



(12) 发明专利

(10) 授权公告号 CN 106909570 B

(45) 授权公告日 2020.12.15

(21) 申请号 201510981357.7

(22) 申请日 2015.12.23

(65) 同一申请的已公布的文献号  
申请公布号 CN 106909570 A

(43) 申请公布日 2017.06.30

(73) 专利权人 创新先进技术有限公司  
地址 开曼群岛大开曼岛乔治镇医院路27号  
开曼企业中心

(72) 发明人 宋忠平

(74) 专利代理机构 北京博思佳知识产权代理有限公司 11415  
代理人 林祥

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

(56) 对比文件

CN 102629263 A, 2012.08.08  
CN 103970905 A, 2014.08.06  
CN 104662535 A, 2015.05.27  
US 2015169667 A1, 2015.06.18  
US 2014019889 A1, 2014.01.16

审查员 王志超

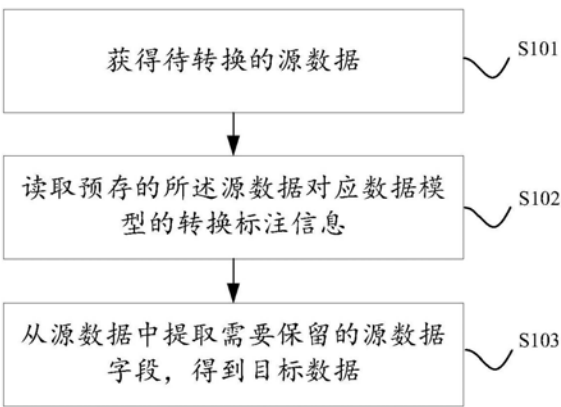
权利要求书2页 说明书6页 附图1页

(54) 发明名称

一种数据转换方法及装置

(57) 摘要

本申请公开了一种数据转换方法及装置。该数据转换方法用于将使用不同数据模型的源数据转换为使用统一需求数据模型的目标数据,该方法包括:获得待转换的源数据;读取预存的所述源数据对应数据模型的转换标注信息,所述转换标注信息中至少包括:需要保留的源数据字段信息;根据所读取的转换标注信息,从所述源数据中提取需要保留的源数据字段,得到目标数据。本申请技术方案与现有技术方案相比,不仅减少了转换逻辑的实现代码数量,而且将原本分散的多套数据转换逻辑集中在一起,从而有效降低了开发和维护成本。



1. 一种数据转换方法,应用于数据收集方,用于调用一套通用的逻辑将使用不同数据模型的源数据转换为使用统一需求数据模型的目标数据,预先存储对应于不同数据模型的不同转换标注信息,以使不同数据模型转换为同一数据模型,其特征在于,该方法包括:

获得待转换的源数据;

读取预存的所述源数据对应数据模型的转换标注信息,所述转换标注信息中至少包括:需要保留的源数据字段名称和字段类型;

根据所读取的转换标注信息,从所述源数据中提取需要保留的源数据字段,得到目标数据。

2. 根据权利要求1所述的方法,其特征在于,所述转换标注信息中还包括:需要保留的源数据字段转换后的目标字段名称;

所述根据所读取的转换标注信息,从所述源数据中提取需要保留的源数据字段,得到目标数据,包括:根据所读取的转换标注信息,从所述源数据中提取需要保留的源数据字段,并将字段名称转换为对应的目标字段名称,得到目标数据。

3. 根据权利要求1所述的方法,其特征在于,所述转换标注信息中还包括:需要保留的源数据字段转换后的目标字段类型;

所述根据所读取的转换标注信息,从所述源数据中提取需要保留的源数据字段,得到目标数据,包括:根据所读取的转换标注信息,从所述源数据中提取需要保留的源数据字段,并将字段类型转换为对应的目标字段类型,得到目标数据。

4. 根据权利要求1所述的方法,其特征在于,所述方法还包括:

将所述目标数据保存为单值数据,所述单值数据的数值中,携带该目标数据每个字段的键值对信息。

5. 根据权利要求4所述的方法,其特征在于,所述单值数据的数值中,还携带该目标数据每个字段的字段类型信息。

6. 一种数据转换装置,应用于数据收集方,用于调用一套通用的逻辑将使用不同数据模型的源数据转换为使用统一需求数据模型的目标数据,预先存储对应于不同数据模型的不同转换标注信息,以使不同数据模型转换为同一数据模型,其特征在于,该装置包括:

源数据获得模块,用于获得待转换的源数据;

标注信息读取模块,用于读取预存的所述源数据对应数据模型的转换标注信息,所述转换标注信息中至少包括:需要保留的源数据字段名称和字段类型;

目标数据生成模块,用于根据所读取的转换标注信息,从所述源数据中提取需要保留的源数据字段,得到目标数据。

7. 根据权利要求6所述的装置,其特征在于,所述转换标注信息中还包括:需要保留的源数据字段转换后的目标字段名称;

所述目标数据生成模块,具体用于:根据所读取的转换标注信息,从所述源数据中提取需要保留的源数据字段,并将字段名称转换为对应的目标字段名称,得到目标数据。

8. 根据权利要求6所述的装置,其特征在于,所述转换标注信息中还包括:需要保留的源数据字段转换后的目标字段类型;

所述目标数据生成模块,具体用于:根据所读取的转换标注信息,从所述源数据中提取需要保留的源数据字段,并将字段类型转换为对应的目标字段类型,得到目标数据。

9. 根据权利要求6所述的装置,其特征在于,所述装置还包括:

保存模块,用于将所述目标数据保存为单值数据,所述单值数据的数值中,携带该目标数据每个字段的键值对信息。

10. 根据权利要求9所述的装置,其特征在于,所述单值数据的数值中,还携带该目标数据每个字段的字段类型信息。

## 一种数据转换方法及装置

### 技术领域

[0001] 本申请涉及数据管理技术领域,尤其涉及一种数据转换方法及装置。

### 背景技术

[0002] 数据管理是指对各种数据进行收集、存储、处理和应用的过程,在大数据时代,为了应对显著增长的数据量,除了需要提供充足的存储空间之外,对数据管理技术也提出了更高的要求。

[0003] 对于数据存储方或数据使用方而言,往往需要从外部的数据源获取数据,而当外部数据源所提供的数据模型与实际使用所需的数据模型不一致时,就需要根据实际需求对外部获取的数据进行转换处理,转换的内容可能包括对字段的取舍、字段名称的变更、字段类型的变更等等。

[0004] 实际应用中,从不同的外部数据源所获取的数据模型是五花八门的,根据现有技术的实现方案,为了满足统一的数据存储及使用需求,需要针对每种外部数据模型分别编写转换逻辑。在数据源数量较多的情况下,需要投入大量的开发及维护成本,而且在很多应用场景中,数据源的数量是在不断增加的,这也使得整体的开发及维护成本难以控制。

### 发明内容

[0005] 针对上述技术问题,本申请提供一种数据转换方法及装置,技术方案如下:

[0006] 一种数据转换方法,用于将使用不同数据模型的源数据转换为使用统一需求数据模型的目标数据,该方法包括:

[0007] 获得待转换的源数据;

[0008] 读取预存的所述源数据对应数据模型的转换标注信息,所述转换标注信息中至少包括:需要保留的源数据字段信息;

[0009] 根据所读取的转换标注信息,从所述源数据中提取需要保留的源数据字段,得到目标数据。

[0010] 一种数据转换装置,用于将使用不同数据模型的源数据转换为使用统一需求数据模型的目标数据,该装置包括:

[0011] 源数据获得模块,用于获得待转换的源数据;

[0012] 标注信息读取模块,用于读取预存的所述源数据对应数据模型的转换标注信息,所述转换标注信息中至少包括:需要保留的源数据字段信息;

[0013] 目标数据生成模块,用于根据所读取的转换标注信息,从所述源数据中提取需要保留的源数据字段,得到目标数据。

[0014] 本申请所提供的技术方案,通过预先对源数据模型进行标注的方式,为不同的源数据模型配置数据转换需求,在需要对源数据进行转换时,只需调用一套通用的、能够识别上述数据转换需求的逻辑,即可实现对使用不同数据模型的源数据进行统一转换。与现有技术方案相比,不仅减少了转换逻辑的实现代码数量,而且将原本分散的多套数据转换逻

辑集中在一起,从而有效降低了开发和维护成本。

[0015] 应当理解的是,以上的一般描述和后文的细节描述仅是示例性和解释性的,并不能限制本申请。

### 附图说明

[0016] 为了更清楚地说明本申请实施例或现有技术中的技术方案,下面将对实施例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,下面描述中的附图仅仅是本申请中记载的一些实施例,对于本领域普通技术人员来讲,还可以根据这些附图获得其他的附图。

[0017] 图1是本申请的数据转换方法的流程示意图;

[0018] 图2是本申请的数据转换装置的第一种结构示意图;

[0019] 图3是本申请的数据转换装置的第二种结构示意图。

### 具体实施方式

[0020] 为了使本领域技术人员更好地理解本申请中的技术方案,下面将结合本申请实施例中的附图,对本申请实施例中的技术方案进行详细地描述,显然,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。基于本申请中的实施例,本领域普通技术人员所获得的所有其他实施例,都应当属于本申请保护的范围。

[0021] 为了解决针对使用多种数据模型分别开发转换逻辑的所带来的问题,本申请提供的方案是:根据应用需求,预先对各种源数据模型的字段进行标注,标注内容可以包括:需要保留哪些字段、是否需要改变字段名称,是否需要改变字段类型,等等。在需要对源数据进行转换时,只需调用一套通用的、能够识别上述数据转换需求的逻辑,即可实现对使用不同数据模型的源数据进行统一转换。

[0022] 图1所示,为本申请提供的数据转换方法的流程图,该方法可以包括以下步骤:

[0023] S101,获得待转换的源数据;

[0024] S102,读取预存的所述源数据对应数据模型的转换标注信息,所述转换标注信息中至少包括:需要保留的源数据字段信息;

[0025] S103,根据所读取的转换标注信息,从所述源数据中提取需要保留的源数据字段,得到目标数据。

[0026] 本申请所提供的数据转换方法,即对应于前述的通用数据转换逻辑,下面将结合具体的实例,对上述数据转换方法进行说明:

[0027]

字段名称	字段描述	字段类型
name	姓名	String
age	年龄	Long
birthday	生日	Date
height	身高	Double

[0028] 表1

[0029]

字段名称	字段描述	字段类型
name	姓名	String

age	年龄	Long
birthday	生日	Date
tel	电话号码	String

[0030] 表2

[0031] 表1和表2分别示出了两种不同的源数据模型的字段信息,其中数据模型1中包含了name、age、birthday、height四个字段,数据模型2中包含了name、age、birthday、tel四个字段。而在实际的应用需求中,仅需要使用name、age、birthday三个字段的数据。那么,根据本申请方案,可以预先对数据模型1和数据模型2的字段分别进行标注,得到两种数据模型的转换标注信息分别如表3和表4所示:

[0032]

字段名称	字段描述	字段类型	标注
name	姓名	String	1
age	年龄	Long	1
birthday	生日	Date	1
height	身高	Double	0

[0033] 表3

[0034]

字段名称	字段描述	字段类型	标注
name	姓名	String	1
age	年龄	Long	1
birthday	生日	Date	1
tel	电话号码	String	0

[0035] 表4

[0036] 在表3和表4中,以“0”和“1”分别表示对应的字段是否保留在转换结果中,可以看出,表3中的“height”与表4中的“tel”字段均被标注为0,表示对应的字段不需要保留在转换结果中。标注完毕后,将每种模型的标注信息与源数据标识对应保存。

[0037] 在转换阶段,假设传入基于数据模型1的源数据如表5所示:

[0038]

name	age	birthday	height
张三	30	1980-12-12	1.87
李四	20	1990-01-12	1.70
王五	15	1995-08-12	1.60

[0039] 表5

[0040] 根据本申请所提供的数据转换方法,获得源数据之后,首先确定该数据使用的模型是数据模型1,进而读取数据模型1的转换标注信息,根据转换标注信息可以确定需要保留的字段包括name、age、birthday,分别提取这三个字段的数据后,得到目标数据如表6所示:

[0041]

name	age	birthday
张三	30	1980-12-12
李四	20	1990-01-12
王五	15	1995-08-12

## [0042] 表6

[0043] 在上述实施例中,转换标注信息仅用于标识源数据的各个字段是否需要保留在转换结果中,也即默认不需要进行其他转换处理。而在实际的转换应用中,除了需要对源数据的字段进行取舍处理之外,可能还需要进行其他转换处理,例如对字段名称进行转换,对字段类型进行转换等等,应用本申请方案,可以通过对转换标注信息进行扩充以实现上述需求。

[0044] 假设对数据模型1进行转换时,除了需要保留name、age、birthday三个字段外,为了与需求数据模型相适应,要求转换的目标数据三个字段的名称分别为“username”、“age”、“birthday”,字段类型分别为string、short、date,可以看出,原name字段的名称、原age字段的类型均发生了变化,上述需求可以通过表7所示的标注方式实现。

## [0045]

字段名称	字段描述	字段类型	标注
name	姓名	String	@Define(targetName="username" targetType="String")
age	年龄	Long	@Define(targetName="age" targetType="Short")
birthday	生日	Date	@Define(targetName="birthday" targetType="Date")
height	身高	Double	Null

## [0046] 表7

[0047] 在表7中,除了转换后不需要保留的字段height(以Null进行标注)之外,对需要保留的3个字段均指定了转换后的目标字段名称targetName以及转换后的目标字段类型targetType,在实际进行数据转换时,先从源数据中提取需要保留的字段,然后按照指定的targetName以及targetType生成目标数据。

[0048] 可以理解的是,除了在标注信息中指定目标字段名称、目标字段类型之外,还可以指定其他转换需求,例如计量单位的转换、数值的四舍五入处理,甚至调用特定的函数对源数据字段的数值进行处理,本领域技术人员可以根据实际需求对转换标注信息进行扩展。另外需要说明的是,本申请所提供的信息标注方式仅用于示意性说明,不应理解为对本申请方案的限制。

[0049] 在实际应用中,对源数据模型的标注可以通过简单的配置操作来实现,而所换取的收益是:仅用一套通用的转换逻辑,即可实现对使用不同数据模型的源数据进行统一转换,相对于现有的多套转换逻辑而言,开发成本明显降低;另外,在通用转换逻辑中,不仅减少了大量冗余代码,而且能够将原本分散的多套数据转换逻辑集中在一起,使得代码的可读性增强,更加易于维护。

[0050] 生成目标数据之后,可以直接将所生成的数据转发给数据应用侧,也可以对所生成的数据进行存储。在本申请所提供的一种具体实施方式中,为了方便对数据进行统一存储,可以将目标数据保存为单值数据的形式。即每行都是单值数据。在这个单值数据中,需要体现出目标数据各个字段的键值对信息。具体而言,针对目标数据的每一行,需要将该行各字段名称以及对应数值写到一个字符串中,同时需要在字符串中提供分割标识以区分各个字段,这个字符串即为单值数据的数值。可选地,还可以在单值数据中添加字段的类型信息。

[0051] 实际应用中,在进行存储操作时,可以直接利用转换标注信息中的目标字段名称、目标字段类型等信息,赋予源数据的数值后得到需要存储的数据,例如,表8所示的3条数

据,其中前4列是基于前述的数据模型1的源数据,根据表7所示的标注信息,最终保存的目标数据结果如表8最后一列所示:

[0052]	姓名	年龄	生日	身高	转换后的目标数据
	张三	30	1980-12-12	1.87	{ "name": "String": "张三", "age": "Short": "30", "birthday": "Date": "1980-12-12" }
	李四	20	1990-01-12	1.70	{ "name": "String": "李四", "age": "Short": "20", "birthday": "Date": "1990-01-12" }
	王五	15	1995-08-12	1.60	{ "name": "String": "王五", "age": "Short": "15", "birthday": "Date": "1995-08-12" }

[0053] 表8

[0054] 采用上述数据保存方式,仅使用一个内容字段就可以存储任意类型、任意数量字段的数据内容,不仅维护方便,而且有效地精简了数据结构,提升了数据写入效率。后续在使用数据时,直接根据单值数据的生成规则,就可以还原得到具体的字段数值。

[0055] 相应于上述方法实施例,本申请还提供一种数据转换装置,参见图2所示,该装置可以包括:

[0056] 源数据获得模块110,用于获得待转换的源数据;

[0057] 标注信息读取模块120,用于读取预存的源数据对应数据模型的转换标注信息,其中,转换标注信息中至少包括:需要保留的源数据字段信息;

[0058] 目标数据生成模块130,用于根据所读取的转换标注信息,从源数据中提取需要保留的源数据字段,得到目标数据。

[0059] 在本申请的一种具体实施方式中,转换标注信息中还可以包括:需要保留的源数据字段转换后的目标字段名称;

[0060] 相应地,目标数据生成模块130可以具体用于:根据所读取的转换标注信息,从源数据中提取需要保留的源数据字段,并将字段名称转换为对应的目标字段名称,得到目标数据。

[0061] 在本申请的一种具体实施方式中,转换标注信息中还可以包括:需要保留的源数据字段转换后的目标字段类型;

[0062] 相应地,目标数据生成模块130可以具体用于:根据所读取的转换标注信息,从源数据中提取需要保留的源数据字段,并将字段类型转换为对应的目标字段类型,得到目标数据。

[0063] 参见图3所示,在本申请的一种具体实施方式中,上述数据转换装置还可以包括保存模块140,用于将目标数据保存为单值数据,单值数据的数值中,携带该目标数据每个字



段的键值对信息。

[0064] 可选地,在单值数据的数值中,还可以携带该目标数据每个字段的字段类型信息。

[0065] 上述装置中各个模块的功能和作用的实现过程具体详见上述方法中对应步骤的实现过程,在此不再赘述。

[0066] 通过以上的实施方式的描述可知,本领域的技术人员可以清楚地了解到本申请可借助软件加必需的通用硬件平台的方式来实现。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,该计算机软件产品可以存储在存储介质中,如ROM/RAM、磁碟、光盘等,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,服务器,或者网络设备等)执行本申请各个实施例或者实施例的某些部分所述的方法。

[0067] 本说明书中的各个实施例均采用递进的方式描述,各个实施例之间相同相似的部分互相参见即可,每个实施例重点说明的都是与其他实施例的不同之处。尤其,对于装置实施例而言,由于其基本相似于方法实施例,所以描述得比较简单,相关之处参见方法实施例的部分说明即可。以上所描述的装置实施例仅仅是示意性的,其中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,在实施本申请方案时可以把各模块的功能在同一个或多个软件和/或硬件中实现。也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。本领域普通技术人员在不付出创造性劳动的情况下,即可以理解并实施。

[0068] 以上所述仅是本申请的具体实施方式,应当指出,对于本技术领域的普通技术人员来说,在不脱离本申请原理的前提下,还可以做出若干改进和润饰,这些改进和润饰也应视为本申请的保护范围。

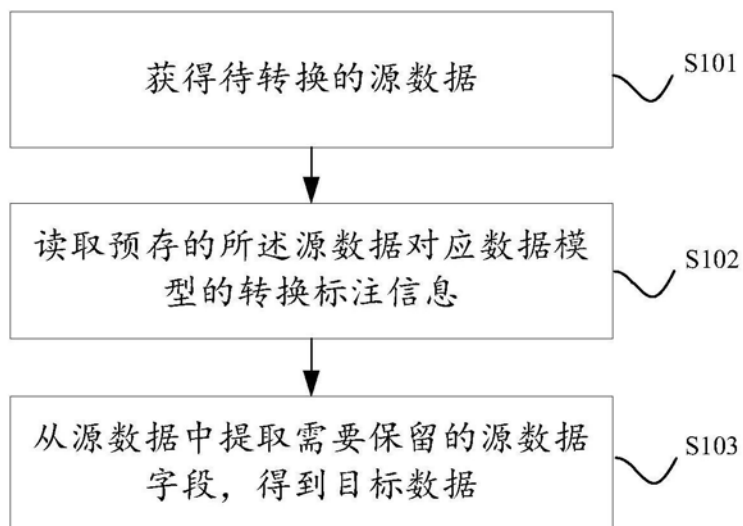


图1

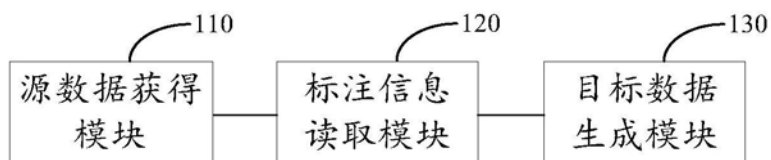


图2

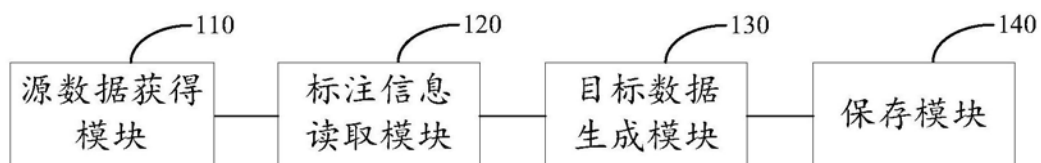


图3