



(12) 发明专利申请

(10) 申请公布号 CN 114048241 A

(43) 申请公布日 2022. 02. 15

(21) 申请号 202210029569.5

(22) 申请日 2022.01.12

(71) 申请人 树根互联股份有限公司
地址 510000 广东省广州市海珠区琶洲大道东路3号303-309房

(72) 发明人 陈明东 梁家铭 蔡孟宏 黄胜

(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
代理人 刘文强

(51) Int. Cl.
G06F 16/2457 (2019.01)
G06F 16/25 (2019.01)

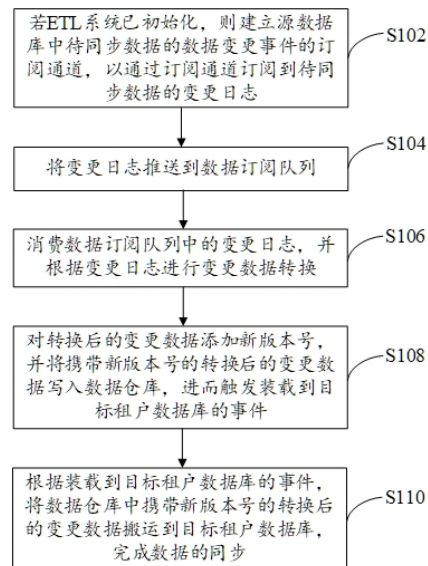
权利要求书2页 说明书11页 附图3页

(54) 发明名称

工业互联网数据的同步方法、装置和电子设备

(57) 摘要

本发明提供了一种工业互联网数据的同步方法、装置和电子设备,通过ETL系统完成源数据库中待同步数据与目标租户数据库之间的数据同步,建立了待同步数据的数据变更事件的订阅通道,当业务系统所关注的字段的数据(即待同步数据)发生变更时,便会主动将变更日志推送到数据订阅队列,进而,对数据订阅队列中的变更日志进行变更数据转换和对转换后的变更数据添加新版本号的操作,最终将携带新版本号的转换后的变更数据搬运到目标租户数据库,完成数据的同步,该种订阅数据变更事件的数据同步方式更加高效,数据同步更加实时,且无需进行API接口的开发,降低了系统被侵入的可能性。



1. 一种工业互联网数据的同步方法,其特征在于,应用于ETL系统,所述方法包括:

若所述ETL系统已初始化,则建立源数据库中待同步数据的数据变更事件的订阅通道,以通过所述订阅通道订阅到所述待同步数据的变更日志,其中,所述源数据库为工业互联网平台的数据库,所述待同步数据为业务系统所关注的字段的数据;

将所述变更日志推送到数据订阅队列;

消费所述数据订阅队列中的变更日志,并根据所述变更日志进行变更数据转换;

对转换后的变更数据添加新版本号,并将携带新版本号的转换后的变更数据写入数据仓库,进而触发装载到目标租户数据库的事件,其中,所述目标租户数据库为所述业务系统的数据库;

根据所述装载到目标租户数据库的事件,将所述数据仓库中携带新版本号的转换后的变更数据搬运到所述目标租户数据库,完成数据的同步。

2. 根据权利要求1所述的方法,其特征在于,建立源数据库中待同步数据的数据变更事件的订阅通道,包括:

通过所述源数据库的change stream API建立所述源数据库中待同步数据的数据变更事件的订阅通道。

3. 根据权利要求1所述的方法,其特征在于,消费所述数据订阅队列中的变更日志,并根据所述变更日志进行数据转换,包括:

读取所述数据订阅队列的偏移量;

根据所述偏移量读取所述数据订阅队列中的目标变更日志,其中,所述目标变更日志为所述偏移量对应的变更日志;

根据所述目标变更日志确定变更操作和所述变更操作对应的变更数据,其中,所述变更操作包括以下至少之一:增加操作和修改操作;

基于所述变更操作和所述变更操作对应的变更数据对所述变更数据进行数据治理,其中,所述数据治理包括:变更数据转换和变更数据消歧;

对数据治理后的变更数据进行级联处理和聚合运算,得到所述转换后的变更数据。

4. 根据权利要求3所述的方法,其特征在于,当所述变更操作为删除操作时,所述方法还包括:

基于所述删除操作对应的变更数据进行级联处理和聚合运算,得到所述转换后的变更数据。

5. 根据权利要求1所述的方法,其特征在于,所述方法还包括:

若所述ETL系统未初始化,则将所述源数据库中待同步数据通过dump方法推送到所述数据订阅队列,以将所述数据订阅队列中的待同步数据经过变更数据转换后作为初始版本同步至所述目标租户数据库。

6. 根据权利要求1所述的方法,其特征在于,根据所述装载到目标租户数据库的事件,将所述数据仓库中携带新版本号的转换后的变更数据搬运到所述目标租户数据库,完成数据的同步,包括:

建立与所述目标租户数据库的连接;

获取所述目标租户数据库对应的当前版本号,并根据所述当前版本号在所述数据仓库中读取携带新版本号的转换后的变更数据,以将所述携带新版本号的转换后的变更数据搬

运到所述目标租户数据库；

将所述新版本号作为所述目标租户数据库的当前版本号，并记录所述当前版本号。

7. 根据权利要求3所述的方法，其特征在于，在根据所述偏移量读取所述数据订阅队列中的目标变更日志之后，所述方法还包括：

记录下一次消费所述数据订阅队列中的变更日志时的偏移量。

8. 一种工业互联网数据的同步装置，其特征在于，应用于ETL系统，所述装置包括：

建立单元，用于若所述ETL系统已初始化，则建立源数据库中待同步数据的数据变更事件的订阅通道，以通过所述订阅通道订阅到所述待同步数据的变更日志，其中，所述源数据库为工业互联网平台的数据库，所述待同步数据为业务系统所关注的字段的数据；

推送单元，用于将所述变更日志推送到数据订阅队列；

数据转换单元，用于消费所述数据订阅队列中的变更日志，并根据所述变更日志进行变更数据转换；

版本号添加单元，用于对转换后的变更数据添加新版本号，并将携带新版本号的转换后的变更数据写入数据仓库，进而触发装载到目标租户数据库的事件，其中，所述目标租户数据库为所述业务系统的数据库；

数据搬运单元，用于根据所述装载到目标租户数据库的事件，将所述数据仓库中携带新版本号的转换后的变更数据搬运到所述目标租户数据库，完成数据的同步。

9. 一种电子设备，包括存储器、处理器及存储在所述存储器上并可在所述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上述权利要求1至7中任一项所述的方法的步骤。

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存储有机器可运行指令，所述机器可运行指令在被处理器调用和运行时，所述机器可运行指令促使所述处理器运行上述权利要求1至7中任一项所述的方法。

工业互联网数据的同步方法、装置和电子设备

技术领域

[0001] 本发明涉及数据同步的技术领域,尤其是涉及一种工业互联网数据的同步方法、装置和电子设备。

背景技术

[0002] 工业互联网平台(即云平台)是管理各个企业的设备数据的平台。在工业互联网平台中,每个企业作为一个租户存在,随着工业企业的不断发展,每个租户下的设备数量不断的快速增长,各租户下的设备通过预先创建的数据模型接入工业互联网平台后,就能将其工作时产生的工况数据上传至工业互联网平台的数据库中。而在实际的应用场景中,设备的工况数据往往还需要在不同的业务系统(例如,工业数据融合系统、指标系统等)中使用,所以,就需要将工业互联网平台的数据库中的设备数据和工况数据同步至各个业务系统。

[0003] 现有的数据同步方法是:工业互联网平台开放多个API接口,每个API接口用于传输一种数据,这样,业务系统在进行数据同步时,便被动的调不同的API接口从工业互联网平台中拉取全部其所需要的数据,将其所需的数据全部拉取完成后,再进行数据的处理。这种数据同步的过程比较低效,当将业务系统所需要全部数据拉取完成后,工业互联网平台可能又产生了新的业务系统所需要的数据,也就是业务系统拉取的数据达不到实时同步的效果,进而导致业务系统同步得到的数据出现错误,最终影响业务系统对设备的有效监控、分析和数据挖掘,甚至影响工业企业的发展以及经济社会的发展,另外,开发多个API接口,工业互联网平台和业务系统被侵入的可能性增大,加大了开发维护成本。

[0004] 综上,现有的工业互联网数据的同步方法存在数据同步不实时、系统被侵入的可能性大的技术问题。

发明内容

[0005] 有鉴于此,本发明的目的在于提供一种工业互联网数据的同步方法、装置和电子设备,以缓解现有的工业互联网数据的同步方法存在数据同步不实时、系统被侵入的可能性大的技术问题。。

[0006] 第一方面,本发明实施例提供了一种工业互联网数据的同步方法,应用于ETL系统,所述方法包括:

若所述ETL系统已初始化,则建立源数据库中待同步数据的数据变更事件的订阅通道,以通过所述订阅通道订阅到所述待同步数据的变更日志,其中,所述源数据库为工业互联网平台的数据库,所述待同步数据为业务系统所关注的字段的数据;

将所述变更日志推送到数据订阅队列;

消费所述数据订阅队列中的变更日志,并根据所述变更日志进行变更数据转换;

对转换后的变更数据添加新版本号,并将携带新版本号的转换后的变更数据写入数据仓库,进而触发装载到目标租户数据库的事件,其中,所述目标租户数据库为所述业务系统的数据库;

根据所述装载到目标租户数据库的事件,将所述数据仓库中携带新版本号的转换后的变更数据搬运到所述目标租户数据库,完成数据的同步。

[0007] 进一步的,建立源数据库中待同步数据的数据变更事件的订阅通道,包括:

通过所述源数据库的change stream API建立所述源数据库中待同步数据的数据变更事件的订阅通道。

[0008] 进一步的,消费所述数据订阅队列中的变更日志,并根据所述变更日志进行数据转换,包括:

读取所述数据订阅队列的偏移量;

根据所述偏移量读取所述数据订阅队列中的目标变更日志,其中,所述目标变更日志为所述偏移量对应的变更日志;

根据所述目标变更日志确定变更操作和所述变更操作对应的变更数据,其中,所述变更操作包括以下至少之一:增加操作和修改操作;

基于所述变更操作和所述变更操作对应的变更数据对所述变更数据进行数据治理,其中,所述数据治理包括:变更数据转换和变更数据消歧;

对数据治理后的变更数据进行级联处理和聚合运算,得到所述转换后的变更数据。

[0009] 进一步的,当所述变更操作为删除操作时,所述方法还包括:

基于所述删除操作对应的变更数据进行级联处理和聚合运算,得到所述转换后的变更数据。

[0010] 进一步的,所述方法还包括:

若所述ETL系统未初始化,则将所述源数据库中待同步数据通过dump方法推送到所述数据订阅队列,以将所述数据订阅队列中的待同步数据经过变更数据转换后作为初始版本同步至所述目标租户数据库。

[0011] 进一步的,根据所述装载到目标租户数据库的事件,将所述数据仓库中携带新版本号的转换后的变更数据搬运到所述目标租户数据库,完成数据的同步,包括:

建立与所述目标租户数据库的连接;

获取所述目标租户数据库对应的当前版本号,并根据所述当前版本号在所述数据仓库中读取携带新版本号的转换后的变更数据,以将所述携带新版本号的转换后的变更数据搬运到所述目标租户数据库;

将所述新版本号作为所述目标租户数据库的当前版本号,并记录所述当前版本号。

[0012] 进一步的,在根据所述偏移量读取所述数据订阅队列中的目标变更日志之后,所述方法还包括:

记录下一次消费所述数据订阅队列中的变更日志时的偏移量。

[0013] 第二方面,本发明实施例还提供了一种工业互联网数据的同步装置,应用于ETL系统,所述装置包括:

建立单元,用于若所述ETL系统已初始化,则建立源数据库中待同步数据的数据变更事件的订阅通道,以通过所述订阅通道订阅到所述待同步数据的变更日志,其中,所述源数据库为工业互联网平台的数据库,所述待同步数据为业务系统所关注的字段的数据;

推送单元,用于将所述变更日志推送到数据订阅队列;

数据转换单元,用于消费所述数据订阅队列中的变更日志,并根据所述变更日志进行变更数据转换;

版本号添加单元,用于对转换后的变更数据添加新版本号,并将携带新版本号的转换后的变更数据写入数据仓库,进而触发装载到目标租户数据库的事件,其中,所述目标租户数据库为所述业务系统的数据库;

数据搬运单元,用于根据所述装载到目标租户数据库的事件,将所述数据仓库中携带新版本号的转换后的变更数据搬运到所述目标租户数据库,完成数据的同步。

[0014] 第三方面,本发明实施例还提供了一种电子设备,包括存储器、处理器及存储在所述存储器上并可在所述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上述第一方面任一项所述的方法的步骤。

[0015] 第四方面,本发明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存储有机器可运行指令,所述机器可运行指令在被处理器调用和运行时,所述机器可运行指令促使所述处理器运行上述第一方面任一项所述的方法。

[0016] 在本发明实施例中,提供了一种工业互联网数据的同步方法,应用于ETL系统,该方法包括:若ETL系统已初始化,则建立源数据库中待同步数据的数据变更事件的订阅通道,以通过订阅通道订阅到待同步数据的变更日志,其中,源数据库为工业互联网平台的数据库,待同步数据为业务系统所关注的字段的数据;将变更日志推送到数据订阅队列;消费数据订阅队列中的变更日志,并根据变更日志进行变更数据转换;对转换后的变更数据添加新版本号,并将携带新版本号的转换后的变更数据写入数据仓库,进而触发装载到目标租户数据库的事件,其中,目标租户数据库为业务系统的数据库;根据装载到目标租户数据库的事件,将数据仓库中携带新版本号的转换后的变更数据搬运到目标租户数据库,完成数据的同步。通过上述描述可知,本发明的工业互联网数据的同步方法是通过ETL系统完成的源数据库中待同步数据与目标租户数据库之间的数据同步,建立了待同步数据的数据变更事件的订阅通道,当业务系统所关注的字段的数据(即待同步数据)发生变更时,便会主动将变更日志推送到数据订阅队列,进而,对数据订阅队列中的变更日志进行变更数据转换和对转换后的变更数据添加新版本号的操作,最终将携带新版本号的转换后的变更数据搬运到目标租户数据库,完成数据的同步,该种订阅数据变更事件的数据同步方式更加高效,无需将业务系统所关注的字段的全部数据同步至目标租户数据库,只需要将业务系统所关注的字段的数据中发生变更的数据同步至目标租户数据库即可,数据同步更加实时,且无需进行API接口的开发,降低了系统被侵入的可能性,缓解了现有的工业互联网数据同步方法数据同步不实时、系统被侵入的可能性大的技术问题。

附图说明

[0017] 为了更清楚地说明本发明具体实施方式或现有技术中的技术方案,下面将对具体实施方式或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图是本发明的一些实施方式,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。

[0018] 图1为本发明实施例提供的一种工业互联网数据的同步方法的流程图;

图2为本发明实施例提供的消费数据订阅队列中的变更日志,并根据变更日志进行数据转换的方法流程图;

图3为本发明实施例提供的将数据仓库中携带新版本号的转换后的变更数据搬运到目标租户数据库的方法流程图;

图4为本发明实施例提供的一种工业互联网数据的同步装置的示意图;

图5为本发明实施例提供的一种电子设备的示意图。

具体实施方式

[0019] 下面将结合实施例对本发明的技术方案进行清楚、完整地描述,显然,所描述的实施例是本发明一部分实施例,而不是全部的实施例。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。

[0020] 传统的API接口对接方式实现的数据同步比较低效,达不到实时的效果,在多数据源的情况下,还需要开发多个API接口,系统被侵入的可能性增大,因此开发维护成本高。

[0021] 基于此,本发明的工业互联网数据的同步方法是通过ETL系统完成的源数据库中待同步数据与目标租户数据库之间的数据同步,建立了待同步数据的数据变更事件的订阅通道,当业务系统所关注的字段的数据(即待同步数据)发生变更时,便会主动将变更日志推送到数据订阅队列,进而,对数据订阅队列中的变更日志进行变更数据转换和对转换后的变更数据添加新版本号的操作,最终将携带新版本号的转换后的变更数据搬运到目标租户数据库,完成数据的同步,该种订阅数据变更事件的数据同步方式更加高效,无需将业务系统所关注的字段的全部数据同步至目标租户数据库,只需要将业务系统所关注的字段的数据中发生变更的数据同步至目标租户数据库即可,数据同步更加实时,且无需进行API接口的开发,降低了系统被侵入的可能性。

[0022] 为便于对本实施例进行理解,首先对本发明实施例所公开的一种工业互联网数据的同步方法进行详细介绍。

[0023] 实施例一:

根据本发明实施例,提供了一种工业互联网数据的同步方法的实施例,需要说明的是,在附图的流程图示出的步骤可以在诸如一组计算机可执行指令的计算机系统中执行,并且,虽然在流程图中示出了逻辑顺序,但是在某些情况下,可以以不同于此处的顺序执行所示出或描述的步骤。

[0024] 图1是根据本发明实施例的一种工业互联网数据的同步方法的流程图,如图1所示,该方法包括如下步骤:

步骤S102,若ETL系统已初始化,则建立源数据库中待同步数据的数据变更事件的订阅通道,以通过订阅通道订阅到待同步数据的变更日志,其中,源数据库为工业互联网平台的数据库,待同步数据为业务系统所关注的字段的数据;

在本发明实施例中,上述工业互联网数据的同步方法应用于ETL系统,ETL(Extract-Transform-Load),用来描述将数据从源端经过抽取(extract)、交互转换(transform)、加载(load)至目的端的过程,ETL是将源数据库中的数据经过抽取、清洗转换之后加载到数据仓库的过程,因而也称为数据仓库技术,其目的是将分散、零乱、标准不统

一的数据整合到一起,为企业的决策提供分析依据。

[0025] 具体的,源数据库为工业互联网平台的数据库,可以为MongoDB数据库(一种非关系型数据库、文档型数据库)。

[0026] 具体的,可以订阅到工业互联网平台的物实例、物模型、实时工况等数据的变更日志。

[0027] 步骤S104,将变更日志推送到数据订阅队列;

步骤S106,消费数据订阅队列中的变更日志,并根据变更日志进行变更数据转换;

具体的,消费数据订阅队列中的变更日志,就会驱动ETL系统的行变更数据转换过程。

[0028] 步骤S108,对转换后的变更数据添加新版本号,并将携带新版本号的转换后的变更数据写入数据仓库,进而触发装载到目标租户数据库的事件,其中,目标租户数据库为业务系统的数据库;

上述变更数据为一条行记录(在工业互联网平台的数据库中,一个设备的数据就是一条行记录),在添加新版本号时,版本号采用原子数字序列,保证每行数据的变更都会有一个新的版本号,并将携带新版本号的转换后的变更数据写入数据仓库,进而触发装载到目标租户数据库的事件。

[0029] 数据仓库是一个面向主题的、集成的、随时间变化的,但信息本身相对稳定的数据集合,用于对管理决策过程的支持。

[0030] 上述业务系统的数据库可以为PostgreSQL数据库,其是一种关系型数据库。

[0031] 步骤S110,根据装载到目标租户数据库的事件,将数据仓库中携带新版本号的转换后的变更数据搬运到目标租户数据库,完成数据的同步。

[0032] 在本发明实施例中,提供了一种工业互联网数据的同步方法,应用于ETL系统,该方法包括:若ETL系统已初始化,则建立源数据库中待同步数据的数据变更事件的订阅通道,以通过订阅通道订阅到待同步数据的变更日志,其中,源数据库为工业互联网平台的数据库,待同步数据为业务系统所关注的字段的数据;将变更日志推送到数据订阅队列;消费数据订阅队列中的变更日志,并根据变更日志进行变更数据转换;对转换后的变更数据添加新版本号,并将携带新版本号的转换后的变更数据写入数据仓库,进而触发装载到目标租户数据库的事件,其中,目标租户数据库为业务系统的数据库;根据装载到目标租户数据库的事件,将数据仓库中携带新版本号的转换后的变更数据搬运到目标租户数据库,完成数据的同步。通过上述描述可知,本发明的工业互联网数据的同步方法是通过ETL系统完成的源数据库中待同步数据与目标租户数据库之间的数据同步,建立了待同步数据的数据变更事件的订阅通道,当业务系统所关注的字段的数据(即待同步数据)发生变更时,便会主动将变更日志推送到数据订阅队列,进而,对数据订阅队列中的变更日志进行变更数据转换和对转换后的变更数据添加新版本号的操作,最终将携带新版本号的转换后的变更数据搬运到目标租户数据库,完成数据的同步,该种订阅数据变更事件的数据同步方式更加高效,无需将业务系统所关注的字段的全部数据同步至目标租户数据库,只需要将业务系统所关注的字段的数据中发生变更的数据同步至目标租户数据库即可,数据同步更加实时,且无需进行API接口的开发,降低了系统被侵入的可能性,缓解了现有的工业互联网数据的数据同步方法数据同步不实时、系统被侵入的可能性大的技术问题。

[0033] 上述内容对本发明的工业互联网数据的同步方法进行了简要介绍,下面对其涉及到的具体内容进行详细描述。

[0034] 在本发明的一个可选实施例中,上述步骤S102,建立源数据库中待同步数据的数据变更事件的订阅通道,具体包括:

通过源数据库的change stream API建立源数据库中待同步数据的数据变更事件的订阅通道。

[0035] 具体的,通过MongoDB数据库的change stream API建立源数据库中待同步数据的数据变更事件的订阅通道。

[0036] API是一些预先定义的接口(如函数、HTTP接口),或指软件系统不同组成部分衔接的约定,用来提供应用程序与开发人员基于某软件或硬件得以访问的一组例程,而又无需访问源码,或理解内部工作机制的细节。

[0037] 在本发明的一个可选实施例中,参考图2,上述步骤S106,消费数据订阅队列中的变更日志,并根据变更日志进行数据转换,具体包括如下步骤:

步骤S201,读取数据订阅队列的偏移量;

具体的,上述偏移量用于表示数据订阅队列中待处理的变更日志的位置。

[0038] 步骤S202,根据偏移量读取数据订阅队列中的目标变更日志,其中,目标变更日志为偏移量对应的变更日志;

步骤S203,根据目标变更日志确定变更操作和变更操作对应的变更数据,其中,变更操作包括以下至少之一:增加操作和修改操作;

步骤S204,基于变更操作和变更操作对应的变更数据对变更数据进行数据治理,其中,数据治理包括:变更数据转换和变更数据消歧;

例如,变更操作为增加操作,增加了一个数据模型,对应的有数据模型的设备信息及各字段的数据信息,那么变更数据即为数据模型的设备信息及各字段的数据信息,增加的信息中有些数据在业务系统中是无法识别的,那么,就需要进行变更数据转换的处理,转换为业务系统能够识别的数据信息;另外,有些字段的数据信息对于业务系统来讲是存在歧义的,如名称字段,在工业互联网平台中保存的名称字段的名称为lastname,而业务系统中需要全名,也就是firstname+lastname,那就需要进行变更数据消歧的处理。

[0039] 上述变更数据转换和变更数据消歧都是根据业务系统的需要提前设置的关于变更数据转换和变更数据消歧的代码。

[0040] 需要说明的是,并不是所有的变更数据都需要进行上述变更数据转换和变更数据消歧的处理,也就是说根据业务系统的需要,在需要进行变更数据转换和变更数据消歧的处理时,再进行上述数据治理,如果无需进行数据治理,则不进行。

[0041] 步骤S205,对数据治理后的变更数据进行级联处理和聚合运算,得到转换后的变更数据。

[0042] 具体的,如果数据治理后的变更数据存在级联处理和聚合运算,则进行级联处理和聚合运算,这里的级联处理和聚合运算也是根据业务系统的需要和数据模型的结构预先设置的代码。

[0043] 例如,空调的数据模型下有10个空调设备,业务需要需要得到空调的数据模型的总告警数,上述变更操作为在空调的数据模型下增加1个空调设备,那么,此时空调的数据

模型下就有11个空调设备,总告警数就应该是这11个空调设备的告警数和,即需要对11个空调设备的告警数进行聚合计算,这里不再对级联处理的过程进行举例,其本质也是根据业务系统的需要、数据模型之间的关系、数据模型与设备之间的关系等进行预先设置的。

[0044] 在本发明的一个可选实施例中,当变更操作为删除操作时,该方法还包括:基于删除操作对应的变更数据进行级联处理和聚合运算,得到转换后的变更数据。

[0045] 具体的,当为删除一个设备时,只有可能触发级联处理和聚合运算的处理,不会触发数据治理的处理,所以,当变更操作为删除操作时,只进行级联处理和聚合运算,便能得到转换后的变更数据。

[0046] 在本发明的一个可选实施例,该方法还包括:

若ETL系统未初始化,则将源数据库中待同步数据通过dump方法推送到数据订阅队列,以将数据订阅队列中的待同步数据经过变更数据转换后作为初始版本同步至目标租户数据库。

[0047] 具体的,dump一般指将数据导出、转存成文件或静态形式,通过dump方法推送到数据订阅队列,即将源数据库中待同步数据导出到数据订阅队列,或,将源数据库中待同步数据转存成文件推送到数据订阅队列,或,将源数据库中待同步数据以其它静态形式推送到数据订阅队列。

[0048] 在本发明的一个可选实施例中,参考图3,上述步骤S110,根据装载到目标租户数据库的事件,将数据仓库中携带新版本号的转换后的变更数据搬运到目标租户数据库,完成数据的同步,具体包括如下步骤:

步骤S301,建立与目标租户数据库的连接;

步骤S302,获取目标租户数据库对应的当前版本号,并根据当前版本号在数据仓库中读取携带新版本号的转换后的变更数据,以将携带新版本号的转换后的变更数据搬运到目标租户数据库;

具体的,目标租户数据库对应的当前版本号是记录在ETL系统中。例如,目标租户数据库对应的当前版本号是V5,那么根据当前版本号在数据仓库中发现是从V5变成了V6版本,那么就将V6版本的那部分转换后的变更数据搬运到目标租户数据库。

[0049] 步骤S303,将新版本号作为目标租户数据库的当前版本号,并记录当前版本号。

[0050] 在本发明的一个可选实施例中,在根据偏移量读取数据订阅队列中的目标变更日志之后,该方法还包括:

记录下一次消费数据订阅队列中的变更日志时的偏移量。

[0051] 具体的,上述记录的当前版本号(步骤S303记录的当前版本号)和记录的偏移量作为ETL系统重启恢复的依据。例如,ETL系统在运行过程中发生了故障,那么需要进行重启恢复,重启恢复时,继续执行通过订阅通道订阅到待同步数据的变更日志,将变更日志推送到数据订阅队列的过程(即数据抽取的过程继续执行),然后,根据记录的偏移量继续消费数据订阅队列中的变更日志,同时,根据目标租户数据库的当前版本号,也启动根据装载到目标租户数据库的事件,将数据仓库中携带新版本号的转换后的变更数据搬运到目标租户数据库的过程,使得ETL系统恢复正常运转的工作,同时同步的过程不丢失遗漏数据。

[0052] ETL系统服务于工业领域的多租户设备数据实时同步,保证业务系统设备监控、分析、数据挖掘的数据实时性、准确性、完整性;实现了根云平台(即工业互联网平台)的接入

与建模服务的多租户数据实时ETL同步到根云平台的工业数据融合系统和指标系统(即业务系统),同时包括了对接入与建模服务的物模型、物实例数据的加工、治理和对异常数据纠偏、消歧,从数据的全量(dump方法)、增量(数据变更事件的订阅通道)、断点续传(当业务系统的数据库不可用时,则不进行装载,当业务系统的数据库恢复时,再继续进行变更数据搬运的过程)等多个场景来实现,保证了根云平台的业务系统对各工业领域的数据进行大数据指标分析,保证分析的结果的准确性,促使工业领域企业的发展,最终将会影响经济社会发展。

[0053] 因此最终目的是解决非关系型数据库到关系型数据库的之间的跨业务系统(如,工业数据融合系统和指标系统即为不同的业务系统,能实现跨这两个业务系统)的数据同步;由源数据库的变更事件来驱动ETL工作,将非关系型数据在ETL中加工、治理和对异常数据纠偏、消歧,再由内部级联处理器和聚合运算器计算;最终处理成规整的业务数据并输出到业务系统中;减少了系统之间的相互依赖(传统方案中工业互联网平台必须要开放API,业务系统才能实现数据同步,且工业互联网平台必须在线,业务系统才能数据同步,而本发明的数据同步,工业互联网平台升级等工作,不影响业务系统的使用,ETL系统是一种补偿的方式)和服务之间的耦合(传统的API对接的方式,当不同的业务系统需要相同的待同步数据时,都需要与工业互联网平台的API进行对接,各自通过API去工业互联网平台拉取数据,即不同的业务系统需要进行重复的工作,而本发明的方案只需ETL系统从源数据库中抽取一次数据,进行数据的加工后,便可同步到不同的业务系统中使用);方便业务系统内部闭环;具有低延时,低业务侵入。

[0054] 本发明的工业互联网数据的同步方法具有以下特点:订阅源数据库数据变更事件实现全量数据和增量数据输出到数据订阅队列;ETL的数据基于版本管理,每次变更的数据产生一个递增的新版本;业务系统通过版本号比对来装载新版本的数据;变更数据时触发级联更新和聚合处理;抽取(即变更日志推送到数据订阅队列),转换(即变更数据转换),装载(携带新版本号的转换后的变更数据搬运到目标租户数据库)过程并行处理,互不干涉,即转换过程不需要等抽取过程全部完成再进行,装载过程也不需要等转换过程全部完成后再进行,三者只要存在要处理的数据,可以同步执行;通过数据订阅队列的偏移量和业务系统的装载版本号来实现ETL系统的故障恢复。

[0055] 实施例二:

本发明实施例还提供了一种工业互联网数据的同步装置,该工业互联网数据的同步装置主要用于执行本发明实施例一中所提供的工业互联网数据的同步方法,以下对本发明实施例提供的工业互联网数据的同步装置做具体介绍。

[0056] 图4是根据本发明实施例的一种工业互联网数据的同步装置的示意图,如图4所示,该装置主要包括:建立单元10、推送单元20、数据转换单元30、版本号添加单元40和数据搬运单元50,其中:

建立单元,用于若ETL系统已初始化,则建立源数据库中待同步数据的数据变更事件的订阅通道,以通过订阅通道订阅到待同步数据的变更日志,其中,源数据库为工业互联网平台的数据库,待同步数据为业务系统所关注的字段的数据;

推送单元,用于将变更日志推送到数据订阅队列;

数据转换单元,用于消费数据订阅队列中的变更日志,并根据变更日志进行变更

数据转换；

版本号添加单元，用于对转换后的变更数据添加新版本号，并将携带新版本号的转换后的变更数据写入数据仓库，进而触发装载到目标租户数据库的事件，其中，目标租户数据库为业务系统的数据库；

数据搬运单元，用于根据装载到目标租户数据库的事件，将数据仓库中携带新版本号的转换后的变更数据搬运到目标租户数据库，完成数据的同步。

[0057] 在本发明实施例中，提供了一种工业互联网数据的同步装置，应用于ETL系统，该装置包括：若ETL系统已初始化，则建立源数据库中待同步数据的数据变更事件的订阅通道，以通过订阅通道订阅到待同步数据的变更日志，其中，源数据库为工业互联网平台的数据库，待同步数据为业务系统所关注的字段的数据；将变更日志推送到数据订阅队列；消费数据订阅队列中的变更日志，并根据变更日志进行变更数据转换；对转换后的变更数据添加版本号，并将携带新版本号的转换后的变更数据写入数据仓库，进而触发装载到目标租户数据库的事件，其中，目标租户数据库为业务系统的数据库；根据装载到目标租户数据库的事件，将数据仓库中携带新版本号的转换后的变更数据搬运到目标租户数据库，完成数据的同步。通过上述描述可知，本发明的工业互联网数据的同步装置是通过ETL系统完成的源数据库中待同步数据与目标租户数据库之间的数据同步，建立了待同步数据的数据变更事件的订阅通道，当业务系统所关注的字段的数据（即待同步数据）发生变更时，便会主动将变更日志推送到数据订阅队列，进而，对数据订阅队列中的变更日志进行变更数据转换和对转换后的变更数据添加新版本号的操作，最终将携带新版本号的转换后的变更数据搬运到目标租户数据库，完成数据的同步，该种订阅数据变更事件的数据同步方式更加高效，无需将业务系统所关注的字段的全部数据同步至目标租户数据库，只需要将业务系统所关注的字段的数据中发生变更的数据同步至目标租户数据库即可，数据同步更加实时，且无需进行API接口的开发，降低了系统被侵入的可能性，缓解了现有的工业互联网数据的数据同步方法数据同步不实时、系统被侵入的可能性大的技术问题。

[0058] 可选地，建立单元还用于：通过源数据库的change stream API建立源数据库中待同步数据的数据变更事件的订阅通道。

[0059] 可选地，数据转换单元还用于：读取数据订阅队列的偏移量；根据偏移量读取数据订阅队列中的目标变更日志，其中，目标变更日志为偏移量对应的变更日志；根据目标变更日志确定变更操作和变更操作对应的变更数据，其中，变更操作包括以下至少之一：增加操作和修改操作；基于变更操作和变更操作对应的变更数据对变更数据进行数据治理，其中，数据治理包括：变更数据转换和变更数据消歧；对数据治理后的变更数据进行级联处理和聚合运算，得到转换后的变更数据。

[0060] 可选地，数据转换单元还用于：当变更操作为删除操作时，基于删除操作对应的变更数据进行级联处理和聚合运算，得到转换后的变更数据。

[0061] 可选地，该装置还用于：若ETL系统未初始化，则将源数据库中待同步数据通过dump方法推送到数据订阅队列，以将数据订阅队列中的待同步数据经过变更数据转换后作为初始版本同步至目标租户数据库。

[0062] 可选地，数据搬运单元还用于：建立与目标租户数据库的连接；获取目标租户数据库对应的当前版本号，并根据当前版本号在数据仓库中读取携带新版本号的转换后的变更

数据,以将携带新版本号的转换后的变更数据搬运到目标租户数据库;将新版本号作为目标租户数据库的当前版本号,并记录当前版本号。

[0063] 可选地,该装置还用于:记录下一次消费数据订阅队列中的变更日志时的偏移量。

[0064] 本发明实施例所提供的装置,其实现原理及产生的技术效果和前述方法实施例相同,为简要描述,装置实施例部分未提及之处,可参考前述方法实施例中相应内容。

[0065] 如图5所示,本申请实施例提供的一种电子设备600,包括:处理器601、存储器602和总线,所述存储器602存储有所述处理器601可执行的机器可读指令,当电子设备运行时,所述处理器601与所述存储器602之间通过总线通信,所述处理器601执行所述机器可读指令,以执行如上述工业互联网数据的同步方法的步骤。

[0066] 具体地,上述存储器602和处理器601能够为通用的存储器和处理器,这里不做具体限定,当处理器601运行存储器602存储的计算机程序时,能够执行上述工业互联网数据的同步方法。

[0067] 处理器601可能是一种集成电路芯片,具有信号的处理能力。在实现过程中,上述方法的各步骤可以通过处理器601中的硬件的集成逻辑电路或者软件形式的指令完成。上述的处理器601可以是通用处理器,包括中央处理器(Central Processing Unit,简称CPU)、网络处理器(Network Processor,简称NP)等;还可以是数字信号处理器(Digital Signal Processing,简称DSP)、专用集成电路(Application Specific Integrated Circuit,简称ASIC)、现成可编程门阵列(Field-Programmable Gate Array,简称FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。可以实现或者执行本申请实施例中的公开的各方法、步骤及逻辑框图。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。结合本申请实施例所公开的方法的步骤可以直接体现为硬件译码处理器执行完成,或者用译码处理器中的硬件及软件模块组合执行完成。软件模块可以位于随机存储器,闪存、只读存储器,可编程只读存储器或者电可擦写可编程存储器、寄存器等本领域成熟的存储介质中。该存储介质位于存储器602,处理器601读取存储器602中的信息,结合其硬件完成上述方法的步骤。

[0068] 对应于上述工业互联网数据的同步方法,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存储有机运行指令,所述计算机可运行指令在被处理器调用和运行时,所述计算机可运行指令促使所述处理器运行上述工业互联网数据的同步方法的步骤。

[0069] 本申请实施例所提供的工业互联网数据的同步装置可以为设备上的特定硬件或者安装于设备上的软件或固件等。本申请实施例所提供的装置,其实现原理及产生的技术效果和前述方法实施例相同,为简要描述,装置实施例部分未提及之处,可参考前述方法实施例中相应内容。所属领域的技术人员可以清楚地了解到,为描述的方便和简洁,前述描述的系统、装置和单元的具体工作过程,均可以参考上述方法实施例中的对应过程,在此不再赘述。

[0070] 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。

[0071] 再例如,附图中的流程图和框图显示了根据本申请的多个实施例的装置、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段或代码的一部分,所述模块、程序段或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意,在有些作为替换的实现方式中,方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如,两个连续的方框实际上可以基本并行地执行,它们有时也可以按相反的顺序执行,这依所涉及的功能而定。也要注意的,框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合,可以用执行规定的功能或动作的专用的基于硬件的系统来实现,或者可以用专用硬件与计算机指令的组合来实现。

[0072] 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。

[0073] 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。

[0074] 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电子设备(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车辆标记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(Read-Only Memory,简称ROM)、随机存取存储器(Random Access Memory,简称RAM)、磁碟或者光盘等各种可以存储程序代码的介质。

[0075] 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。

[0076] 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权利要求的保护范围为准。

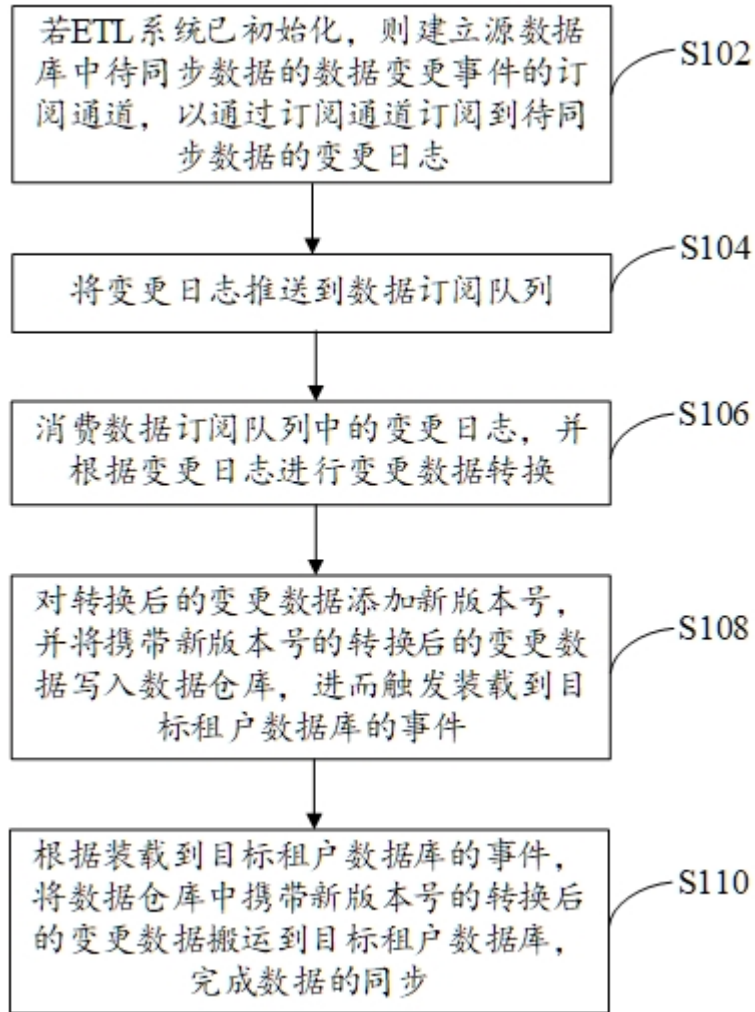


图1

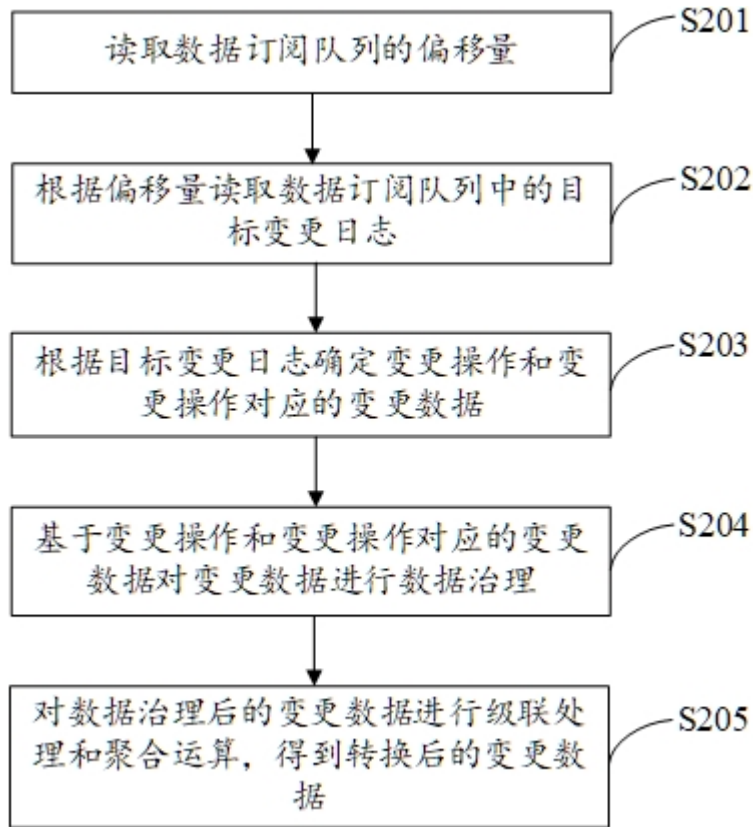


图2

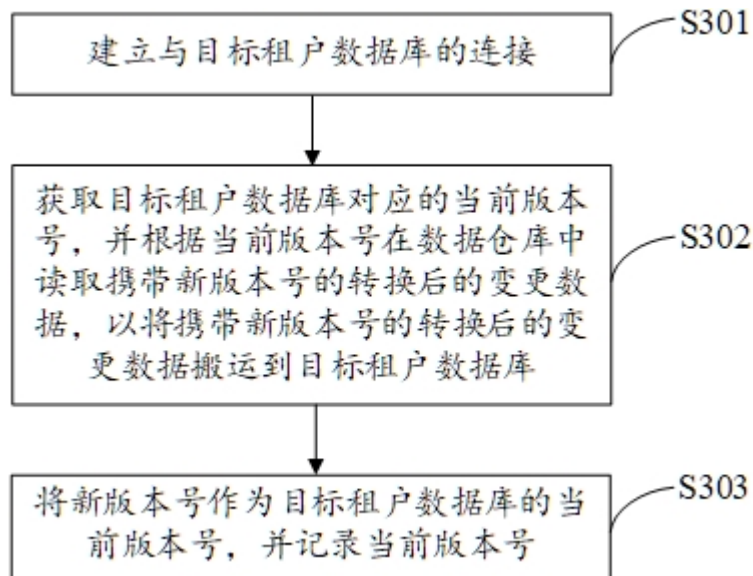


图3

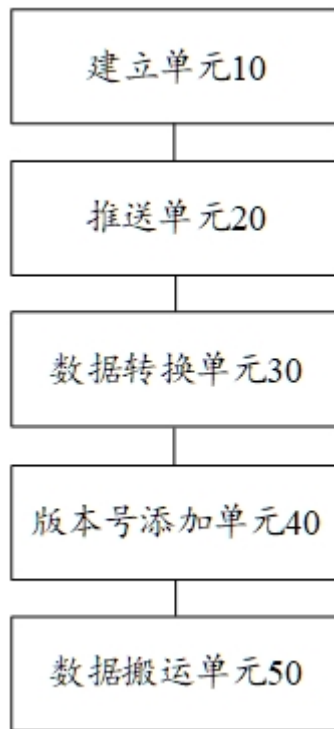


图4

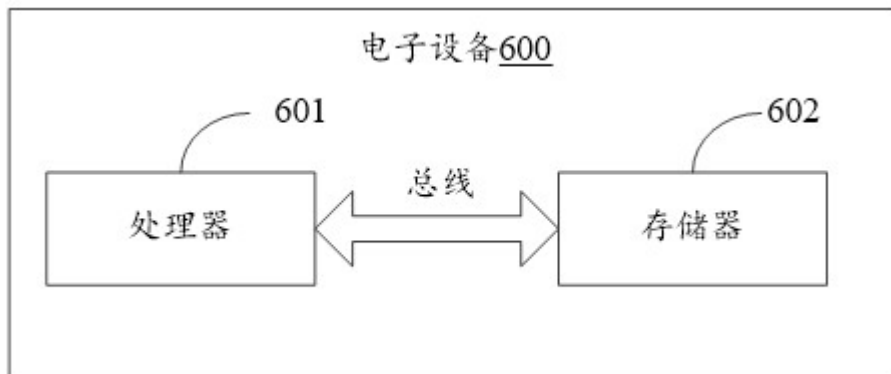


图5