数据率SDRAM(DDRSDRAM)、增强型SDRAM(ESDRAM)、同步链路(Synchlink)DRAM(SLDRAM)、存储器总线(Rambus)直接RAM(RDRAM)、直接存储器总线动态RAM(DRDRAM)、以及存储器总线动态RAM(RDRAM)等。

[0145] 以上实施例的各技术特征可以进行任意的组合,为使描述简洁,未对上述实施例中的各个技术特征所有可能的组合都进行描述,然而,只要这些技术特征的组合不存在矛盾,都应当认为是本说明书记载的范围。

[0146] 以上所述实施例仅表达了本申请的几种实施方式,其描述较为具体和详细,但并不能因此而理解为对发明专利范围的限制。应当指出的是,对于本领域的普通技术人员来说,在不脱离本申请构思的前提下,还可以做出若干变形和改进,这些都属于本申请的保护范围。因此,本申请专利的保护范围应以所附权利要求为准。

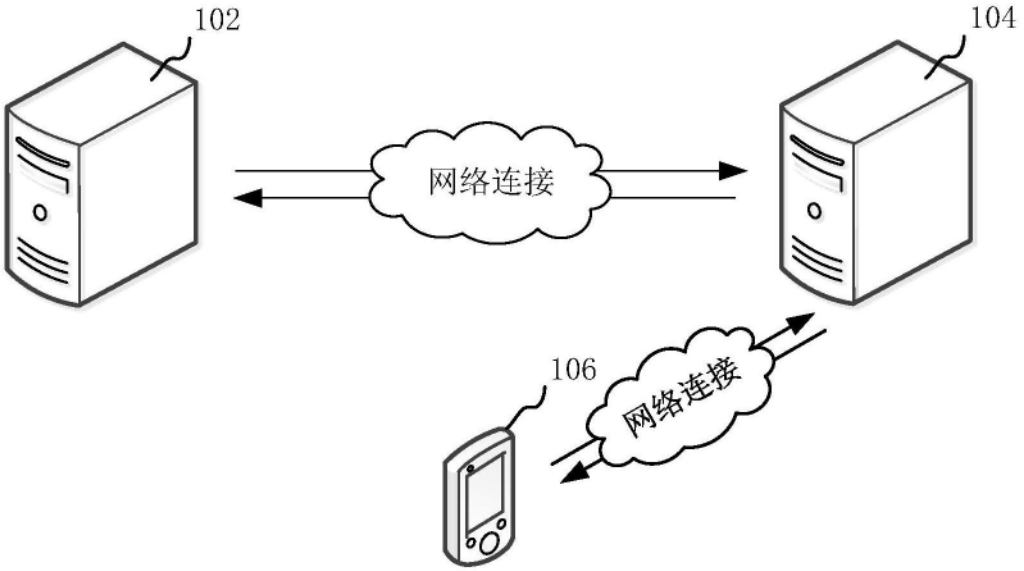


图1a

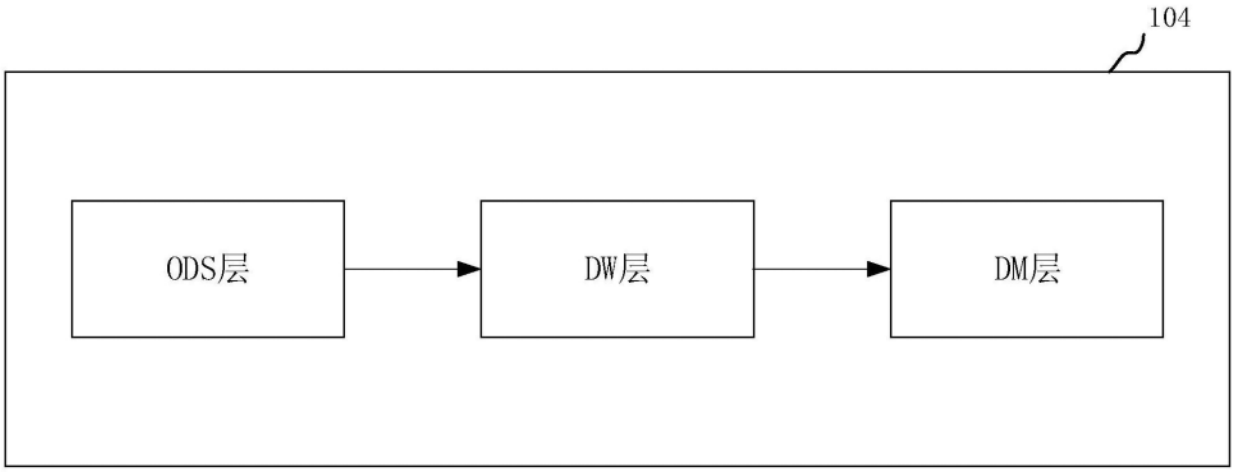


图1b

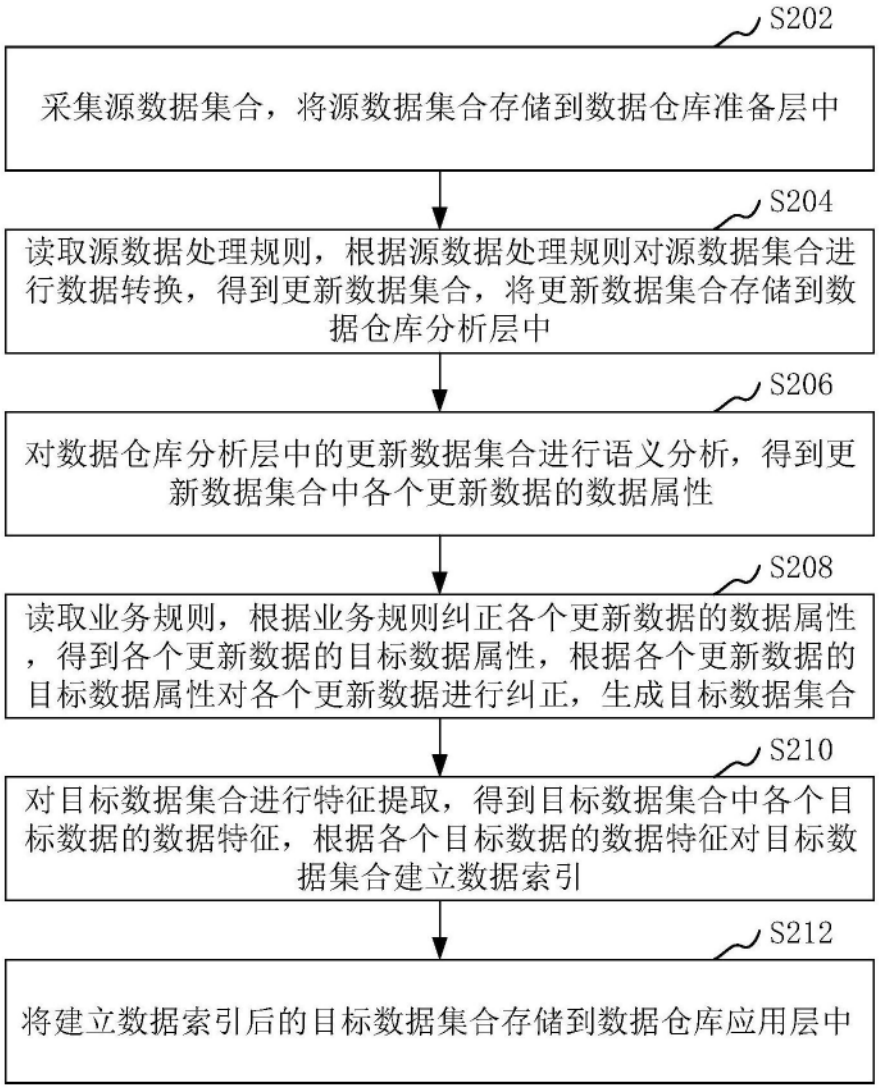

S202
采集源数据集合，将源数据集合存储到数据仓库准备层中
S204
读取源数据处理规则，根据源数据处理规则对源数据集合进行数据转换，得到更新数据集合，将更新数据集合存储到数据仓库分析层中
S206
对数据仓库分析层中的更新数据集合进行语义分析，得到更新数据集合中各个更新数据的数据属性
S208
读取业务规则，根据业务规则纠正各个更新数据的数据属性，得到各个更新数据的目标数据属性，根据各个更新数据的目标数据属性对各个更新数据进行纠正，生成目标数据集合
S210
对目标数据集合进行特征提取，得到目标数据集合中各个目标数据的数据特征，根据各个目标数据的数据特征对目标数据集合建立数据索引
S212
将建立数据索引后的目标数据集合存储到数据仓库应用层中

图2

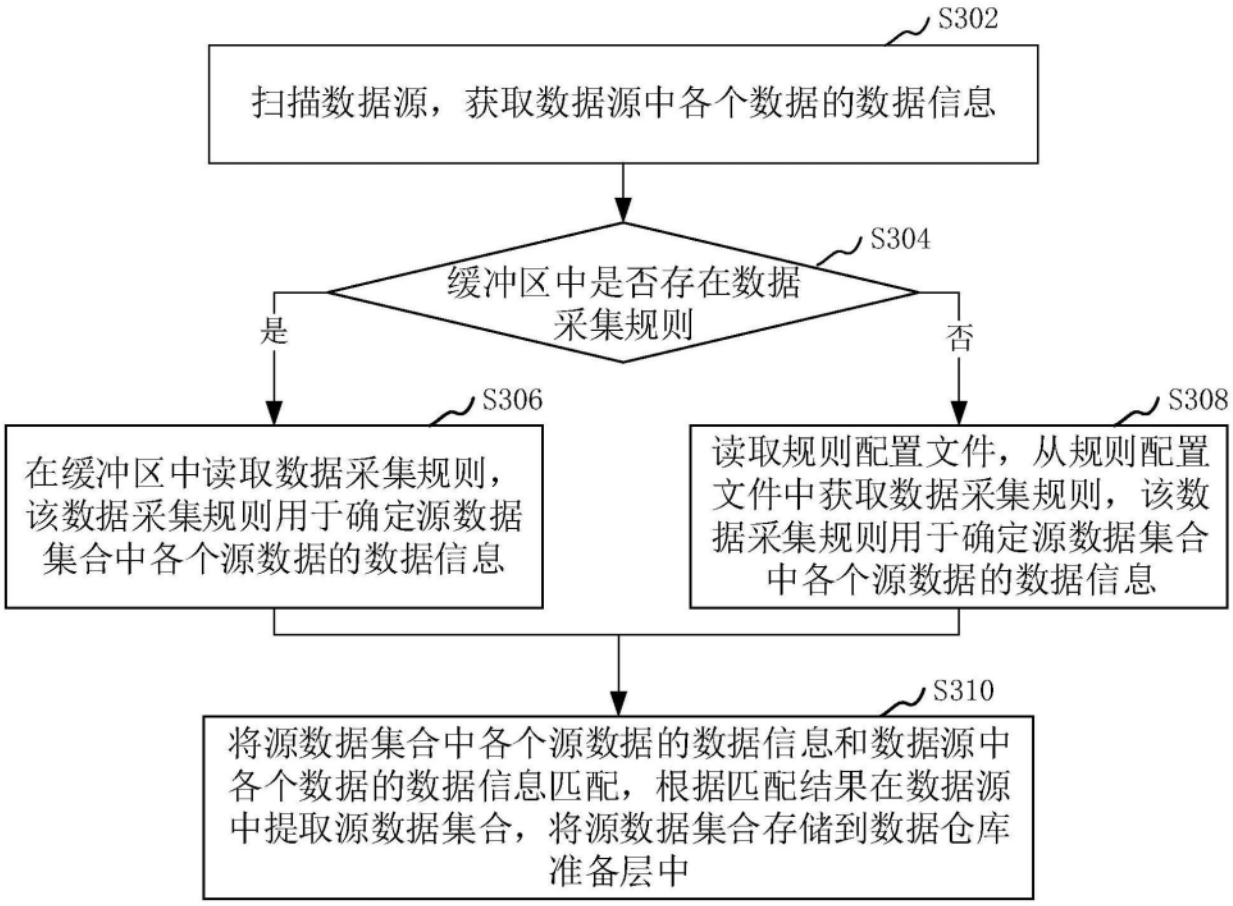

S302
扫描数据源，获取数据源中各个数据的数据信息
S304
缓冲区中是否存在数据采集规则
是
否
S306
在缓冲区中读取数据采集规则，该数据采集规则用于确定源数据集合中各个源数据的数据信息
S308
读取规则配置文件，从规则配置文件中获取数据采集规则，该数据采集规则用于确定源数据集合中各个源数据的数据信息
S310
将源数据集合中各个源数据的数据信息和数据源中各个数据的数据信息匹配，根据匹配结果在数据源中提取源数据集合，将源数据集合存储到数据仓库准备层中

图3

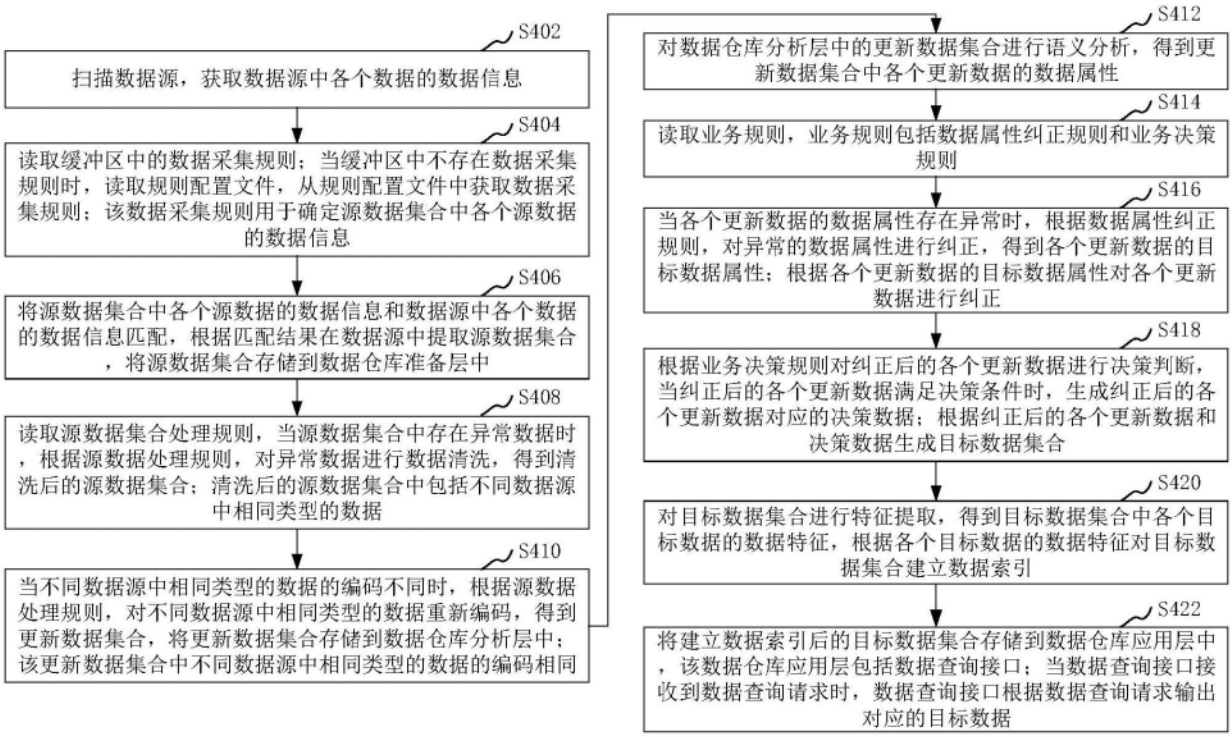

S402
扫描数据源，获取数据源中各个数据的数据信息
S404
读取缓冲区中的数据采集规则；当缓冲区中不存在数据采集规则时，读取规则配置文件，从规则配置文件中获取数据采集规则；该数据采集规则用于确定源数据集合中各个源数据的数据信息
S406
将源数据集合中各个源数据的数据信息和数据源中各个数据的数据信息匹配，根据匹配结果在数据源中提取源数据集合，将源数据集合存储到数据仓库准备层中
S408
读取源数据集合处理规则，当源数据集合中存在异常数据时，根据源数据处理规则，对异常数据进行数据清洗，得到清洗后的源数据集合；清洗后的源数据集合中包括不同数据源中相同类型的数据
S410
当不同数据源中相同类型的数据的编码不同时，根据源数据处理规则，对不同数据源中相同类型的数据重新编码，得到更新数据集合，将更新数据集合存储到数据仓库分析层中；该更新数据集合中不同数据源中相同类型的数据的编码相同
S412
对数据仓库分析层中的更新数据集合进行语义分析，得到更新数据集合中各个更新数据的数据属性
S414
读取业务规则，业务规则包括数据属性纠正规则和业务决策规则
S416
当各个更新数据的数据属性存在异常时，根据数据属性纠正规则，对异常的数据属性进行纠正，得到各个更新数据的目标数据属性；根据各个更新数据的目标数据属性对各个更新数据进行纠正
S418
根据业务决策规则对纠正后的各个更新数据进行决策判断，当纠正后的各个更新数据满足决策条件时，生成纠正后的各个更新数据对应的决策数据；根据纠正后的各个更新数据和决策数据生成目标数据集合
S420
对目标数据集合进行特征提取，得到目标数据集合中各个目标数据的数据特征，根据各个目标数据的数据特征对目标数据集合建立数据索引
S422
将建立数据索引后的目标数据集合存储到数据仓库应用层中，该数据仓库应用层包括数据查询接口；当数据查询接口接收到数据查询请求时，数据查询接口根据数据查询请求输出对应的目标数据

图4

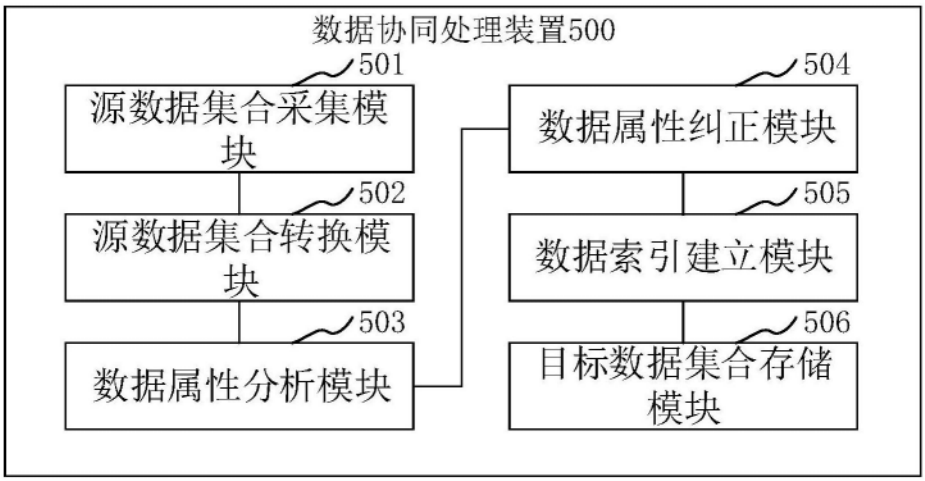

数据协同处理装置500
501
源数据集合采集模块
502
源数据集合转换模块
503
数据属性分析模块
504
数据属性纠正模块
505
数据索引建立模块
506
目标数据集合存储模块

图5

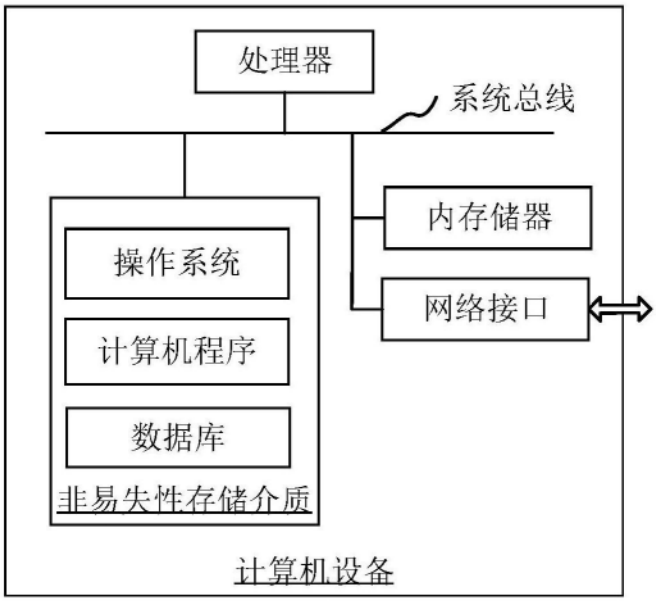


图6