



(12)发明专利申请

(10)申请公布号 CN 110413669 A
(43)申请公布日 2019. 11. 05

(21)申请号 201910542628.7

(22)申请日 2019.06.21

(71)申请人 深圳中琛源科技股份有限公司
地址 518000 广东省深圳市南山区粤海街道高新区社区科技南路16号深圳湾科技生态园11栋B1101

(72)发明人 黄义 努力曼·阿布拉 罗应军

(74)专利代理机构 广州市越秀区哲力专利商标事务所(普通合伙) 44288
代理人 齐则琳 张雷

(51)Int.Cl.
G06F 16/25(2019.01)

权利要求书2页 说明书7页 附图4页

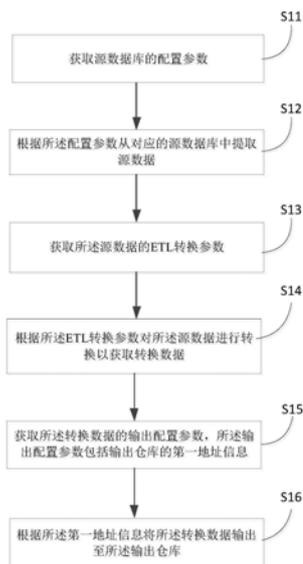
(54)发明名称

一种基于ETL的数据处理方法、装置、终端设备及存储介质

(57)摘要

本发明公开了一种基于ETL的数据处理方法、装置、终端设备及存储介质,其中该基于ETL的数据处理方法应用于终端设备,所述方法包括:获取源数据库的配置参数;根据所述配置参数从对应的源数据库中提取源数据;获取所述源数据的ETL转换参数;根据所述ETL转换参数对所述源数据进行转换以获取转换数据;获取所述转换数据的输出配置参数,所述输出配置参数包括输出仓库的第一地址信息;根据所述第一地址信息将所述转换数据输出至所述输出仓库。本发明所提供的基于ETL的数据处理方法能够实现转换数据的自动化管理,减少了人工的干预,提高了转换数据处理效率。

CN 110413669 A



1. 一种基于ETL的数据处理方法,应用于终端设备,其特征在于,所述方法包括:
 - 获取源数据库的配置参数;
 - 根据所述配置参数从对应的源数据库中提取源数据;
 - 获取所述源数据的ETL转换参数;
 - 根据所述ETL转换参数对所述源数据进行转换以获取转换数据;
 - 获取所述转换数据的输出配置参数,所述输出配置参数包括输出仓库的第一地址信息;
 - 根据所述第一地址信息将所述转换数据输出至所述输出仓库。
2. 如权利要求1所述的方法,其特征在于,所述配置参数包括源数据库的第二地址信息,所述根据所述配置参数从对应的源数据库中获取源数据,包括:
 - 根据所述第二地址信息向对应的源数据库发送数据请求,所述数据请求包括输出仓库需向终端设备发送的源数据信息;
 - 接收源数据,所述源数据为所述对应的源数据库响应所述数据请求后发送的数据。
3. 如权利要求1所述的方法,其特征在于,所述ETL转换参数包括数据映射关系,所述根据所述ETL转换参数对所述源数据进行转换以获取转换数据,包括:
 - 根据所述数据映射关系将所述源数据转换成转换数据,所述数据映射关系为所述源数据与所述转换数据的映射关系。
4. 如权利要求3所述的方法,其特征在于,所述方法还包括:
 - 判断所述转换数据是否转换成功;
 - 若所述转换数据未转换成功,则获取转换数据转换未成功的错误信息,并将所述错误信息写入记录文档中。
5. 如权利要求4所述的方法,其特征在于,所述错误信息包括至少一个错误类型,每一错误类别对应一记录文档,所述将所述错误信息写入记录文档中包括:
 - 识别所述错误信息的错误类别;
 - 根据所述错误类别,将所述错误信息写入与所述错误类别对应的记录文档中。
6. 如权利要求4所述的方法,其特征在于,所述方法还包括:若所述转换数据转换成功,则生成提示信息。
7. 一种基于ETL的数据处理装置,其特征在于,所述基于ETL的数据处理装置包括:
 - 第一获取模块,用于获取源数据库的配置参数;
 - 提取模块,用于根据所述配置参数从对应的源数据库中提取源数据;
 - 第二获取模块,用于获取所述源数据的ETL转换参数;
 - 转换模块,用于根据所述ETL转换参数对所述源数据进行转换以获取转换数据;
 - 第三获取模块,获取所述转换数据的输出配置参数,所述输出配置参数包括输出仓库的第一地址信息;
 - 输出模块,用于根据所述第一地址信息将所述转换数据输出至所述输出仓库。
8. 如权利要求7所述的数据处理装置,其特征在于:提取模块还用于根据所述第二地址信息向对应的源数据库发送数据请求;接收源数据,所述源数据为所述对应的源数据库响应所述数据请求后发送的数据。
9. 一种终端设备,其特征在于,所述终端设备包括:

存储器,所述存储器用于储存计算机可执行的基于ETL的数据处理程序;以及
处理器,所述处理器调用所述基于ETL的数据处理程序时,执行如权利要求1-6任意一
项所述的基于ETL的数据处理方法。

10.一种存储介质,其特征在于:所述存储介质上存储有转换数据处理程序,所述基于
ETL的数据处理程序可被一个或者多个处理器执行,以实现如权利要求1-6中任一项所述的
基于ETL的数据处理方法。

一种基于ETL的数据处理方法、装置、终端设备及存储介质

技术领域

[0001] 本发明涉及数据仓库技术领域,尤其涉及一种基于ETL的数据处理方法、装置、终端设备及存储介质。

背景技术

[0002] ETL(Extract-Transform-Load,抽取、转换、装载)是一种数据仓库技术,用户通过ETL任务,可以从数据源抽取出所需的数据,经过数据清洗、相互转换等过程,将数据加载到目的端。

[0003] 目前,对于ETL源数据的处理大都采用用户干预的方式,即用户在配置转换数据处理逻辑规则的同时,还必须人工指定每一步ETL操作的输入数据格式和输出数据格式。这种基于用户显式干预的源数据管理方法的智能性较差,操作起来较为繁琐,不利于推广应用。

发明内容

[0004] 本发明的主要目的在于提供一种基于ETL的数据处理方法、装置、终端设备及存储介质,旨在实现源数据管理的自动化,减少用户干预。

[0005] 为实现上述目的,本发明提供一种基于ETL的数据处理方法,应用于终端设备,所述方法包括:

[0006] 获取源数据库的配置参数;

[0007] 根据所述配置参数从对应的源数据库中提取源数据;

[0008] 获取所述源数据的ETL转换参数;

[0009] 根据所述ETL转换参数对所述源数据进行转换以获取转换数据;

[0010] 获取所述转换数据的输出配置参数,所述输出配置参数包括输出仓库的第一地址信息;

[0011] 根据所述第一地址信息将所述转换数据输出至所述输出仓库。

[0012] 优选地,所述配置参数包括源数据库的第二地址信息,所述根据所述配置参数从对应的源数据库中获取源数据,包括:

[0013] 根据所述第二地址信息向对应的源数据库发送数据请求,所述数据请求包括输出仓库需向终端设备发送的源数据信息;

[0014] 接收源数据,所述源数据为所述对应的源数据库响应所述数据请求后发送的数据。

[0015] 优选地,所述ETL转换参数包括数据映射关系,所述根据所述ETL转换参数对所述源数据进行转换以获取转换数据,包括:

[0016] 根据所述数据映射关系将所述源数据转换成转换数据,所述数据映射关系为所述源数据与所述转换数据的映射关系。

[0017] 优选地,所述方法还包括:判断所述转换数据是否转换成功;

[0018] 若所述转换数据未转换成功,则获取转换数据转换未成功的错误信息,并将所述

错误信息写入记录文档中。

[0019] 优选地,所述错误信息包括至少一个错误类型,每一错误类别对应一记录文档,所述将所述错误信息写入记录文档中包括:

[0020] 识别所述错误信息的错误类别;

[0021] 根据所述错误类别,将所述错误信息写入与所述错误类别对应的记录文档中。

[0022] 优选地,若所述转换数据转换成功,则生成提示信息。

[0023] 为实现上述目的,本发明还提供一种基于ETL的数据处理装置,所述基于ETL的数据处理装置包括:

[0024] 第一获取模块,用于获取源数据库的配置参数;

[0025] 提取模块,用于根据所述配置参数从对应的源数据库中提取源数据;

[0026] 第二获取模块,用于获取所述源数据的ETL转换参数;

[0027] 转换模块,用于根据所述ETL转换参数对所述源数据进行转换以获取转换数据;

[0028] 第三获取模块,获取所述转换数据的输出配置参数,所述输出配置参数包括输出仓库的第一地址信息;

[0029] 输出模块,用于根据所述第一地址信息将所述转换数据输出至所述输出仓库。

[0030] 优选地,提取模块还用于根据所述源数据库地址信息向对应的源数据库发送数据请求;以及接收源数据,所述源数据为所述对应的源数据库响应所述数据请求后发送给所述终端设备的数据。

[0031] 为实现上述目的,本发明还提供一种终端设备,所述终端设备包括:

[0032] 存储器,所述存储器用于储存计算机可执行的基于ETL的数据处理程序;以及

[0033] 处理器,所述处理器调用所述基于ETL的数据处理程序时,执行前述的基于ETL的数据处理方法。

[0034] 为实现上述目的,本发明还提供一种存储介质,所述存储介质上存储有基于ETL的数据处理程序,所述基于ETL的数据处理程序可被一个或者多个处理器执行,以实现前述的基于ETL的数据处理方法。

[0035] 与现有设计相比,本发明提供的基于ETL的数据处理方法、装置、终端设备及存储介质,通过获取源数据库的配置参数;根据所述配置参数从对应的源数据库中提取源数据;获取所述源数据的ETL转换参数;根据所述ETL转换参数对所述源数据进行转换以获取转换数据;获取所述转换数据的输出配置参数,所述输出配置参数包括输出仓库的第一地址信息;根据所述第一地址信息将所述转换数据输出至所述输出仓库。从而能够实现转换数据的自动化管理,减少了人工的干预,提高了转换数据处理效率。

附图说明

[0036] 图1为本发明第一实施例提供的一种基于ETL的数据处理方法的步骤流程图。

[0037] 图2为图1中步骤S12的一种具体实施方式流程图。

[0038] 图3为图1中步骤S14的一种具体实施方式流程图。

[0039] 图4为本发明第二实施例所提供的一种基于ETL的数据处理装置的结构框图。

[0040] 图5为本发明第三实施例所提供的一种终端设备的结构框图。

具体实施方式

[0041] 为了使本发明的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本发明进行进一步详细说明。应当理解,此处所描述的具体实施例仅用以解释本发明,并不用于限定本发明。基于本发明中的实施例,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,都属于本发明保护的范围。

[0042] 本申请的说明书和权利要求书及上述附图中的术语“第一”、“第二”、“第三”、“第四”等(如果存在)是用于区别类似的对象,而不必用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换,以便这里描述的实施例能够以除了在这里图示或描述的内容以外的顺序实施。此外,术语“包括”和“具有”以及他们的任何变形,意图在于覆盖不排他的包含,例如,包含了一系列步骤或单元的过程、方法、系统、产品或设备不必限于清楚地列出的那些步骤或单元,而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它步骤或单元。

[0043] 需要说明的是,在本发明中涉及“第一”、“第二”等的描述仅用于描述目的,而不能理解为指示或暗示其相对重要性或者隐含指明所指示的技术特征的数量。由此,限定有“第一”、“第二”的特征可以明示或者隐含地包括至少一个该特征。另外,各个实施例之间的技术方案可以相互结合,但是必须是以本领域普通技术人员能够实现为基础,当技术方案的结合出现相互矛盾或无法实现时应当认为这种技术方案的结合不存在,也不在本发明要求的保护范围之内。

[0044] 本发明一种基于ETL的数据处理方法、装置、终端设备及存储介质,通过获取源数据库的配置参数;根据所述配置参数从对应的源数据库中提取源数据;获取所述源数据的ETL转换参数;根据所述ETL转换参数对所述源数据进行转换以获取转换数据;获取所述转换数据的输出配置参数,所述输出配置参数包括输出仓库的第一地址信息;根据所述第一地址信息将所述转换数据输出至所述输出仓库。从而实现了转换数据的自动化管理,减少了人工的干预,提高了转换数据处理效率。

[0045] 请参阅图1,图1为本发明提供的一种基于ETL的数据处理方法,该基于ETL的数据处理方法应用于终端设备,该终端设备可以是服务器或PC机等,该基于ETL的数据处理方法包括步骤S11-S16。

[0046] 其中,步骤S11,获取源数据库的配置参数。

[0047] 源数据库的配置参数用于使得该终端设备可以根据该配置参数向对应的源数据库获取相应的源数据。该配置参数为用户通过键盘、鼠标等输入设备进行设置或选择。

[0048] 步骤S12:根据所述配置参数从对应的源数据库中提取源数据。

[0049] 终端设备根据该配置参数从对应的源数据库中提取对应的源数据,其中,配置参数包括源数据库名称、IP地址、登陆用户名或密码中的一者或多者。

[0050] 请参阅图2,在部分实施例中,配置参数包括源数据库的第二地址信息,步骤S12包括:步骤S121-S122。

[0051] 步骤S121:根据所述第二地址信息向对应的源数据库发送数据请求,所述数据请求包括输出仓库需向终端设备发送的源数据信息;

[0052] 步骤S122:接收源数据,所述源数据为所述对应的源数据库响应所述数据请求后发送给所述终端设备的数据。

[0053] 终端设备运行有一操作平台,当需要将对应的源数据添加到该操作平台时,可以依据该源数据保存的第二地址信息,即源数据库的地址信息,向该源数据库发送数据请求,该数据请求中包括了该数据库需发送的源数据信息,以使对应源数据库响应该数据请求将对应的源数据发送给该终端设备,以使终端设备将该源数据导入到该操作平台中。

[0054] 步骤S13:获取所述源数据的ETL转换参数。

[0055] 该ETL转换参数为用户根据需要通过键盘、鼠标等输入设备进行设置,以使终端设备根据该ETL转换参数对对应的源数据进行转换。

[0056] 步骤S14:根据所述ETL转换参数对所述源数据进行转换以获取转换数据。

[0057] 请参阅图3,在部分实施例中,所述ETL转换参数包括数据映射关系,步骤S14包括:

[0058] 步骤S141:根据所述数据映射关系将所述源数据转换成转换数据,所述数据映射关系为所述源数据与所述转换数据的映射关系。

[0059] 在需要对源数据进行转换时,用户可以通过键盘、鼠标等输入装置设置ETL转换参数,该ETL转换参数包括数据映射关系,数据映射关系即为所述源数据与所述转换数据的映射关系。终端设备根据该数据映射关系将对应的源数据转换成转换数据。

[0060] 在部分实施例中,步骤S14还包括:

[0061] 步骤S142:判断所述转换数据是否转换成功。

[0062] 终端设备实时监测并获取该转换数据的转换情况数据,以根据该转换情况数据判断是否将所述待转换的源数据成功转换成转换数据。

[0063] 步骤S143:若所述转换数据未转换成功,则获取转换数据转换未成功的错误信息,并将所述错误信息写入记录文档中。

[0064] 在转换数据转换过程中,终端设备实时获取该转换数据的转换情况数据并生成对应的记录文档,如log日志文件。若转换数据转换失败,则将转换失败生成的错误信息记录到该记录文档中,以便工作人员根据该记录文档可以及时获知转换失败的原因以及及时进行调试修正。

[0065] 在部分实施例中,错误信息包括多个错误类型,记录文档包括至少一个与错误类型对应的记录文档,所述将所述错误信息写入记录文档中包括:

[0066] 识别所述错误信息的错误类别;

[0067] 根据所述错误类别,将所述错误信息写入与所述错误类别对应的记录文档中。

[0068] 每一错误信息对应唯一的标识,通过标识识别错误信息的错误类别,并根据错误类别和预设的错误类别和对应记录文档的映射关系,将对应的错误信息写入对应的记录文档中,从而工作人员可以根据对应的记录文档快速识别错误类别。

[0069] 步骤S144:若所述转换数据转换成功,则生成提示信息。

[0070] 该提示信息可以是文字提示信息,声音提示信息或图像提示信息中的任意一者、任意两者或三者的组合。

[0071] 步骤S15:获取所述转换数据的输出配置参数,所述输出配置参数包括输出仓库的第一地址信息。

[0072] 步骤S16:根据所述第一地址信息将所述转换数据输出至所述输出仓库。

[0073] 用户通过键盘、鼠标等输入设备对转换数据的输出配置参数进行设置,该输出配置参数包括转换数据的输出名称、转换数据的输出类型、转换数据的访问方式、输出仓库的

第一地址信息、输出仓库的名称、输出仓库的登陆用户名称、密码中的任意一者或多者。

[0074] 终端设备根据所述第一地址信息,将转换数据输出至预设的输出仓库。

[0075] 请参阅图4,图4为本发明提供了一种基于ETL的数据处理装置10,该基于ETL的数据处理装置10包括:

[0076] 第一获取模块101,用于获取源数据库的配置参数;

[0077] 提取模块102,用于根据所述配置参数从对应的源数据库中提取源数据;

[0078] 第二获取模块103,用于获取所述源数据的ETL转换参数;

[0079] 转换模块104,用于根据所述ETL转换参数对所述源数据进行转换以获取转换数据;

[0080] 第三获取模块105,获取所述转换数据的输出配置参数,所述输出配置参数包括输出仓库的第一地址信息;以及

[0081] 输出模块106,用于根据所述第一地址信息将所述转换数据输出至所述输出仓库。

[0082] 在部分实施例中,所述配置参数包括源数据库的第二地址信息,提取模块102还用于根据所述第二地址信息向对应的源数据库发送数据请求,所述数据请求包括输出仓库需向终端设备发送的源数据信息;接收源数据,所述源数据为所述对应的源数据库响应所述数据请求后发送的数据。

[0083] 在部分实施例中,所述ETL转换参数包括待转换的源数据信息,转换模块104还用于根据所述待转换的源数据信息从所述源数据中获取待转换的源数据;根据预设关系将所述待转换的源数据转换成转换数据,所述预设关系为待转换的源数据与所述转换数据的映射关系。

[0084] 在部分实施例中,所述ETL转换参数包括数据映射关系,转换模块104还用于根据所述数据映射关系将所述源数据转换成转换数据,所述数据映射关系为所述源数据与所述转换数据的映射关系。

[0085] 在部分实施例中,转换模块104还用于判断所述转换数据是否转换成功;

[0086] 若所述转换数据未转换成功,则获取转换数据转换未成功的错误信息,并将所述错误信息写入记录文档中。

[0087] 在部分实施例中,所述错误信息包括多个错误类型,每一错误类别对应一记录文档,转换模块104还用于识别所述错误信息的错误类别;根据所述错误类别,将所述错误信息写入与所述错误类别对应的记录文档中。

[0088] 在部分实施例中,转换模块104还用于若所述转换数据转换成功,则生成提示信息。

[0089] 请参阅图5,在部分实施例中,终端设备20可以是PC(Personal Computer,个人电脑),也可以是平板电脑、掌上电脑、便携计算机、矿机或服务器。

[0090] 该终端设备20包括存储器201、处理器201和总线203,存储器201通过总线203与处理器201电连接。

[0091] 其中,存储器201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)、磁性存储器、磁盘、光盘等。存储器201在一些实施例中可以是终端设备20的内部存储单元,例如该终端设备20的硬盘。存储器201在另一些实施例中也可以是终端设备20的外部存储设备,例如终端设备20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。进一步地,存储器201还可以既包括终端设备20的内部存储单元也包括外部存储设备。存储器201不仅可以用于存储安装于终端设备20的应用软件及各类数据,例如计算机可读程序的基于ETL的数据处理程序等,也即该存储器201可以作为存储介质。

[0092] 处理器201在一些实施例中可以是一中央处理器(Central Processing Unit,CPU)、控制器、微控制器、微处理器或其他数据处理芯片,用于运行存储器201中存储的程序代码或处理数据,以实现如下步骤:

[0093] 获取源数据库的配置参数;

[0094] 根据所述配置参数从对应的源数据库中提取源数据;

[0095] 获取所述源数据的ETL转换参数;根据所述ETL转换参数对所述源数据进行转换以获取转换数据;

[0096] 获取所述转换数据的输出配置参数,所述输出配置参数包括输出仓库的第一地址信息;

[0097] 根据所述第一地址信息将所述转换数据输出至所述输出仓库。

[0098] 在部分实施例中,所述配置参数包括源数据库的第二地址信息,处理器201还用于实现以下步骤:

[0099] 根据所述第二地址信息向对应的源数据库发送数据请求,所述数据请求包括输出仓库需向终端设备发送的源数据信息;

[0100] 接收源数据,所述源数据为所述对应的源数据库响应所述数据请求后发送的数据。

[0101] 在部分实施例中,所述ETL转换参数包括待转换的源数据信息,处理器201还用于实现以下步骤:

[0102] 根据所述待转换的源数据信息从所述源数据中获取待转换的源数据;

[0103] 根据预设关系将所述待转换的源数据转换成转换数据,所述预设关系为待转换的源数据与所述转换数据的映射关系。

[0104] 在部分实施例中,所述ETL转换参数包括数据映射关系,处理器201还用于实现以下步骤:根据所述数据映射关系将所述源数据转换成转换数据,所述数据映射关系为所述源数据与所述转换数据的映射关系。

[0105] 在部分实施例中,处理器201还用于实现以下步骤:

[0106] 判断所述转换数据是否转换成功;

[0107] 若所述转换数据未转换成功,则获取转换数据转换未成功的错误信息,并将所述错误信息写入记录文档中。

[0108] 在部分实施例中,所述错误信息包括至少一个错误类型,每一错误类别对应一记录文档,处理器201还用于实现以下步骤:

[0109] 识别所述错误信息的错误类别;

[0110] 根据所述错误类别,将所述错误信息写入与所述错误类别对应的记录文档中。

[0111] 在部分实施例中,处理器201还用于实现以下步骤:

[0112] 若所述转换数据转换成功,则生成提示信息。

[0113] 本发明还提供一种存储介质,所述存储介质上存储有基于ETL的数据处理程序,所

述基于ETL的数据处理程序可被一个或者多个处理器执行,以实现前述的基于ETL的数据处理方法。

[0114] 以上仅为本发明的优选实施例,并非因此限制本发明的专利范围,凡是利用本发明说明书及附图内容所作的等效结构或等效流程变换,或直接或间接运用在其他相关的技术领域,均同理包括在本发明的专利保护范围内。

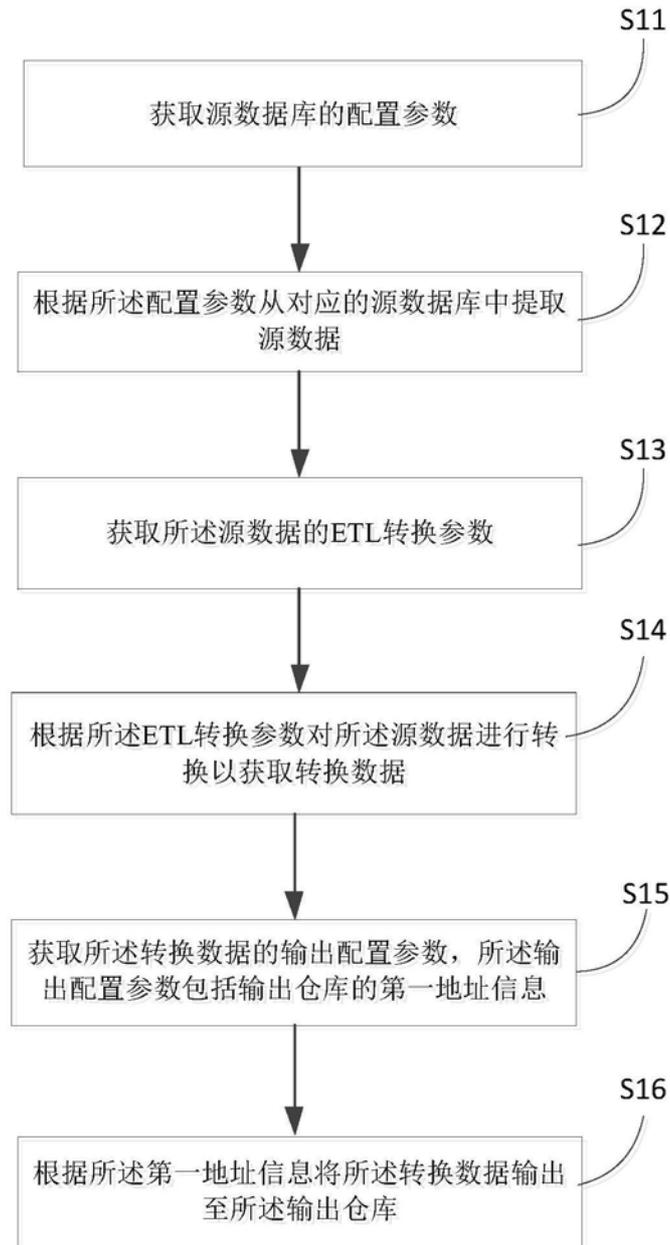


图1

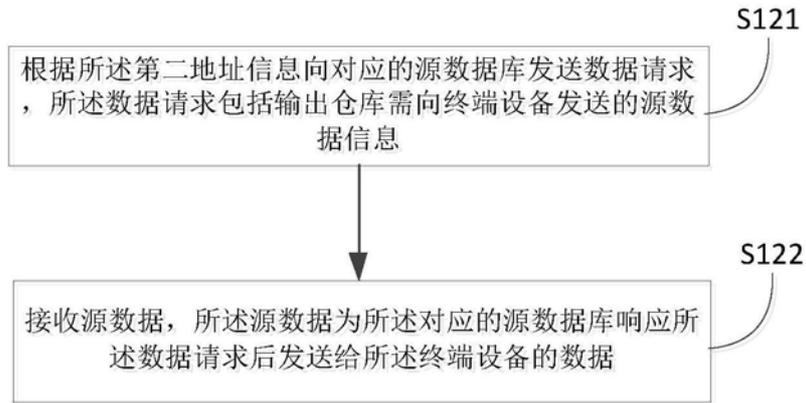


图2

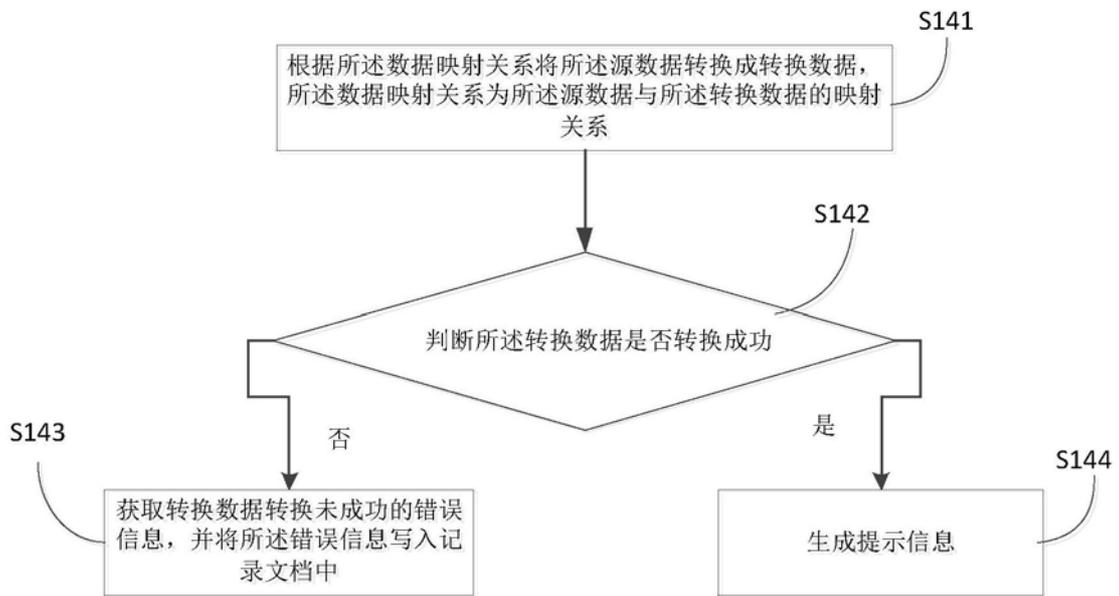


图3

10

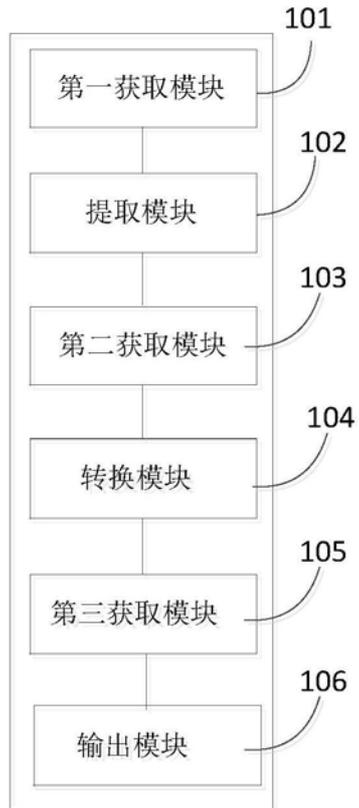


图4

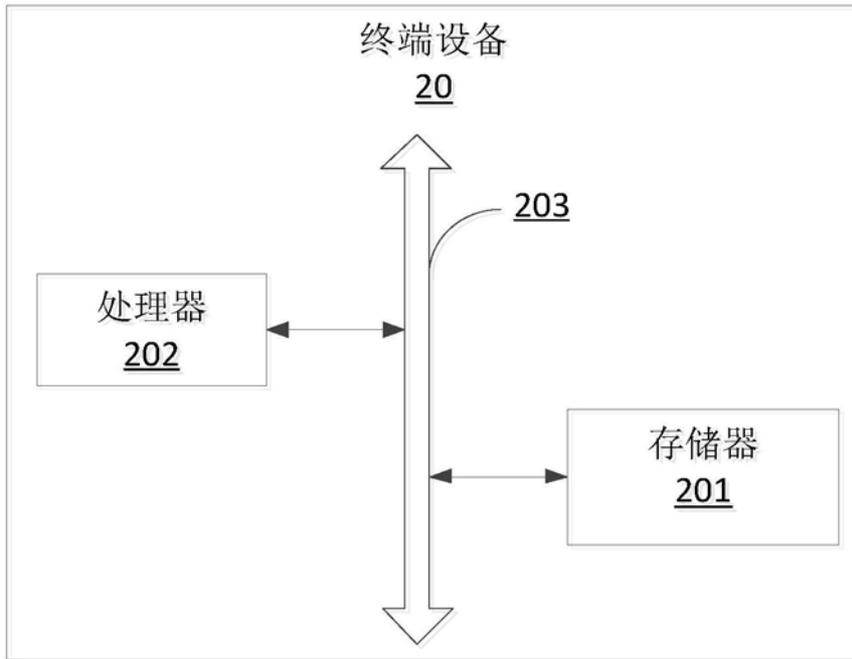


图5