



(12) 发明专利

(10) 授权公告号 CN 115269552 B

(45) 授权公告日 2025. 06. 06

(21) 申请号 202210902836.5

(22) 申请日 2022.07.29

(65) 同一申请的已公布的文献号
申请公布号 CN 115269552 A

(43) 申请公布日 2022.11.01

(73) 专利权人 广东电网有限责任公司
地址 510062 广东省广州市越秀区东风东
路757号

(72) 发明人 冯歆尧 梁盈威 周旺 王金贺
朱泰鹏 彭泽武 苏华权 谢瀚阳
潘定

(74) 专利代理机构 广州市华学知识产权代理有
限公司 44245
专利代理师 郑秋松

(51) Int.Cl.

G06F 16/21 (2019.01)

G06F 16/22 (2019.01)

G06F 16/23 (2019.01)

G06F 16/2458 (2019.01)

G06F 16/28 (2019.01)

(56) 对比文件

CN 112084182 A, 2020.12.15

CN 112732686 A, 2021.04.30

审查员 罗晓彤

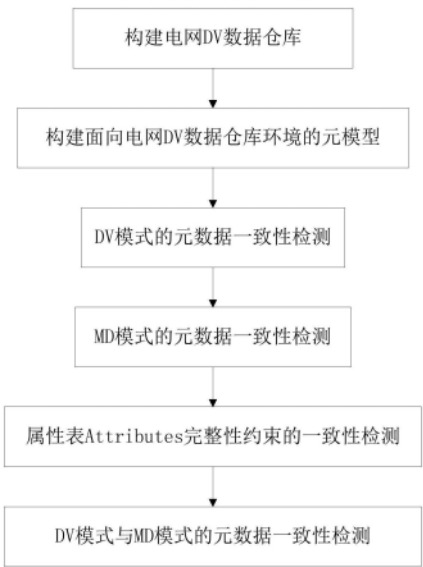
权利要求书3页 说明书10页 附图5页

(54) 发明名称

一种电网数据仓库多版本元数据存储及一
致性检测方法

(57) 摘要

本发明公开了一种电网数据仓库多版本元数据存储及一致性检测方法,该方法包括下述步骤:构建电网Data Vault (DV) 数据仓库;构建面向电网DV数据仓库环境的元模型,包括DV模式和多维 (MD) 模式部分;进行DV模式的元数据一致性检测;进行MD模式的元数据一致性检测;进行属性表Attributes完整性约束的一致性检测;进行DV模式与MD模式的元数据一致性检测;本发明利用电网元数据存储库的元模型,分别检验数据仓库区、数据集市区,以及数据仓库和数据集市之间的多版本元数据一致性关系。本发明能够在电网元数据层面自动发现元数据缺失、重复和冲突情况,从而达到提升电网数据质量的目的。



1. 一种电网数据仓库多版本元数据存储及一致性检测方法,其特征在于,包括下述步骤:

构建电网DV数据仓库,电网DV模式中采用中心点、链接、附属三类表分别存储电网业务实体、关系、中心点或链接的属性数据,电网MD模式中采用事实表、维度表分别存储电网客户的用电情况以及业务实体的属性;

构建面向电网DV数据仓库环境的元模型,包括DV模式和MD模式的元数据存储表,用于存储对应的元数据,DV模式的元数据存储表包括:中心点版本表、链接版本表、链接要素表、附属版本表和附属要素表,MD模式的元数据存储表包括:事实版本表、维度版本表、层次结构版本表、层节点要素表和层节点版本表,还设有公共部分的基本表,包括:全局版本表、属性表、属性约束表和完整性约束表,公共部分的基本表由DV模式和MD模式共享;

进行电网数据仓库的元数据一致性检测,包括DV模式的元数据一致性检测以及MD模式的元数据一致性检测,所述DV模式的元数据一致性检测包括中心点及其附属的一致性检测、链接及其附属的一致性检测;

进行属性表Attributes完整性约束的一致性检测,检测完整性约束表中是否存在相对应的完整性约束记录;

进行DV模式与MD模式的元数据一致性检测,检测链接版本表中是否存在当前版本对应的记录,输出一致性检测结果;基于当前版本的每个事实标识,获取维度版本表中对应的全部维度标识记录集合,检测在中心点版本表中是否存在对应的记录,输出一致性检测结果;

所述进行DV模式与MD模式的元数据一致性检测,具体步骤包括:

从全局标识开始,检测事实版本表中当前版本每个记录的链接标识和版本标识外键,是否在链接版本表存在一个对应的记录,如果没有则输出警告提示;

从全局标识开始,利用事实版本表中当前版本的每个事实标识,获取维度版本表中对应的全部维度标识记录集合,检测其每个记录中的业务键和事实版本标识外键,是否在中心点版本表中存在对应的记录,如果没有则输出警告提示。

2. 根据权利要求1所述的电网数据仓库多版本元数据存储及一致性检测方法,其特征在于,所述中心点及其附属的一致性检测的步骤具体包括:

从全局标识开始,获得中心点版本表中该全局标识的、当前中心点版本标识的全部业务键,检测是否在数据仓库中都有对应的保存这些业务键的中心点表存在,没有则输出错误提示,检测是否在数据仓库中都有对应的中心点表存在且该表的记录来源相符,不相符则输出错误提示;

对获得的全部业务键,检测与当前中心点版本标识和业务键外键相等的附属版本表中存储的元数据记录,是否在数据仓库中都有对应的附属表存在,没有则输出错误提示;

若在数据仓库中存在对应的保存这些业务键的附属表,则获取与该附属版本表记录对应的若干附属要素表记录,检测每个附属要素表记录对应的属性表中是否包含对应记录,没有则输出错误提示。

3. 根据权利要求1所述的电网数据仓库多版本元数据存储及一致性检测方法,其特征在于,所述链接及其附属的一致性检测的步骤具体包括:

从全局标识开始,获得链接版本表中该全局标识的、当前链接版本标识的全部链接标识,检测是否在数据仓库中都有对应的链接表存在且该链接表记录来源相符,没有或记录

来源不相符则输出错误提示；

基于当前链接版本标识的全部链接标识,经过链接要素表查找若干业务键和中心点版本标识,判断是否都存于中心点版本表中,没有则输出错误提示；

基于当前链接版本标识的全部链接标识,检测与这两个标识外键相关的附属版本表中存储的元数据,是否在数据仓库中都有对应的附属表存在,没有则输出错误提示；

若在数据仓库中有对应的链接表存在,则获取与该附属版本表记录对应的若干附属要素表记录,并检测对应属性表中是否包含对应记录,若没有则输出错误提示。

4.根据权利要求1所述的电网数据仓库多版本元数据存储及一致性检测方法,其特征在于,所述进行MD模式的元数据一致性检测,具体步骤包括:

从全局标识开始,获得事实版本表中该全局标识的、当前事实版本标识的全部事实标识,检测是否在数据集市中有对应的事实表存在,没有则输出错误提示；

对获得的全部事实标识,检测与当前事实版本标识和事实标识外键相等的每个维度版本表,是否在数据集市中有对应的维度表存在,没有则输出错误提示；

若在数据集市中有对应的维度表存在,则获取与维度版本表对应的若干层次结构,再获取与该层次结构记录对应的若干层节点要素表记录,检测其父子关系是否成立,若没有则输出错误提示,检测是否存在对应的层节点标识记录,若没有则输出错误提示；

对于层节点标识记录,通过属性表中的层节点标识外键,检测对应属性表中是否包含对应记录,若没有则输出错误提示；

对于属性表中的事实版本标识和事实标识外键,检测对应事实版本表中是否包含对应记录,若没有则输出警告提示,使用事实版本标识和事实标识,检测在数据集市中有对应的事实表存在,没有则输出错误提示。

5.根据权利要求1所述的电网数据仓库多版本元数据存储及一致性检测方法,其特征在于,所述进行属性表Attributes完整性约束的一致性检测,具体步骤包括:

对于属性表中的每个属性标识,获取与该属性标识对应的若干完整性约束记录,检测完整性约束表中是否存在相关记录成立,若没有则输出错误提示。

6.根据权利要求1所述的电网数据仓库多版本元数据存储及一致性检测方法,其特征在于,当某个元数据发生变化时,版本更新限制在DV模式或MD模式的局部,对应生成一个关于中心点版本标识、链接版本标识或事实版本标识的新版本标识。

7.根据权利要求1所述的电网数据仓库多版本元数据存储及一致性检测方法,其特征在于,元数据采用DV或MD局部批量更新的方式。

8.根据权利要求1所述的电网数据仓库多版本元数据存储及一致性检测方法,其特征在于,链接版本表与事实版本表之间存在对应的关联关系,MD模式中的事实表由DV模式中的0或1个链接表生成;中心点版本表与维度版本表之间存在对应的关联关系,MD模式中的维度表由DV模式中的0或1个中心点表生成。

9.一种电网数据仓库多版本元数据存储及一致性检测系统,其特征在于,用于实现权利要求1-8任一项所述的电网数据仓库多版本元数据存储及一致性检测方法,包括:电网DV数据仓库构建模块、元模型构建模块、电网数据仓库元数据一致性模块、完整性约束一致性检测模块、DV模式与MD模式元数据一致性检测模块；

所述电网DV数据仓库构建模块用于构建电网DV数据仓库,电网DV模式中采用中心点、

链接、附属三类表分别存储电网业务实体、关系、中心点或链接的属性数据,电网MD模式中采用事实表、维度表分别存储电网客户的用电情况以及业务实体的属性;

所述元模型构建模块用于构建面向电网DV数据仓库环境的元模型,包括DV模式和MD模式的元数据存储表,用于存储对应的元数据,DV模式的元数据存储表包括:中心点版本表、链接版本表、链接要素表、附属版本表和附属要素表,MD模式的元数据存储表包括:事实版本表、维度版本表、层次结构版本表、层节点要素表和层节点版本表,还设有公共部分的基本表,包括:全局版本表、属性表、属性约束表和完整性约束表,其中,属性表由DV模式和MD模式共享;

所述电网数据仓库元数据一致性模块用于进行电网数据仓库的元数据一致性检测,包括DV模式的元数据一致性检测以及MD模式的元数据一致性检测,所述DV模式的元数据一致性检测包括中心点及其附属的一致性检测、链接及其附属的一致性检测;

所述完整性约束一致性检测模块用于进行属性表Attributes完整性约束的一致性检测,检测完整性约束表中是否存在相对应的完整性约束记录;

所述DV模式与MD模式元数据一致性检测模块用于进行DV模式与MD模式的元数据一致性检测,检测链接版本表中是否存在当前版本对应的记录,基于当前版本的每个事实标识,获取维度版本表中对应的全部维度标识记录集合,检测在中心点版本表中是否存在对应的记录,输出一致性检测结果。

一种电网数据仓库多版本元数据存储及一致性检测方法

技术领域

[0001] 本发明涉及数据仓库元数据处理技术领域,具体涉及一种电网数据仓库多版本元数据存储及一致性检测方法。

背景技术

[0002] 在电网数据仓库建设中,电网企业一直面临着如何高效组织海量数据的技术难题。现有的以多维数据模型及其关联来实现数据仓库的传统解决方案,已经无法满足大数据时代对海量数据的组织要求,因此,电网企业开始探索引入Data Vault (DV) 建模方法,满足电网企业对大数据组织的特殊要求。

[0003] 现有的公开了一种基于表查询装置和建表装置,按表逻辑关系自动生成数据仓库的Data Vault模型,以及完成数据仓库的初始化的自动化构建方法及装置,但其业务逻辑关系十分复杂,表逻辑关系难以完全覆盖全部模式生成,且并未涉及元数据的相关问题;

[0004] 现有的公开了一种数据仓库构建装载方法和装载系统,主要包括:(1)模型输入模块,用于供用户输入模型定义并生成相应的模型;(2)模型命名模块,用于根据命名规范,输出库、表、字段的名称;(3)建表模块,用于生成相应的库和表的初始化语句,该方案能够实现数据抽取、数据装载、人物分析,以及任务调度,但并未涉及元数据存储管理的相关问题。

[0005] 在现有技术中,有的研究了数据仓库环境中面向模式演化的多维(MD)模型的多版本元数据,给出了从MD模型多版本元数据到关系模式的映射,以及标准SQL中的OLAP操作,但这种元模型是针对模式演化产生的多版本元数据建立的存储架构,而且仅涉及数据仓库中的MD模型,未涉及DV数据仓库的元数据版本管理问题,也不考虑元数据一致性检测算法问题,并且元模型中建立MAP表群用于关联新老版本,版本管理过于繁琐复杂;

[0006] 也有基于DW2.0架构中各类数据区的版本管理需求提出一种元数据版本管理元模型,用于支持版本的演化管理,但该元模型也是仅仅涉及数据仓库中的MD模型,不涉及DV模式,而且构建的结构完整性检验算法主要检查MD模型中的事实、维度、层次和属性表之间的结构关系是否符合MD模型的要求,未涉及元数据一致性的全面检测。

[0007] 在建设电网数据仓库层数据组织时,主要涉及3个数据区,首先是暂存区,在此将源业务系统数据拷贝到这里,数据经确认后按DV模型加载入数据仓库区,为了方便最终用户的使用,按多维数据模型将数据从数据仓库区,转换到数据集市区,某个时刻,多个存储区中可能形成相同数据的多个版本。在基于DV技术建设的电网数据仓库,必须协调管理多数据区、多种数据集和元数据集,这需要解决多版本元数据存储并保证其一致性的技术问题。而对多版本元数据存储需要设计相应的元模型,指导元数据存储库的构建。进一步地,检查利用这种元模型存储的元数据是否存在元数据表间记录的缺失、重复、脱节和冲突等不一致情况,以及是否存在元数据表记录与对应数据表之间的缺失、重复和脱节等不一致情况。

[0008] 综上,针对电网企业对海量数据组织具备集中化、高扩展、高可用性的平台,支持全网型、跨域数据整合,以及动态按需供应和实时分配资源的需求,传统的基于多维数据模

型的数据仓库解决方案无法做到;DV建模方法能够满足电网企业的上述大数据组织结构和运作效率要求,而在现有技术都未公开面向DV数据仓库的多版本元模型及其一致性检测技术,现有仅依靠传统数据仓库的多版本元数据存储和整体结构检测方法,对元数据的版本管理过于繁琐复杂,具有操作庞杂、运行效率低等缺陷,因此,亟需一种面向电网数据仓库多版本元数据解决其存储及一致性检测问题的技术方案。

发明内容

[0009] 为了克服现有技术存在的缺陷与不足,本发明提供一种电网数据仓库多版本元数据存储及一致性检测方法,本发明面向电网DV数据仓库环境,解决了数据仓库区和数据集市区的多版本元数据的存储和一致性检测问题,在元模型中增加了DV模式、DV模式与MD模式之间的关联,并简化了表关系,可以满足电网DV数据仓库的版本管理需求,实现电网多版本元数据存储及一致性检测。

[0010] 为了达到上述目的,本发明采用以下技术方案:

[0011] 本发明提供一种电网数据仓库多版本元数据存储及一致性检测方法,包括下述步骤:

[0012] 构建电网DV数据仓库,电网DV模式中采用中心点、链接、附属三类表分别存储电网业务实体、关系、中心点或链接的属性数据,电网MD模式中采用事实表、维度表分别存储电网客户的用电情况以及业务实体的属性;

[0013] 构建面向电网DV数据仓库环境的元模型,包括DV模式和MD模式的元数据存储表,用于存储对应的元数据,DV模式的元数据存储表包括:中心点版本表、链接版本表、链接要素表、附属版本表和附属要素表,MD模式的元数据存储表包括:事实版本表、维度版本表、层次结构版本表、层节点要素表和层节点版本表,还设有公共部分的基本表,包括:全局版本表、属性表、属性约束表和完整性约束表,公共部分的基本表由DV模式和MD模式共享;

[0014] 进行电网数据仓库的元数据一致性检测,包括DV模式的元数据一致性检测以及MD模式的元数据一致性检测,所述DV模式的元数据一致性检测包括中心点及其附属的一致性检测、链接及其附属的一致性检测;

[0015] 进行属性表Attributes完整性约束的一致性检测,检测完整性约束表中是否存在相对应的完整性约束记录;

[0016] 进行DV模式与MD模式的元数据一致性检测,检测链接版本表中是否存在当前版本对应的记录,输出一致性检测结果;基于当前版本的每个事实标识,获取维度版本表中对应的全部维度标识记录集合,检测在中心点版本表中是否存在对应的记录,输出一致性检测结果。

[0017] 作为优选的技术方案,所述中心点及其附属的一致性检测的步骤具体包括:

[0018] 从全局标识开始,获得中心点版本表中该全局标识的、当前中心点版本标识的全部业务键,检测是否在数据仓库中都有对应的保存这些业务键的中心点表存在,没有则输出错误提示,检测是否在数据仓库中都有对应的中心点表存在且该表的记录来源相符,不相符则输出错误提示;

[0019] 对获得的全部业务键,检测与当前版本标识和业务键外键相等的附属版本表中存储的元数据记录,是否在数据仓库中都有对应的附属表存在,没有则输出错误提示;

[0020] 若在数据仓库中存在对应的保存这些业务键的附属表,则获取与该附属版本表记录对应的若干附属要素表记录,检测每个附属要素表记录对应的属性表中是否包含对应记录,没有则输出错误提示。

[0021] 作为优选的技术方案,所述链接及其附属的一致性检测的步骤具体包括:

[0022] 从全局标识开始,获得链接版本表中该全局标识的、当前链接版本标识的全部链接标识,检测是否在数据仓库中都有对应的链接表存在且该链接表记录来源相符,没有或记录来源不相符则输出错误提示;

[0023] 基于当前链接版本标识的全部链接标识,经过链接要素表查找若干业务键和中心点版本标识,判断是否都存于中心点版本表中,没有则输出错误提示;

[0024] 基于当前链接版本标识的全部链接标识,检测与这两个标识外键相关的附属版本表中存储的元数据,是否在数据仓库中都有对应的附属表存在,没有则输出错误提示;

[0025] 若在数据仓库中有对应的链接表存在,则获取与该附属版本表记录对应的若干附属要素表记录,并检测对应属性表中是否包含对应记录,若没有则输出错误提示。

[0026] 作为优选的技术方案,所述进行MD模式的元数据一致性检测,具体步骤包括:

[0027] 从全局标识开始,获得事实版本表中该全局标识的、当前事实版本标识的全部事实标识,检测是否在数据集中都有对应的事实表存在,没有则输出错误提示;

[0028] 对获得的全部事实标识,检测与当前事实版本标识和事实标识外键相等的每个维度版本表,是否在数据集中都有对应的维度表存在,没有则输出错误提示;

[0029] 若在数据集中有对应的维度表存在,则获取与维度版本表对应的若干层次结构,再获取与该层次结构记录对应的若干层节点要素表记录,检测其父子关系是否成立,若没有则输出错误提示,检测是否存在对应的层节点标识记录,若没有则输出错误提示;

[0030] 对于层节点标识记录,通过属性表中的层节点标识外键,检测对应属性表中是否包含对应记录,若没有则输出错误提示;

[0031] 对于属性表中的事实版本标识和事实标识外键,检测对应事实版本表中是否包含对应记录,若没有则输出警告提示,使用事实版本标识和事实标识,检测在数据集中是否有对应的事实表存在,没有则输出错误提示。

[0032] 作为优选的技术方案,所述进行属性表Attributes完整性约束的一致性检测,具体步骤包括:

[0033] 对于属性表中的每个属性标识,获取与该属性标识对应的若干完整性约束记录,检测完整性约束表中是否存在相关记录成立,若没有则输出错误提示。

[0034] 作为优选的技术方案,所述进行DV模式与MD模式的元数据一致性检测,具体步骤包括:

[0035] 从全局标识开始,检测事实版本表中当前版本每个记录的链接标识和版本标识外键,是否在链接版本表存在一个对应的记录,如果没有则输出警告提示;

[0036] 从全局标识开始,利用事实版本表中当前版本的每个事实标识,获取维度版本表中对应的全部维度标识记录集合,检测其每个记录中的业务键和事实版本标识外键,是否在中心点版本表中存在对应的记录,如果没有则输出警告提示。

[0037] 作为优选的技术方案,当某个元数据发生变化时,版本更新限制在DV模式或MD模式的局部,对应生成一个关于中心点版本标识、链接版本标识或事实版本标识新版本标识。

[0038] 作为优选的技术方案,元数据采用DV或MD局部批量更新的方式。

[0039] 作为优选的技术方案,链接版本表与事实版本表之间存在对应的关联关系,MD模式中的事实表由DV模式中的0或1个链接表生成;中心点版本表与维度版本表之间存在对应的关联关系,MD模式中的维度表由DV模式中的0或1个中心点表生成。

[0040] 本发明还提供一种电网数据仓库多版本元数据存储及一致性检测系统,包括:电网DV数据仓库构建模块、元模型构建模块、电网数据仓库元数据一致性模块、完整性约束一致性检测模块、DV模式与MD模式元数据一致性检测模块;

[0041] 所述电网DV数据仓库构建模块用于构建电网DV数据仓库,电网DV模式中采用中心点、链接、附属三类表分别存储电网业务实体、关系、中心点或链接的属性数据,电网MD模式中采用事实表、维度表分别存储电网客户的用电情况以及业务实体的属性;

[0042] 所述元模型构建模块用于构建面向电网DV数据仓库环境的元模型,包括DV模式和MD模式的元数据存储表,用于存储对应的元数据,DV模式的元数据存储表包括:中心点版本表、链接版本表、链接要素表、附属版本表和附属要素表,MD模式的元数据存储表包括:事实版本表、维度版本表、层次结构版本表、层节点要素表和层节点版本表,还设有公共部分的基本表,包括:全局版本表、属性表、属性约束表和完整性约束表,其中,属性表由DV模式和MD模式共享;

[0043] 所述电网数据仓库元数据一致性模块用于进行电网数据仓库的元数据一致性检测,包括DV模式的元数据一致性检测以及MD模式的元数据一致性检测,所述DV模式的元数据一致性检测包括中心点及其附属的一致性检测、链接及其附属的一致性检测;

[0044] 所述完整性约束一致性检测模块用于进行属性表Attributes完整性约束的一致性检测,检测完整性约束表中是否存在相对应的完整性约束记录;

[0045] 所述DV模式与MD模式元数据一致性检测模块用于进行DV模式与MD模式的元数据一致性检测,检测链接版本表中是否存在当前版本对应的记录,基于当前版本的每个事实标识,获取维度版本表中对应的全部维度标识记录集合,检测在中心点版本表中是否存在对应的记录,输出一致性检测结果。

[0046] 本发明与现有技术相比,具有如下优点和有益效果:

[0047] (1) 本发明面向电网DV数据仓库环境,解决了数据仓库区和数据集市区的多版本元数据存储的技术问题,在元模型中增加了DV模式、DV模式与MD模式之间的关联,并简化了表关系,可以满足电网DV数据仓库的版本管理需求,达到了电网数据仓库多版本元数据的高效存储。

[0048] (2) 本发明面向电网DV数据仓库环境,采用上述元模型存储相应的多版本元数据,并根据元数据类型的定义及其关系特征,解决了元数据一致性检测的技术问题,达到了电网多版本元数据的自动一致性验证,能够在元数据层面自动发现数据缺失、重复和冲突情况,从而提升电网数据质量。

附图说明

[0049] 图1为本发明电网多版本元数据存储及一致性检测方法的流程示意图;

[0050] 图2(a)为本发明电网数据仓库环境中的电网DV模式的片段示意图;

[0051] 图2(b)为本发明电网数据仓库环境中的电网MD模式的片段示意图;

- [0052] 图3为本发明面向电网DV数据仓库环境的元模型示意图；
[0053] 图4为本发明DV模式的元数据一致性检测的流程示意图；
[0054] 图5为本发明MD模式的元数据一致性检测的流程示意图；
[0055] 图6为本发明属性表Attributes完整性约束的一致性检测的流程示意图；
[0056] 图7为本发明DV模式与MD模式的元数据一致性检测的流程示意图。

具体实施方式

[0057] 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仅用以解释本发明,并不用于限定本发明。

[0058] 实施例1

[0059] 如图1所示,本实施例提供一种电网数据仓库多版本元数据存储及一致性检测方法,包括下述步骤:

[0060] S1:构建电网DV数据仓库;

[0061] 为了组织电网的海量数据,针对电网的数据组织要求:具备集中化、高扩展、高可用性的平台,支持全网型、跨域数据整合,以及动态按需供应和实时分配资源,构建电网DV数据仓库,具体包括:

[0062] 1)如图2(a)所示,用中心点、链接、附属三类表分别保存电网业务实体、关系和中心点/链接的属性数据;在图中有电网客户中心点表Hub_Customer(有客户附属表Sat_Customer和客户地址附属表Sat_CustAddr)、用电合同中心点表Hub_Contract(有合同附属表Sat_Contract)和服务中心点表Hub_Service(有服务附属表Sat_Service),还有客户-合同链接表Lnk_Cust-Cont和合同-服务链接表Lnk_Cont-Serv(有合同-服务附属表Sat_Cont-Serv);

[0063] 2)中心点表以业务键为基础连接链接表、附属表,链接表保存多对多关系;

[0064] 3)一个中心点或链路可能有多个附属,每个附属按时间戳(Load_Date)形成相关属性的不同时期的历史记录;

[0065] 4)中心点、链接、附属三类表中保存有记录来源(Rec_Source)属性,可以保存所有源电网业务系统的数据集。

[0066] 电网DV模式适合高效存储电网的海量数据,但不利于用户使用,所以在DV数据仓库环境中,还包括电网MD模式,主要有事实表和维度表构成的多维模型,用于支持最终用户分析数据,如图2(b)所示,具体包括:

[0067] 1)用事实表保存电网客户的用电情况,有事实表MD_Usage,其中包括两个度量属性用电量和金额,在主键区有连接4个维度表的外键;

[0068] 2)维度表保存业务实体的属性,有客户维度MD_Customer、合同维度MD_Contract、服务维度MD_Service,以及公用日期维度MD_Data,这些维度表的主键共同确定事实表中的某个记录,即用电量和金额;

[0069] 3)维度表中可以包含层次结构,层次结构由层节点构成,每个层节点对应一个属性,如公用日期维度MD_Data中有三个层节点构成层次结构“日期-月份-年份”,客户维度MD_Customer中有四个层节点构成层次结构“客户姓名-用电地址-城市-省市”。

[0070] 上述对电网海量数据的组织最大限度地满足集中化、高扩展、高可用性的电网数据平台的需求,同时支持全网型、跨域数据整合,以及动态按需供应和实时分配资源。

[0071] S2:构建面向电网DV数据仓库环境的元模型;

[0072] 在本实施例中,元模型用于存储与数据仓库和数据集市有关的多版本元数据,其中,一个元数据的模式版本由DV模式和MD模式的表构成。

[0073] 如图3所示,整个元模型分成三个部分,DV模式部分包括5个基本元数据存储表:中心点版本表HUB_Vers、链接版本表Lnk_Vers、链接要素表Lnk_Eles、附属版本表Sat_Vers表和附属要素表Sat_Eles;MD模式部分包括5个基本元数据存储表:事实版本表FACT_Vers、维度版本表Dim_Vers、层次结构版本表Hier_Vers、层节点要素表Hier_Eles、层节点版本表Lev_Vers;公共部分包括4个基本表:全局版本表Versions、属性表Attributes、属性约束表Att_Cons和完整性约束表Int_Cons,其中,图中的线条表示各个表之间的联系,虚线表示非依赖关系,实线表示依赖关系,线条端点的黑圆点直连表示多对1的关系,菱形表示0或1;

[0074] 全局版本表Versions存储的元数据是关于电网全局版本的,主要涉及中心点、链接和事实表,包括一个唯一的版本标识、名字、开始和结束的有效时间,以及状态(是否一个版本是递交的或是正在开发的)等。形成一个稳定版本后,当某个元数据发生变化时,若只影响到DV或MD局部,版本更新就限制其在DV模式或MD模式的局部,对应生成一个关于中心点版本标识Hub_vers_id、链接版本标识Lnk_vers_id或事实版本标识FV_id的新版本标识,而不生成Versions表的全局标识VER_id。为避免在DV模式或MD模式局部频繁生成全局新版本标识,这里采用局部批量更新元数据的方式,简化多版本元数据管理。

[0075] 为了保存源系统相关的元数据,DV模式中的中心点版本表HUB_Vers、链接版本表Lnk_Vers、附属版本表Sat_Vers分别包含着中心点、链接和附属数据表中的元数据,可以与DV数据表中的源表和数据加载元数据相互印证,如这些表中的记录来源和加载时间。所有中心点的元数据存储于中心点版本表HUB_Vers中,主要有业务键Bus_key和中心点版本标识Hub_vers_id,如电网DV模式中的Hub_Customer、Hub_Contract和Hub_Service表中的业务键。这里的中心点版本标识Hub_vers_id主要反映其附属的属性变化形成的不同版本,然后,各个附属的元数据存储于附属版本表Sat_Vers中,如客户附属表Sat_Customer、客户地址附属表Sat_CustAddr和合同附属表Sat_Contract等,附属属性的元数据,经过附属要素表Sat_Eles表存储于属性表Attributes中,如姓名、电话和邮件等。

[0076] 对链接版本表Lnk_Vers,首先将各个链接表的链接标识Lnk_id、版本标识Lnk_vers_id等存储于链接版本表Lnk_Vers表;其次,利用业务键(如Bus_key中存储的具体业务键Cust_key和Cont_key等)通过链接版本表Lnk_Eles与中心点版本表HUB_Vers连接,保存相关中心点表的多个业务键等元数据。然后,如果该链接有附属表,可以将各个附属的元数据存储于附属版本表Sat_Vers中,附属属性的元数据,经过附属要素表Sat_Eles存储于属性表Attributes中。

[0077] 按照最终用户的多维分析任务,MD模式从事实表开始,经过维度表连接某些层次结构表,再到层节点表。关于事实版本的元数据存储于事实版本表FACT_Vers中,包括一个唯一的事实标识FV_id、事实版本标识FV_vers_id、事实名称、事实系统标识等。

[0078] 关于维度版本的元数据存储于维度版本表Dim_Vers中,包括一个唯一的维度标识DimV_id、事实标识FV_id、版本标识FV_vers_id,以及维度名称等。

[0079] 描述层次结构及其相关维度的元数据存于层次结构版本表Hier_Vers。层次结构版本由若干层节点版本构成。层节点版本的描述保存于层节点版本表Lev_Vers,包括层节点标识LV_id、版本标识FV_id和FV_vers_id、名字、层节点类型等。层次结构中多个层节点的联系信息保存于层节点要素表Hier_Eles,其中包含层节点的父子标识属性。

[0080] 属性元数据属于该元模型的公共部分,属性表Attributes可以由DV模式和MD模式共享。前面已明确了属性表Attributes可以保存DV模式的附属属性元数据。同样地,在MD模式中,每个事实版本和层节点版本都含有各自的属性集,即度量属性和维度属性,并存储于属性表Attributes。关于属性的完整性约束,经过完整性约束表Int_Cons,存储于属性约束表Att_Cons。完整性约束表Int_Cons保存完整性约束名、类型和定义。

[0081] 在元模型中,不考虑业务键Bus_key和属性Attribute的版本问题。

[0082] 以上分别说明了DV模式和MD模式的多版本元数据的存储情况。DV模式和MD模式之间也存在关联关系:①通过链接版本表Lnk_Vers与事实版本表FACT_Vers之间的关系,即链接表及其附属属性与事实表相关属性实际上存在对应的关联关系,该关系表明MD模式数中的某个事实表由DV模式中的0或1个链接表生成;②通过中心点版本表HUB_Vers与维度版本表Dim_Vers之间的关系,即中心点及其附属属性与维度表相关属性实际上存在对应的关联关系,该关系表明MD模式中的某个维度表由DV模式中的0或1个中心点表生成。从这两个虚线的非依赖关联关系,可以反映数据仓库和数据集市区域之间的元数据对应关系。

[0083] S3:进行电网数据仓库的元数据一致性检测,具体包括:

[0084] S31:进行DV模式的元数据一致性检测,如图4所示,具体包括中心点及其附属的一致性检测和链接及其附属的一致性检测;

[0085] 中心点及其附属的一致性检测步骤具体包括:

[0086] ①从全局标识VER_id开始,获得中心点版本表HUB_Vers中该全局标识VER_id的、当前中心点版本标识Hub_vers_id的全部业务键Bus_key,检测是否在数据仓库中都有对应的保存这些业务键的中心点表存在(电网DV数据仓库中有一张业务键与中心点表名、附属表名的引用表,利用该表可确定数据仓库中是否存在对应的中心点表),没有则输出错误,然后检测是否在数据仓库中都有对应的中心点表存在且该表的记录来源为Rec_Source(按前述方法找到中心点表,在该中心表对应的记录中有业务键和Rec_Source字段,应确认这两个记录来源相符),不相符则输出错误;

[0087] ②对①中获得的全部业务键Bus_key,检测与当前版本标识Hub_vers_id和业务键Bus_key外键相等的附属版本表Sat_Vers中存储的元数据记录,是否在数据仓库中都有对应的附属表存在(利用①中说明的引用表,判断方法相同),没有则输出错误;

[0088] ③若在数据仓库中存在对应的保存这些业务键的附属表,则获取与该附属版本表Sat_Vers记录对应的若干附属要素表Sat_Eles记录,检测每个附属要素表Sat_Eles记录对应的属性表Attributes中是否包含对应记录,若没有则输出错误;

[0089] 链接及其附属的一致性检测步骤具体包括:

[0090] ④从全局标识VER_id开始,获得链接版本表Lnk_Vers中该全局标识VER_id的、当前链接版本标识Lnk_vers_id的全部链接标识Lnk_id,检测是否在数据仓库中都有对应的链接表存在且该链接表记录来源为Rec_Source(电网DV数据仓库中有一张Lnk_id与链接表名、附属表名的引用表,利用该表可确定数据仓库中是否存在对应的链接表。另外电网DV数

据仓库的链接表中也保存有Rec_Source,应确认这两个记录来源相符),没有或记录来源不相符则输出错误;

[0091] ⑤使用④中的当前链接版本标识Lnk_vers_id的全部链接标识Lnk_id,经过链接要素表Lnk_Eles找到若干业务键Bus_key和中心点版本标识Hub_vers_id,判断是否都存于中心点版本表HUB_Vers中,没有则输出错误;

[0092] ⑥使用④中的当前链接版本标识Lnk_vers_id的全部链接标识Lnk_id,检测与这两个标识外键相关的附属版本表Sat_Vers中存储的元数据,是否在数据仓库中都有对应的附属表存在(利用④中说明的引用表,判断方法相同),没有则输出错误;

[0093] ⑦若在数据仓库中有对应的链接表存在,则获取与该附属版本表Sat_Vers记录对应的若干附属要素表Sat_Eles记录,并检测对应属性表Attributes中是否包含对应记录,若没有则输出错误。

[0094] S32:如图5所示,进行MD模式的元数据一致性检测:

[0095] ①使用上述相同全局标识VER_id开始,获得事实版本表FACT_Vers中该全局标识VER_id的、当前事实版本标识FV_vers_id的全部事实标识FV_id,检测是否在数据集市中有对应的事实表存在(利用DV数据仓库中的标识与表名的引用表,以下类似不再赘述),没有则输出错误;

[0096] ②对①中获得的全部事实标识FV_id,检测与当前事实版本标识FV_vers_id和事实标识FV_id外键相等的每个维度版本表DimV_id,是否在数据集市中有对应的维度表存在,没有则输出错误;

[0097] ③若在数据集市中有对应的维度表存在,则获取与维度版本表Dim_Vers对应的若干层次结构,即Hier_id,再获取与该Hier_id记录对应的若干层节点要素表Hier_Eles记录,首先检测其父子关系是否成立,若没有则输出错误;然后检测是否存在对应的层节点标识LV_id记录,若没有则输出错误;

[0098] ④对于③中的层节点标识LV_id记录,通过属性表Attributes中的层节点标识LV_id外键,检测对应属性表Attributes中是否包含对应记录,若没有则输出错误;

[0099] ⑤对于属性表Attributes中的事实版本标识FV_vers_id和事实标识FV_id外键,检测对应事实版本表FACT_Vers中是否包含对应记录,若没有则输出错误;再使用事实版本标识FV_vers_id和事实标识FV_id,检测在数据集市中有对应的事实表存在,没有则输出错误;

[0100] S33:如图6所示,进行属性表Attributes完整性约束的一致性检测;

[0101] 对于属性表Attributes中的每个属性标识attr_id,获取与该属性标识attr_id对应的若干完整性约束ic_id记录,检测完整性约束表Int_Cons中是否存在相关记录成立,若没有则输出错误。

[0102] S34:如图7所示,进行DV模式与MD模式的元数据一致性检测;

[0103] 本实施例主要检测在MD模式中建立的数据集市表,是否来源于DV模式中的相应数据表。考虑到构建MD模式时可能会应最终用户的要求,变更了部分模型,所以MD模式中可能存在与DV模式不对应的数据表元数据。

[0104] ①从全局标识VER_id开始,检测事实版本表FACT_Vers中当前版本每个记录的链接标识Lnk_id和版本标识Lnk_vers_id外键,是否在链接版本表Lnk_Vers存在一个对应的

记录,如果没有则输出提示可能出错信息;

[0105] ②从全局标识VER_id开始,利用事实版本表FACT_Vers中当前版本的每个事实标识FV_id,获取维度版本表Dim_Vers中对应的全部维度标识DimV_id记录集合,检测其每个记录中的业务键Bus_key和事实版本标识FV_vers_id外键,是否在中心点版本表HUB_Vers中存在对应的记录,如果没有则输出提示可能出错信息。

[0106] 通过上述检测步骤,实现元模型存储的元数据实施一致性检测。对元模型中三部分的基本表中存储的元数据,按当前版本逐个执行相互关联关系的检测,从而保证其中存储的多版本元数据与数据仓库和数据集市保存的具体数据存在一一对应,有效提升电网DV数据仓库环境的数据质量。

[0107] 在本实施例元模型中,不考虑为电网DV模型中的业务键Bus_key建立多版本,也不考虑为属性表Attributes建立多版本。

[0108] 本实施例为面向电网DV数据仓库环境的元模型和元数据一致性检测,解决构建元数据存储库的技术问题,为完善现有电网DV数据仓库环境的基础设施,建设数据质量的基本保障平台奠定基础。

[0109] (1) 元模型:给出多版本元数据的存储结构,解决电网DV数据仓库环境中构建元数据存储库的信息模型的技术问题;

[0110] (2) 元数据一致性检测算法:针对电网元数据存储库中存储的元数据质量问题,检查存储库的多个表中的元数据是否存在缺失、重复、冲突和不一致情况,解决已存入元数据质量保障的技术问题。

[0111] 实施例2:

[0112] 本实施例提供一种电网数据仓库多版本元数据存储及一致性检测系统,包括:电网DV数据仓库构建模块、元模型构建模块、电网数据仓库元数据一致性模块、完整性约束一致性检测模块、DV模式与MD模式元数据一致性检测模块;

[0113] 在本实施例中,电网DV数据仓库构建模块用于构建电网DV数据仓库,电网DV模式中采用中心点、链接、附属三类表分别存储电网业务实体、关系、中心点或链接的属性数据,电网MD模式中采用事实表、维度表分别存储电网客户的用电情况以及业务实体的属性;

[0114] 在本实施例中,元模型构建模块用于构建面向电网DV数据仓库环境的元模型,包括DV模式和MD模式的元数据存储表,用于存储对应的元数据,DV模式的元数据存储表包括:中心点版本表、链接版本表、链接要素表、附属版本表和附属要素表,MD模式的元数据存储表包括:事实版本表、维度版本表、层次结构版本表、层节点要素表和层节点版本表,还设有公共部分的基本表,包括:全局版本表、属性表、属性约束表和完整性约束表,公共部分的基本表由DV模式和MD模式共享;

[0115] 在本实施例中,电网数据仓库元数据一致性模块用于进行电网数据仓库的元数据一致性检测,包括DV模式的元数据一致性检测以及MD模式的元数据一致性检测,所述DV模式的元数据一致性检测包括中心点及其附属的一致性检测、链接及其附属的一致性检测;

[0116] 在本实施例中,完整性约束一致性检测模块用于进行属性表Attributes完整性约束的一致性检测,检测完整性约束表中是否存在相对应的完整性约束记录;

[0117] 在本实施例中,DV模式与MD模式元数据一致性检测模块用于进行DV模式与MD模式的元数据一致性检测,检测链接版本表中是否存在当前版本对应的记录,基于当前版本的

每个事实标识,获取维度版本表中对应的全部维度标识记录集合,检测在中心点版本表中是否存在对应的记录,输出一致性检测结果。

[0118] 上述实施例为本发明较佳的实施方式,但本发明的实施方式并不受上述实施例的限制,其他的任何未背离本发明的精神实质与原理下所作的改变、修饰、替代、组合、简化,均应为等效的置换方式,都包含在本发明的保护范围之内。

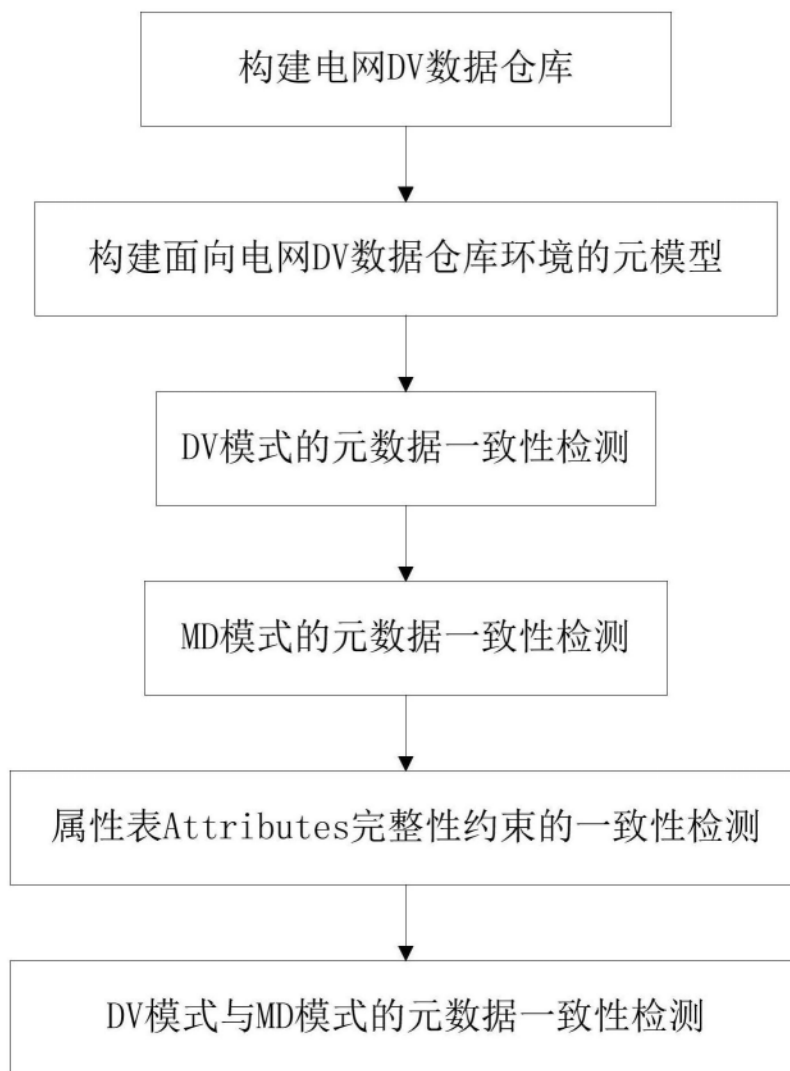


图1

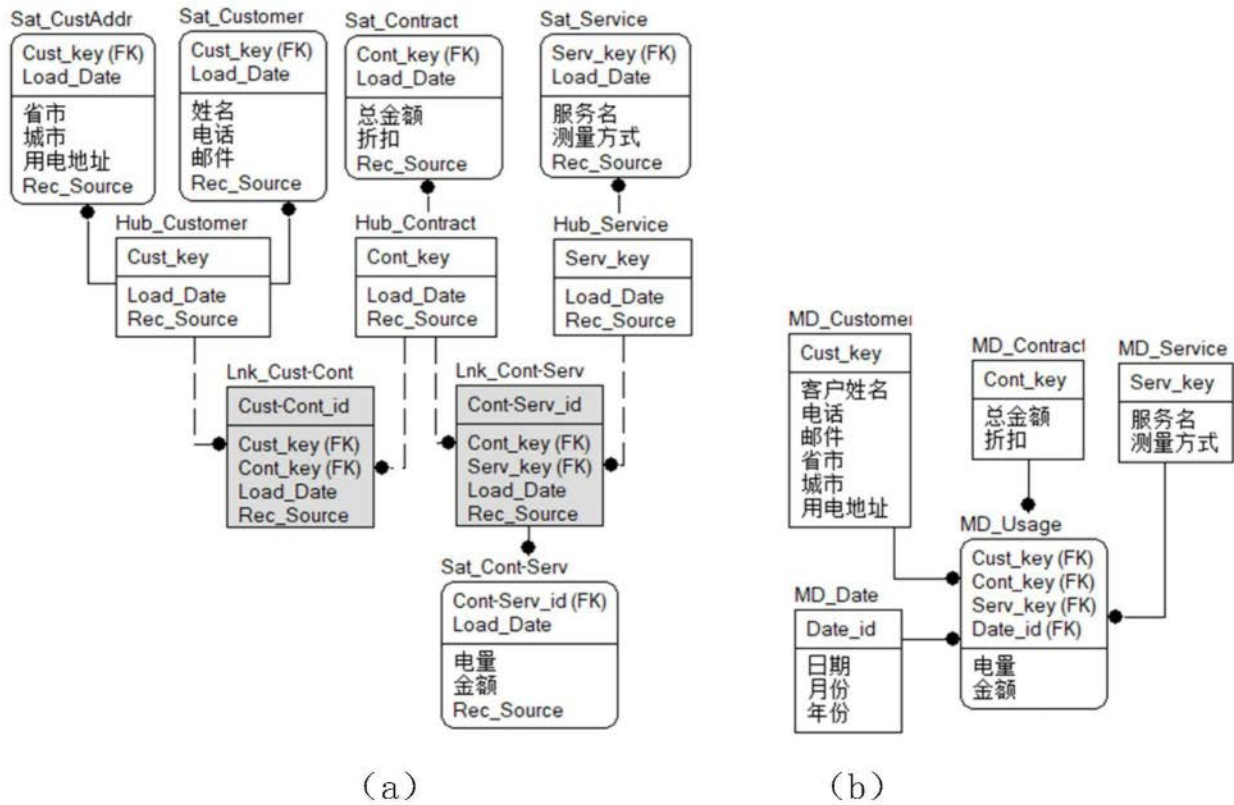


图2

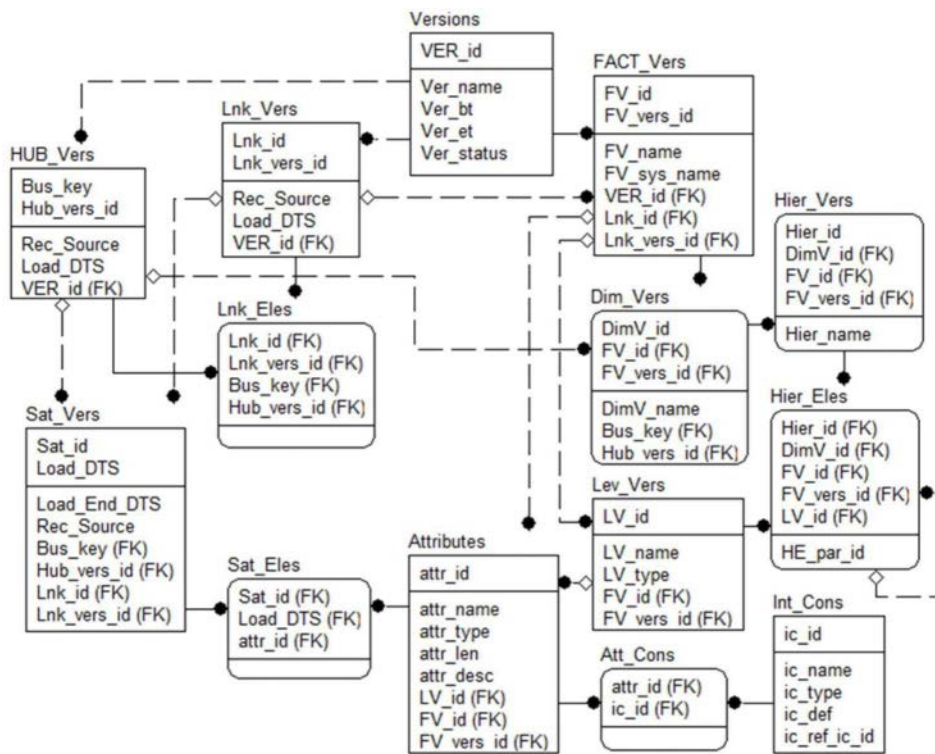


图3

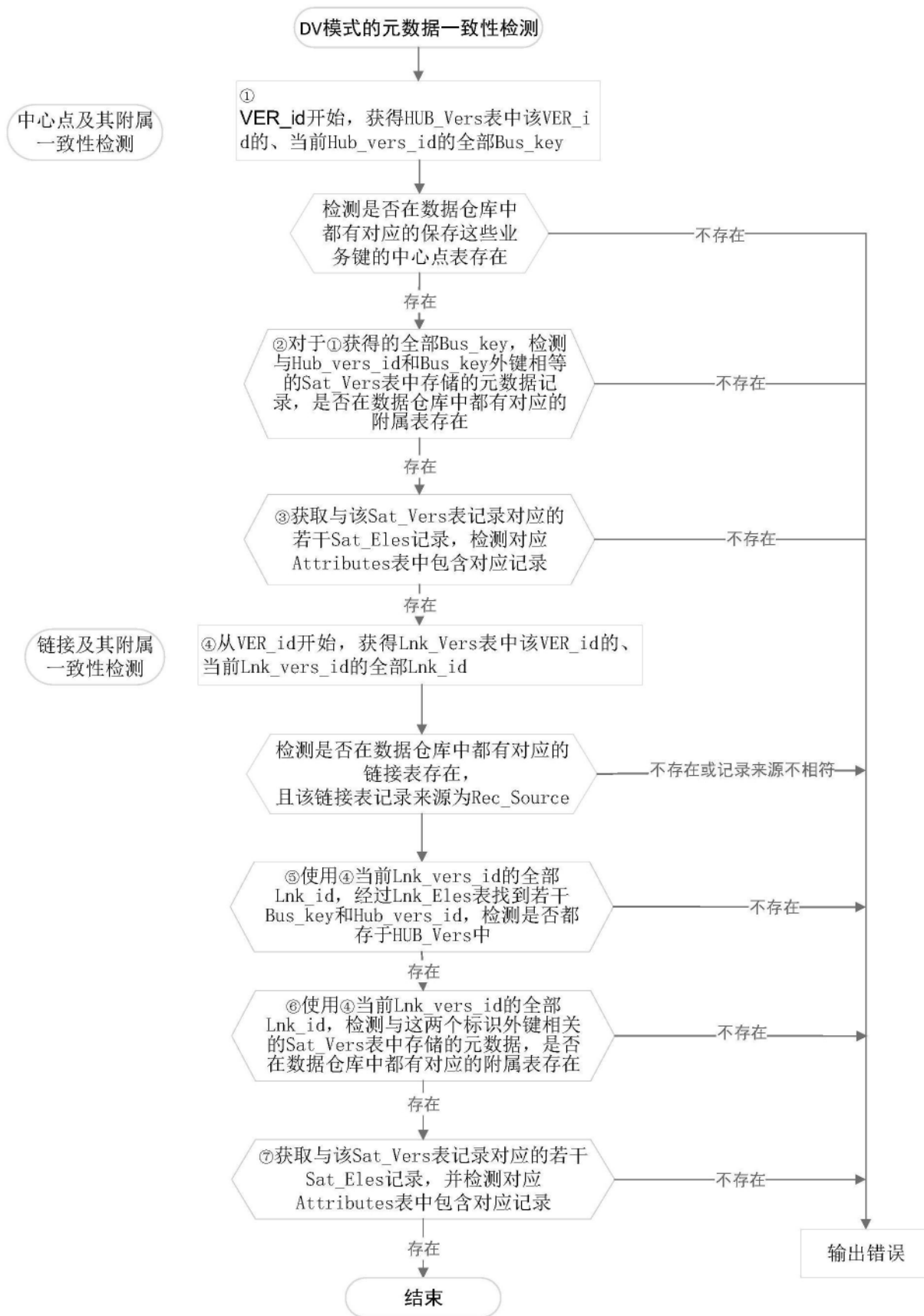


图4

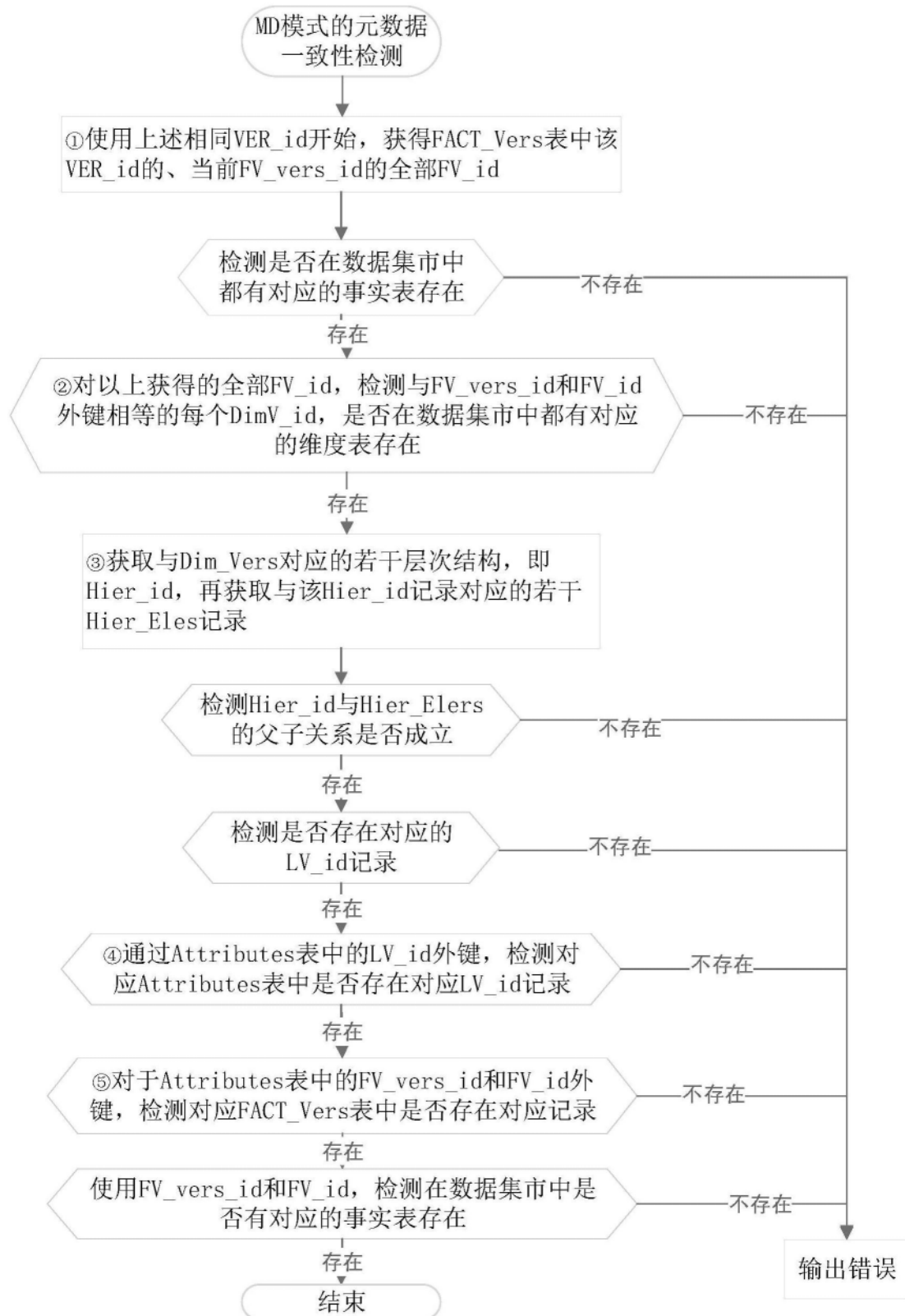


图5

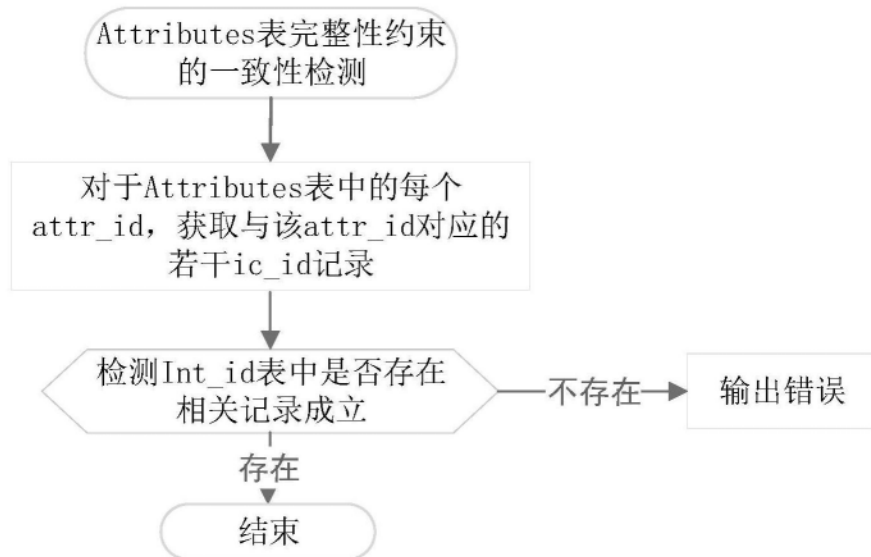


图6

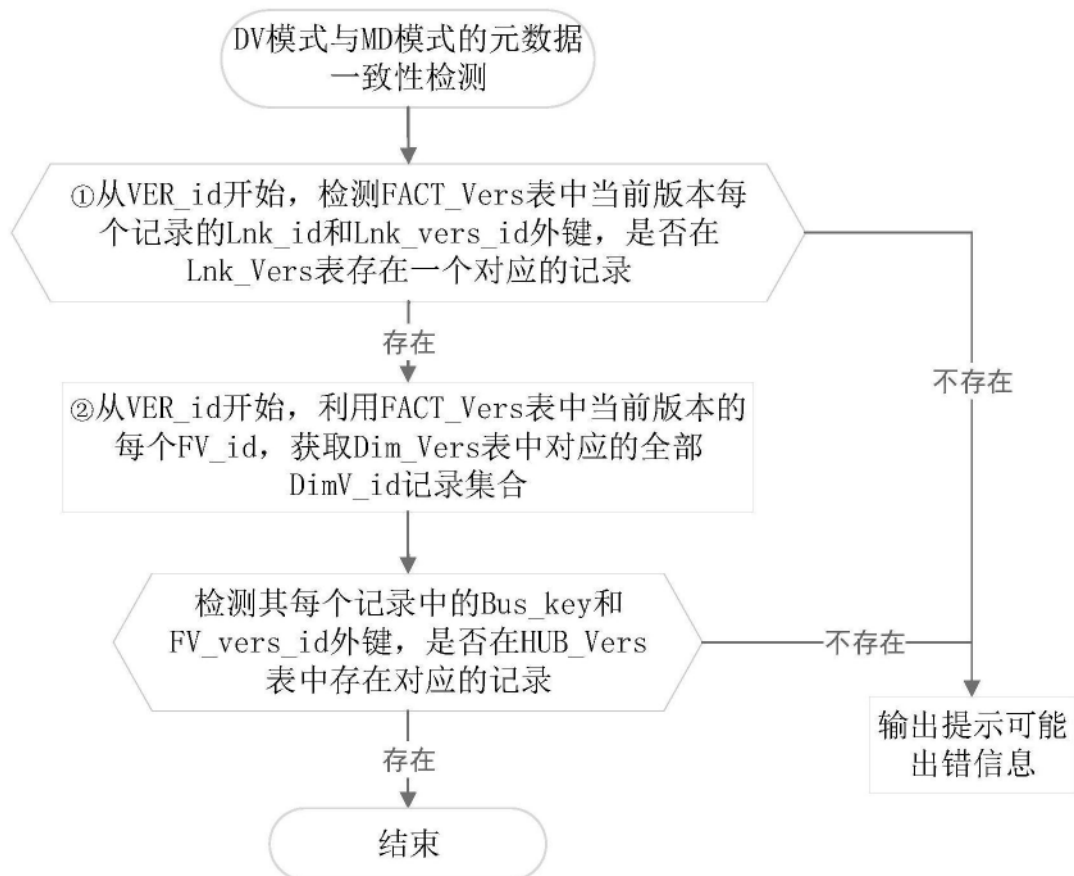


图7