



(12) 发明专利申请

(10) 申请公布号 CN 115687438 A

(43) 申请公布日 2023. 02. 03

(21) 申请号 202110823573.4

(22) 申请日 2021.07.21

(71) 申请人 中移(苏州)软件技术有限公司
地址 215163 江苏省苏州市高新区昆仑山路58号1幢中移软件园
申请人 中国移动通信集团有限公司

(72) 发明人 胡皓月 杨文峰 杨涛 尹腾飞

(74) 专利代理机构 北京派特恩知识产权代理有限公司 11270
专利代理师 刘恋 张颖玲

(51) Int. Cl.
G06F 16/2458 (2019.01)
G06F 18/2415 (2023.01)

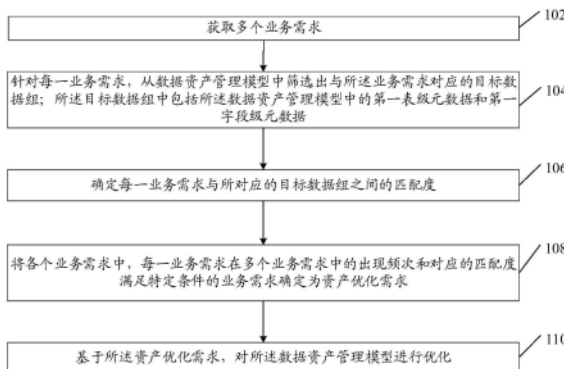
权利要求书4页 说明书23页 附图4页

(54) 发明名称

数据资产的运营方法、装置、电子设备和存储介质

(57) 摘要

本申请公开了一种数据资产的运营方法、装置、电子设备和存储介质,所述方法包括:获取多个业务需求;针对每一业务需求,从数据资产管理模型中筛选出与业务需求对应的目标数据组;目标数据组中包括数据资产管理模型中的第一表级元数据和第一字段级元数据;确定每一业务需求与所对应的目标数据组之间的匹配度;将各个业务需求中,每一业务需求在多个业务需求中的出现频次和对应的匹配度满足特定条件的业务需求确定为资产优化需求;基于资产优化需求,对数据资产管理模型进行优化。根据业务需求,从数据资产管理模型中筛选目标数据组,可以帮助需求者快速找到合适的数



1. 一种数据资产的运营方法,其特征在于,所述方法包括:

获取多个业务需求;

针对每一业务需求,从数据资产管理模型中筛选出与所述业务需求对应的目标数据组;所述目标数据组中包括所述数据资产管理模型中的第一表级元数据和第一字段级元数据;

确定每一业务需求与所对应的目标数据组之间的匹配度;

将各个业务需求中,每一业务需求在多个业务需求中的出现频次和对应的匹配度满足特定条件的业务需求确定为资产优化需求;

基于所述资产优化需求,对所述数据资产管理模型进行优化。

2. 根据权利要求1所述的方法,其特征在于,所述业务需求包括目标表级元数据和目标字段级元数据;所述针对每一业务需求,从数据资产管理模型中筛选出与所述业务需求对应的目标数据组,包括:

根据每一所述业务需求的目标表级元数据,从数据资产管理模型中筛选出第一候选数据表;

根据每一所述业务需求的目标字段级元数据,从所述第一候选数据表中筛选出第二候选数据表,并从所述第二候选数据表中筛选出候选元数据;

对所述第二候选数据表进行排列组合,得到多个候选数据组;

根据所述多个候选数据组中每一所述候选数据组的各第二候选数据表的数据时延和数据层级,从所述多个候选数据组中确定目标数据组。

3. 根据权利要求2所述的方法,其特征在于,所述目标表级元数据包括所述业务需求的目标所属层级、目标业务场景和目标数据时效;所述目标字段级元数据包括目标类别、目标范围和目标粒度;

所述根据每一所述业务需求的目标表级元数据,从所述数据资产管理模型中筛选出第一候选数据表,包括:根据每一所述业务需求的目标所属层级、目标业务场景和目标数据时效,从所述数据资产管理模型中筛选出第一候选数据表;

所述根据每一所述业务需求的目标字段级元数据,从所述第一候选数据表中筛选出第二候选数据表,并从所述第二候选数据表中筛选出候选元数据,包括:根据每一所述业务需求的目标类别、目标范围和目标粒度,从所述第一候选数据表中筛选出第二候选数据表,并从所述第二候选数据表中筛选出候选元数据;所述候选元数据为所述第二候选数据表中与所述目标类别、目标范围和目标粒度相关的元数据。

4. 根据权利要求2所述的方法,其特征在于,所述根据所述多个候选数据组中每一所述候选数据组的各第二候选数据表的数据时延和数据层级,从所述多个候选数据组中确定目标数据组,包括:

确定每一所述候选数据组中各第二候选数据表对应的数据层级;

根据每一所述数据层级中第二候选数据表的个数和对应数据层级的层级系数,确定第一推荐指数;

根据每一所述候选数据组中各第二候选数据表的数据时延,确定第二推荐指数;

根据每一所述第一推荐指数和对应的第二推荐指数,从所述多个候选数据组中确定目标数据组。

5. 根据权利要求4所述的方法,其特征在于,所述根据每一所述第一推荐指数和对应的第二推荐指数,从所述多个候选数据组中确定目标数据组,包括:

将所述第一推荐指数和所述第二推荐指数之和最小的候选数据组,确定为目标数据组;

或者,

根据每一所述第一推荐指数和对应的第二推荐指数,对所述多个候选数据组进行排序;根据排序结果,从所述多个候选数据组中确定多个第一数据组;响应于接收到的选取指令,从所述多个第一数据组中确定目标数据组。

6. 根据权利要求1至5中任一项所述的方法,其特征在于,所述将所述多个业务需求中出现频次和对应的匹配度满足特定条件的至少一个业务需求确定为资产优化需求,包括:

将所述多个业务需求中出现频次最高且匹配度最低的业务需求,确定为资产优化需求;

或者,

根据所述多个业务需求中每一所述业务需求的出现频次和对应的匹配度,确定对应业务需求的优先级;将优先级满足特定条件的业务需求确定为资产优化需求。

7. 根据权利要求6所述的方法,其特征在于,所述业务需求包括目标表级元数据和目标字段级元数据;

所述基于所述资产优化需求,对所述数据资产管理模型进行优化,包括:

基于所述资产优化需求,建立第一数据表;将所述第一数据表加入所述数据资产管理模型中;

或者,

基于所述资产优化需求的目标表级元数据和目标字段级元数据,对所述数据资产管理模型中的目标数据表的第一表级元数据和第一字段级元数据进行更新。

8. 根据权利要求1至5中任一项所述的方法,其特征在于,所述方法还包括:

根据终端信令数据的三要素,分别生成存储于数据仓库层、数据集市层和数据应用层中的数据表;所述三要素包括用户标识信息、用户位置信息和时间点;

根据所述终端信令数据的三要素,建立每一所述数据表对应的表级元数据和字段级元数据,其中,所述数据仓库层用于存放经过预处理的所述终端信令数据的元数据中的明细数据;所述明细数据以用户标识信息和时间点匹配用户位置信息,或者以用户位置信息和时间点匹配用户标识信息;所述数据集市层用于存放经过所述数据仓库层处理后的所述终端信令数据的元数据中的标签数据;所述标签数据按主题分为用户标识信息、用户位置信息和时间点;所述数据应用层用于存放经过所述数据仓库层或者所述数据集市层处理后的所述终端信令数据的元数据中的常用的中间计算过程表;

建立包括所述数据仓库层、所述数据集市层和所述数据应用层的所述数据资产管理模型。

9. 根据权利要求8所述的方法,其特征在于,所述根据终端信令数据的三要素,分别建立存储于数据仓库层、数据集市层和数据应用层中的数据表,包括:

在所述数据仓库层中,根据终端信令数据的用户标识信息和用户位置信息,建立多个基础实时数据表;根据所述终端信令数据的用户标识信息、用户位置信息和时间点,建立离

线事实数据表;根据所述基础实时数据表和所述离线事实数据表,建立与所述基础实时数据表、所述离线事实数据表相关的其他数据表;

在所述数据集市层中,对所述数据仓库层中的多个基础实时数据表、所述离线事实数据表和所述其他数据表中的数据进行汇总处理,建立包括多种标签数据的汇总标签数据表;

在所述数据应用层中,对所述数据仓库层中的多个基础实时数据表、所述离线事实数据表和所述其他数据表中的数据,和/或所述数据集市层中的汇总标签数据表进行处理,建立包含多种粒度数据的汇总粒度数据表。

10. 根据权利要求8所述的方法,其特征在于,所述表级元数据包括所属层级、业务场景、数据时效和数据时延;所述字段级元数据包括类别、范围和粒度;所述根据所述终端信令数据的三要素,建立每一所述数据表对应的表级元数据和字段级元数据,包括:

根据所述终端信令数据的用户标识信息的所属层级、用户位置信息的所属层级和时间点的所属层级,建立每一所述数据表的所属层级;所述所属层级为所述数据仓库层、所述数据集市层和所述数据应用层中的一种;

根据所述终端信令数据的用户标识信息、用户位置信息和时间点之间的检索关系,从预建立的各业务场景中选取至少一种业务场景作为每一所述数据表的业务场景;

根据所述终端信令数据的时间点的所属类别,建立每一所述数据表的数据时效;所述数据时效为离线表或实时表;

根据所述终端信令数据的时间点的最大值,建立每一所述数据表的数据时延;所述数据时延为从原始数据到对应数据表的平均处理时长;

根据所述终端信令数据的用户标识信息的所属类别、用户位置信息的所属类别和时间点的所属类别,建立每一所述数据表的人物类别、地点类别、时间类别和标签属性;

根据所述终端信令数据的用户标识信息的所属范围、用户位置信息的所属范围和时间点的所属范围,建立每一所述数据表的人物范围、地点范围、时间范围和标签值;

根据所述终端信令数据的用户标识信息的最小单元、用户位置信息的最小单元和时间点的最小单元,建立每一所述数据表的人物粒度、地点粒度和时间粒度。

11. 一种数据资产的运营装置,其特征在于,所述装置包括:

获取模块,用于获取多个业务需求;

筛选模块,用于针对每一业务需求,从数据资产管理模型中筛选出与所述业务需求对应的目标数据组;所述目标数据组中包括所述数据资产管理模型中的第一表级元数据和第一字段级元数据;

第一确定模块,用于确定每一业务需求与所对应的目标数据组之间的匹配度;

第二确定模块,用于将各个业务需求中,每一业务需求在多个业务需求中的出现频次和对应的匹配度满足特定条件的业务需求确定为资产优化需求;

优化模块,用于基于所述资产优化需求,对所述数据资产管理模型进行优化。

12. 一种电子设备,包括存储器和处理器,所述存储器存储有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程序时实现权利要求1至10任一项所述数据资产的运营方法中的步骤。

13. 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,该计算机程序被

处理器执行时实现权利要求1至10任一项所述数据资产的运营方法中的步骤。

数据资产的运营方法、装置、电子设备和存储介质

技术领域

[0001] 本申请涉及大数据技术领域,涉及但不限于一种数据资产的运营方法、装置、电子设备和存储介质。

背景技术

[0002] 移动通信的普及使公用移动通信基站的数量成倍增长,利用这些基站所进行的人类通信活动,产生了大量的终端信令数据,这些数据因关联了人和基站的位置信息,天然满足了“人-地-时”三要素特征,是一项重要的移动运营商数据资产,可以简称为移动位置数据。该数据资产相比于其他来源的时空位置数据,天然具备远距离覆盖、全天候采集的优势。但由于数据体量庞大,移动位置数据的价值并未被充分挖掘。

发明内容

[0003] 有鉴于此,本申请实施例提供一种数据资产的运营方法、装置、电子设备和存储介质。

[0004] 第一方面,本申请实施例提供一种数据资产的运营方法,所述方法包括:获取多个业务需求;针对每一业务需求,从数据资产管理模型中筛选出与所述业务需求对应的目标数据组;所述目标数据组中包括所述数据资产管理模型中的第一表级元数据和第一字段级元数据;确定每一业务需求与所对应的目标数据组之间的匹配度;将各个业务需求中,每一业务需求在多个业务需求中的出现频次和对应的匹配度满足特定条件的业务需求确定为资产优化需求;基于所述资产优化需求,对所述数据资产管理模型进行优化。

[0005] 第二方面,本申请实施例提供一种数据资产的运营装置,包括:获取模块,用于获取多个业务需求;筛选模块,用于针对每一业务需求,从数据资产管理模型中筛选出与所述业务需求对应的目标数据组;所述目标数据组中包括所述数据资产管理模型中的第一表级元数据和第一字段级元数据;第一确定模块,用于确定每一业务需求与所对应的目标数据组之间的匹配度;第二确定模块,用于将各个业务需求中,每一业务需求在多个业务需求中的出现频次和对应的匹配度满足特定条件的业务需求确定为资产优化需求;优化模块,用于基于所述资产优化需求,对所述数据资产管理模型进行优化。

[0006] 第三方面,本申请实施例提供一种电子设备,包括存储器和处理器,所述存储器存储有可在处理器上运行的计算机程序,所述处理器执行所述程序时实现本申请实施例任一所述数据资产的运营方法中的步骤。

[0007] 第四方面,本申请实施例提供一种计算机可读存储介质,其上存储有计算机程序,该计算机程序被处理器执行时实现本申请实施例任一所述数据资产的运营方法中的步骤。

[0008] 本申请实施例中,通过根据业务需求,可以从数据资产管理模型中筛选出目标数据组,可以帮助数据需求者能快速找到合适使用的数据;根据业务需求,可以动态地对数据资产管理模型进行优化,可以提升数据资产的价值。

附图说明

- [0009] 图1为本申请实施例一种数据资产的运营方法的流程示意图；
- [0010] 图2为本申请实施例一种数据资产管理模型的建立方法的流程示意图；
- [0011] 图3为本申请实施例一种数据资产管理模型的数据表建模的示意图；
- [0012] 图4为本申请实施例一种数据资产的循环运营管理方法的流程示意图；
- [0013] 图5为图4所示的循环运营管理方法的详细执行流程示意图；
- [0014] 图6为本申请实施例一种数据资产的运营装置的组成结构示意图；
- [0015] 图7为本申请实施例一种数据资产管理模型的建立装置的组成结构示意图；
- [0016] 图8为本申请实施例电子设备的一种硬件实体示意图。

具体实施方式

- [0017] 下面结合附图和实施例对本申请的技术方案进一步详细阐述。
- [0018] 图1为本申请实施例一种数据资产的运营方法的流程示意图,如图1所示,该方法包括:
- [0019] 步骤102:获取多个业务需求;
- [0020] 其中,业务需求又可以称为数据应用需求、数据需求,简称需求,业务需求可以是数据需求方对于数据的读写需求;举例说明,业务需求可以是求2020年每月西安市各大景点的日人流量月度峰值;业务需求还可以是西安市雁塔区某商场男性顾客一日内的偏好购物时段。
- [0021] 步骤104:针对每一业务需求,从数据资产管理模型中筛选出与所述业务需求对应的目标数据组;所述目标数据组中包括所述数据资产管理模型中的第一表级元数据和第一字段级元数据;
- [0022] 其中,所述数据资产管理模型可以是用于对数据资产进行管理和维护的数据模型,所述数据资产管理模型为数据库的基础;所述数据资产是指由个人或企业拥有或者控制的,能够为企业带来未来经济利益的,以物理或电子的方式记录的数据资源;所述数据资源可以是终端信令数据的元数据;所述数据资产管理模型用于通过数据表对终端信令数据的元数据进行管理;所述数据表可以是逻辑表。
- [0023] 所述数据资产管理模型可以包括数据仓库层、数据集市层和数据应用层;所述数据仓库层可以存放经过预处理、规整好的移动位置明细数据;所述数据集市层,可以存放由数据仓库层的数据加工处理而成的、经过汇总的,可供应用直接调用的标签数据;所述数据应用层可以存放由数据仓库层和/或数据集市层的数据加工处理而成的数据;所述数据应用层可以存放常用调用需求的中间计算过程表。
- [0024] 终端信令数据可以是终端用户与发射基站或者微站之间的通信数据;所述终端信令数据又可以称为移动位置数据;终端信令数据的元数据可以是用于描述所述终端信令数据的数据,主要用于描述终端信令数据的数据属性的信息,用来支持如指示存储位置、历史数据、资源查找、文件记录等功能;在做数据资产管理时,可涉及的终端信令数据的元数据的种类众多,如终端信令数据的技术元数据和业务元数据,所述数据资产管理模型可以通过数据表对终端信令数据的业务元数据进行管理。
- [0025] 所述数据仓库层、所述数据集市层和所述数据应用层中包括多个所述数据表;每

一所述数据表中包括表级元数据和字段级元数据。所述表级元数据可以包括所属层级、业务场景、数据时效和标准时延(又可以称为数据时延);所述字段级元数据可以包括类别、范围和粒度;其中,所述类别可以是三要素关键词(例如可以是人物、地点或时间)、标签属性、其他等;所述范围可以是区域范围、时间跨度(又可以称为时间范围)、人物范围(按号码归属地、某类属性划分)、标签值(例如:职业标签,包括学生、公务员、教师等标签值);所述粒度可以是地点粒度、时间粒度或人物粒度。

[0026] 步骤106:确定每一业务需求与所对应的目标数据组之间的匹配度;

[0027] 其中,参见步骤104,可以根据每一所述业务需求,从数据资产管理模型中筛选出目标数据组,因此,可确定每一业务需求与对应目标数据组之间的匹配度,并统计每一所述业务需求的出现频次。

[0028] 步骤108:将各个业务需求中,每一业务需求在多个业务需求中的出现频次和对应的匹配度满足特定条件的业务需求确定为资产优化需求;

[0029] 其中,需要说明的是,在一个实施例中,可以将出现频次和对应的匹配度满足特定条件的至少一个业务需求确定为资产优化需求。

[0030] 在另一个实施例中,还可以根据业务需求之间的相似程度,对各项业务需求进行分类,并确定每一类业务需求出现的频次(即同类业务需求的出现频次),并根据同类业务需求的出现频次和该同类业务需求与对应的目标数据组之间的匹配度,从所述多个业务需求中确定出至少一类业务需求,作为至少一个资产优化需求;每类业务需求中可包括至少一个业务需求;假设某条业务需求项为 D^i ,该条业务需求项的对应的目标数据组为 G_a , D^i 与 G_a 之间的匹配度为 M^i , D^i 的出现频次为 m , D^i 的同类需求为 F^i , F^i 的出现频次为 n ,则可以将 G_a 作为 F^i 对应的目标数据组,若 F^i 与 G_a 之间的匹配度为 N^i ,则可以确定该类业务需求的出现频次为 $m+n$,该类业务需求与对应的目标数据组之间的匹配度为 M^i+N^i ,则可以根据该类业务需求的出现频次和对应的匹配度,确定至少一个资产优化需求。

[0031] 步骤110:基于所述资产优化需求,对所述数据资产管理模型进行优化。

[0032] 本申请实施例中,通过根据业务需求,可以从数据资产管理模型中筛选出目标数据组,可以帮助数据需求者能快速找到合适使用的数据;根据业务需求可以确定资产优化需求,可以动态地对数据资产管理模型进行优化,可以提升数据资产的价值。

[0033] 本申请实施例还提供一种数据资产的运营方法,所述方法包括以下步骤:

[0034] 步骤S202:获取多个业务需求;

[0035] 其中,为了易于理解,可以列举三项需求实例进行阐述:业务需求1:求2019年每月苏州市虎丘区各大景点日人流量月度峰值;业务需求2:求国内某个不法分子正在逃跑的路径;业务需求3:求7月份北京市朝阳区某商场女性顾客一日内的偏好购物时段;所述业务需求包括目标表级元数据和目标字段级元数据。

[0036] 步骤S204:根据每一所述业务需求的目标表级元数据,从所述数据资产管理模型中筛选出第一候选数据表;

[0037] 其中,所述目标表级元数据包括所述业务需求所属的原子场景(即目标业务场景)和所属的时效性要求;所述原子场景又可以称为业务场景;所述时效性要求包括时效类别参数(即目标数据时效)和目标数据时延,数据时效是实时还是离线。可以根据业务需求中的原子场景和数据时效中的时效类别参数,查询所述数据资产管理模型中的各类表级元数

据,筛选出所有符合条件的数据表作为第一候选数据表。

[0038] 在一个实施例中,可以根据每一所述业务需求的目标所属层级、目标业务场景和目标数据时效,从所述数据资产管理模型中筛选出第一候选数据表。

[0039] 可以先根据业务需求,选择针对每项需求选择所需的原子场景。每项业务需求可以由一个或多个原子场景组合;所述原子场景可以包括:场景1:已知区域和时间范围,查找人(群);场景2:已知人(群)和时间范围,查询地点/区域;场景3:已知人(群)和区域范围,查询时间(时段专题报告);场景4:查找关于人的标签属性(性别、年龄、职业、常住地);场景5:查找关于区域(地点)的标签属性;场景6:查找关于时间段的“人地关系”标签属性。针对上述三项需求,分别选择的原子场景为:业务需求1对应场景1;业务需求2对应场景2;业务需求3对应场景1和场景4的组合。

[0040] 再根据业务需求,选择每项需求对数据的时效性要求,业务需求对数据的时效性要求包括两个方面:一方面,数据是实时处理即时出结果还是离线处理批量出结果的,即时效率类别是实时还是离线;另一方面,数据时延允许的最大值是多少;在一个实施例中,业务需求2需要实时数据,满足当前的正在进行的即时查询,对数据时延的期望尽可能小,最大允许10分钟;业务需求1和业务需求3主要针对历史数据进行统计,仅需调用离线数据进行计算;业务需求3仅关注一个月的数据,数据申请是一次性的,对于数据到达应用调用方所花费的时间,由两方人员在数据申请与授权对接时勤沟通即可把控。而针对业务需求1,显然需要每个月出一份上个月的人流量峰值数据统计报表,当一个月结束后,下个月的第几日能拿到这份报表反映了数据时延的要求,假设从管理的角度,希望月中前完成,即数据时延允许的最大值可能是15天。

[0041] 步骤S206:根据每一所述业务需求的目标字段级元数据,从所述第一候选数据表中筛选出第二候选数据表,并从所述第二候选数据表中筛选出候选元数据;

[0042] 其中,所述业务需求的目标字段级元数据为所述业务需求的需求参数,所述需求参数可以包括类别(又可以称为目标类别)、范围(又可以称为目标范围)和粒度(又可以称为目标粒度);所述候选元数据为所述第二候选数据表中与所述目标字段级元数据相关的元数据;可以根据业务需求中的各项需求参数,从第一候选数据表中匹配各个字段的元数据属性,筛选出符合需求参数的表,并将挑选出相关的数据元、删除无关数据元后的符合需求参数的表作为第二候选数据表。其中,在进行数据元匹配时,范围类型的元数据,需满足需求范围小于等于数据范围;需求粒度大于等于数据粒度。

[0043] 在一个实施例中,可以根据每一所述业务需求的目标类别、目标范围和目标粒度,从所述第一候选数据表中筛选出第二候选数据表,并从所述第二候选数据表中筛选出候选元数据;所述候选元数据为所述第二候选数据表中与所述目标类别、目标范围和目标粒度相关的元数据。

[0044] 在所述类别为地点的情况下,所述范围可以为区域范围,所述粒度可以为区域粒度;可以选择需求的最小区域单元作为区域粒度,例如:基站/小区、自定义区域、省、市、区/县;可以根据区域粒度选择需求涵盖的最大区域范围,例如:全国、省、市、区/县等。

[0045] 在所述类别为时间的情况下,所述范围可以为时间范围,所述粒度可以为时间粒度;可以选择需求的最小时间单元作为时间粒度;时间粒度根据数仓的实际表信息提供给使用者选择,可选的时间粒度例如:年、月、日、小时、15分钟;可以根据时间粒度选择需求涵

盖的最大时间段长度(即时间范围),例如:2020年1月至2020年9月。

[0046] 在所述类别为人物的情况下,所述范围可以为任务范围,所述粒度可以为人物粒度;可以选择需求的最小人物单元作为人物粒度,例如:人、团队、组织;可以根据人物粒度选择需求涵盖的人物范围,例如:所有江苏省移动用户。

[0047] 另外,针对某种标签需求(标签类的业务需求),可以在数据集市层中的标签库中选择需求匹配的标签类别,或补充填写标签库中没有的标签类别。所述标签需求对应的范围为标签范围,针对每一种标签类别,列举对应所需的标签值,例如:对于年龄段标签,需求涉及的标签范围仅限60-80岁的老人。表1是一张业务需求1至业务需求3的需求参数表,可配置的需求参数如表1所示:

[0048] 表1

业务需求	业务需求1	业务需求2	业务需求3
区域粒度	区/县	基站小区	基站小区
区域范围	虎丘区	全国	朝阳区某商场
时间粒度	日	小时	小时
时间范围	2019年全年	7*24小时	7月每天24小时
人物粒度	人	人	人
人物范围	不限	可能有重点人员名单	不限
标签类别	无	无	性别、群体
标签范围	无	无	女、商场顾客

[0050] 综上所述,根据业务需求的目标字段级元数据(即目标类别、目标范围和目标粒度),可以从所述第一候选数据表中筛选出第二候选数据表,并从所述第二候选数据表中筛选出候选元数据,将无关元数据从所述第二候选数据表中删除。

[0051] 步骤S208:对所述第二候选数据表进行排列组合,得到多个候选数据组;

[0052] 其中,可以对删除无关数据元后的所述第二候选数据表进行自由排列组合,得到多个候选数据组,每个候选数据组内的所有表级元数据和字段级元数据均可满足业务需求;一个第二候选数据表可以重复出现在多个候选数据组中。

[0053] 步骤S210:根据所述多个候选数据组中每一所述候选数据组的各第二候选数据表的数据时延和数据层级,从所述多个候选数据组中确定目标数据组;

[0054] 其中,所述数据层级可以是数据仓库层、数据集市层和数据应用层中的一种。

[0055] 步骤S212:确定每一业务需求与所对应的目标数据组之间的匹配度;

[0056] 其中,所述匹配度可以是所述业务需求中的目标字段级元数据,与目标数据组的各目标数据表中字段级元数据之间一致的字段级元数据的个数。

[0057] 步骤S214:将各个业务需求中,每一业务需求在多个业务需求中的出现频次和对应的匹配度满足特定条件的业务需求确定为资产优化需求;

[0058] 步骤S216:基于所述资产优化需求,对所述数据资产管理模型进行优化。

[0059] 本申请实施例中,由于终端信令数据来源于通信运营领域,其字段格式具有一定的专业独特性,数据的产生的频率间隔也富含很多现实意义。通过根据业务需求中的目标表级元数据进行数据表的筛选,再根据目标字段级元数据进行数据元的匹配,从而能够使得筛选出的目标数据组中的数据更准确、符合用户的数据需求。

[0060] 本申请实施例还提供一种数据资产的运营方法,所述方法包括以下步骤:

[0061] 步骤S302:获取多个业务需求;

[0062] 步骤S304:根据每一所述业务需求的目标表级元数据,从所述数据资产管理模型中筛选出第一候选数据表;

[0063] 步骤S306:根据每一所述业务需求的目标字段级元数据,从所述第一候选数据表中筛选出第二候选数据表,并从所述第二候选数据表中筛选出候选元数据;

[0064] 步骤S308:对所述第二候选数据表进行排列组合,得到多个候选数据组;

[0065] 其中,假设所述多个候选数据组中包括候选数据组 G_a , G_a 中包括第二候选数据表 $Table_1$ 至 $Table_5$,则 $G_a \in \{Table_1, Table_2, Table_3, Table_4, Table_5\}$ 。

[0066] 步骤S310:确定每一所述候选数据组中各第二候选数据表对应的数据层级;

[0067] 步骤S312:根据每一所述数据层级中第二候选数据表的个数和对应数据层级的层级系数,确定第一推荐指数;

[0068] 其中,假设 I_{a1} 表示第一推荐指数, $Table_{L_x}$ 表示候选数据组中层级系数为 L_x 的表;

$Num(Table_{L_x})$ 表示候选数据组中层级系数为 L_x 的表的个数,则第一推荐指数可以用公式

(1)表示:

$$[0069] \quad I_{a1} = \sum_1^3 L_x * (Num(Table_{L_x})) \quad (1);$$

[0070] 步骤S314:根据每一所述候选数据组中各第二候选数据表的数据时延,确定第二推荐指数;

[0071] 其中,假设 I_{a2} 表示第二推荐指数,Delay表示第二候选数据表的数据时延,则第二推荐指数可以用公式(2)表示:

$$[0072] \quad I_{a2} = \sum_1^i Delay \quad (2);$$

[0073] 步骤S316:根据每一所述第一推荐指数和对应的第二推荐指数,从所述多个候选数据组中确定目标数据组;

[0074] 其中,假设 I_a 表示第一推荐指数和对应的第二推荐指数之和,则 I_a 可以用公式(3)表示:

$$[0075] \quad I_a = I_{a1} + I_{a2} = \sum_1^3 L_x * (Num(Table_{L_x})) + \sum_1^i Delay \quad (3);$$

[0076] 步骤S318:确定每一业务需求与所对应的目标数据组之间的匹配度;

[0077] 步骤S320:将各个业务需求中,每一业务需求在多个业务需求中的出现频次和对应的匹配度满足特定条件的业务需求确定为资产优化需求;

[0078] 步骤S322:基于所述资产优化需求,对所述数据资产管理模型进行优化。

[0079] 本申请实施例中,根据每一所述第一推荐指数和对应的第二推荐指数,从所述多个候选数据组中确定目标数据组,能够使得目标数据组的确定更加准确。

[0080] 本申请实施例还提供一种数据资产的运营方法,所述方法包括以下步骤:

[0081] 步骤S402:获取多个业务需求;

[0082] 步骤S404:根据每一所述业务需求的目标表级元数据,从所述数据资产管理模型中筛选出第一候选数据表;

[0083] 步骤S406:根据每一所述业务需求的目标字段级元数据,从所述第一候选数据表中筛选出第二候选数据表,并从所述第二候选数据表中筛选出候选元数据;

[0084] 步骤S408:对所述第二候选数据表进行排列组合,得到多个候选数据组;

[0085] 步骤S410:确定每一所述候选数据组中各第二候选数据表对应的数据层级;

[0086] 步骤S412:根据每一所述数据层级中第二候选数据表的个数和对应数据层级的层级系数,确定第一推荐指数;

[0087] 步骤S414:根据每一所述候选数据组中各第二候选数据表的数据时延,确定第二推荐指数;

[0088] 步骤S416:将所述第一推荐指数和所述第二推荐指数之和最小的候选数据组,确定为目标数据组;

[0089] 其中,关于从候选数据组中选取目标数据组,可以考虑数据时延和数据层级两个因素;数据时延越短、数据所属数据层级越接近上层应用越优先推荐,因此,参见公式(3),可以将 I_a 最小的候选数据组,确定为目标数据组。

[0090] 假设候选数据组 $G_a \in \{Table_1, Table_2, Table_3, Table_4, Table_5\}$,且 $Table_1$ 和 $Table_2$ 为数据集市层的表, $Table_3$ 为数据仓库层的表, $Table_4$ 和 $Table_5$ 为数据应用层的表, $Table_1$ 至 $Table_5$ 的数据时延依次为 $Delay_1$ 至 $Delay_5$,即数据仓库层的表的个数为1,数据集市层和数据应用层的表的个数为2,则候选数据组 G_a 的推荐指数 I_a 可以用公式(4)表示:

[0091] $I_a = L_1 * 1 + L_2 * 2 + L_3 * 2 + Delay_1 + Delay_2 + Delay_3 + Delay_4 + Delay_5$ (4);

[0092] 则可以在 I_a 最小的情况下,将 G_a 选为目标推荐数据组。

[0093] 步骤S418:确定每一业务需求与所对应的目标数据组之间的匹配度;

[0094] 步骤S420:将各个业务需求中,每一业务需求在多个业务需求中的出现频次最高且匹配度最低的业务需求,确定为资产优化需求;

[0095] 其中,业务需求或者同类业务需求的出现频次越高,建设满足此类业务需求的中间过程表越迫切;每一项业务需求与对应的目标数据组之间的匹配度,匹配度差距越大,表明现有的MID(数据集市层)或APP层(数据应用层)中的表满足业务需求的程度偏低,因此,可以将出现频次最高且匹配度最低的业务需求,确定为资产优化需求。

[0096] 另外,由于多个业务需求的出现频次可能都不高,或者匹配度都不低,则即使是从多个业务需求中选择出现频次最高、且匹配度最低的业务需求,建设该业务需求的中间过程表的需求也不够迫切,且现有的MID(数据集市层)或APP层(数据应用层)中的表满足业务需求的程度偏高,因此,选出的资产优化需求对于优化数据资产管理模型的贡献并不大,因此,可以先确定所述多个业务需求中是否存在出现频次大于频次阈值,或匹配度小于匹配度阈值的至少一个业务需求,若存在,则将该至少一个业务需求确定为目标业务需求,并将至少一个所述目标业务需求中出现频次最高且匹配度最低的业务需求确定为资产优化需求。

[0097] 步骤S422:基于所述资产优化需求,建立对应的第一数据表;

[0098] 其中,可以根据所述资产优化需求中的目标表级元数据和目标字段级元数据,建立第一数据表。

[0099] 步骤S424:将第一数据表加入所述数据资产管理模型中。

[0100] 本申请实施例中,通过考虑数据时延和数据层级两个因素,优先推荐时延短并且

数据层级接近上层应用的候选数据组,一方面,可以尽可能推荐用户使用非明细数据(即数据集市层或数据应用层的数据),可以避免用户对于数据仓库层的敏感类明细数据的访问,保证数据安全;另一方面,可以提高数据的读取效率;另外,通过将出现频次最高且匹配度最低的业务需求,确定为资产优化需求,基于所述资产优化需求建立第一数据表加入数据资产管理模型中,可以提高有迫切需求的业务需求与数据资产管理模型中的数据之间的匹配度。

[0101] 本申请实施例还提供一种数据资产的运营方法,所述方法包括以下步骤:

[0102] 步骤S502:获取多个业务需求;

[0103] 步骤S504:根据每一所述业务需求的目标表级元数据,从所述数据资产管理模型中筛选出第一候选数据表;

[0104] 步骤S506:根据每一所述业务需求的目标字段级元数据,从所述第一候选数据表中筛选出第二候选数据表,并从所述第二候选数据表中筛选出候选元数据;

[0105] 步骤S508:对所述第二候选数据表进行排列组合,得到多个候选数据组;

[0106] 步骤S510:确定每一所述候选数据组中各第二候选数据表对应的数据层级;

[0107] 步骤S512:根据每一所述数据层级中第二候选数据表的个数和对应数据层级的层级系数,确定第一推荐指数;

[0108] 步骤S514:根据每一所述候选数据组中各第二候选数据表的数据时延,确定第二推荐指数;

[0109] 步骤S516:根据每一所述第一推荐指数和对应的第二推荐指数,对所述多个候选数据组进行排序;

[0110] 其中,可以根据第一推荐指数和第二推荐指数之和,对多个候选数据组进行排序,例如可以按照 I_a 从小到大排序。

[0111] 步骤S518:根据排序结果,从所述多个候选数据组中确定多个第一数据组;

[0112] 其中,可以从多个候选数据组中选择 I_a 排序靠前的几项候选数据组作为第一数据组。

[0113] 步骤S520:响应于接收到的选取指令,从所述多个第一数据组中确定目标数据组;

[0114] 其中,所述第一数据组可以供使用者参考选用;所述使用者可以对所述第一数据组进行目标数据组的选取操作,处理器在接收到选取操作后生成选取指令,从第一数据组中确定出目标数据组。

[0115] 步骤S522:确定每一业务需求与所对应的目标数据组之间的匹配度;

[0116] 步骤S524:根据每一业务需求在多个业务需求中的的出现频次和对应的匹配度,确定对应业务需求的优先级;

[0117] 其中,出现频次越高且对应的匹配度越低,对应业务需求的优先级可以越高。

[0118] 步骤S526:将优先级满足特定条件的业务需求确定为资产优化需求;

[0119] 其中,可以按照优先级由高到低的顺序,对业务需求进行排序,并将优先级靠前的几个业务需求确定为资产优化需求;另外,可以先按照优先级由高到低的顺序,对业务需求进行排序,并将优先级靠前的几个业务需求确定为目标业务需求,在所述目标业务需求的出现频次大于频次阈值,或匹配度小于匹配度阈值的情况下,将所述目标业务需求确定为资产优化需求。

[0120] 步骤S528:基于所述至少一个资产优化需求中每一所述资产优化需求的目标表级元数据和目标字段级元数据,对所述数据资产管理模型中的对应目标数据表的第一表级元数据和第一字段级元数据进行更新。

[0121] 本申请实施例中,还可以根据第一推荐指数和第二推荐指数对候选数据组进行排序,并将排序靠前的候选数据组供用户选择,从而能够提高用户选择目标数据组的自主性和灵活性;通过根据业务需求的出现频次和匹配度确定业务需求的优先级,并按照优先级顺序,确定出多个资产优化需求,并根据多个资产优化需求中的元数据,对数据资产管理模型进行更新,既提高了资产优化需求确定的多样性和灵活性,也提高了数据资产管理模型更新方式的灵活性。

[0122] 图2为本申请实施例一种数据资产管理模型的建立方法的流程示意图,如图2所示,该方法包括:

[0123] 步骤202:根据终端信令数据的三要素,分别建立存储于数据仓库层、数据集市层和数据应用层中的数据表;所述三要素包括用户标识信息、用户位置信息和时间点;

[0124] 其中,所述数据仓库层用于存放经过预处理的所述终端信令数据的元数据中的明细数据;所述明细数据以用户标识信息和时间点匹配用户位置信息,或者以用户位置信息和时间点匹配用户标识信息;所述数据集市层用于存放经过所述数据仓库层处理后的所述终端信令数据的元数据中的标签数据;所述标签数据按主题分为用户标识信息、用户位置信息和时间点;所述数据应用层用于存放经过所述数据仓库层或者所述数据集市层处理后的所述终端信令数据的元数据中的常用的中间计算过程表;所述数据表可以用于表征数据之间的逻辑关系,所述数据表可以是逻辑表。

[0125] 所述用户标识信息又可以称为人或人群,可以是各类用户ID(Identity document,身份标识号),所述用户位置信息又可以称为地点,可以是各类区域编码,如省、市和小区等的编码,所述时间点又可以称为时间,既可以是时刻,也可以是时间段,如节假日编码(例如春节对应的编码)、报表类型编码等。

[0126] 步骤204:根据所述终端信令数据的三要素,建立每一所述数据表对应的表级元数据和字段级元数据;

[0127] 步骤206:建立包括所述数据仓库层、所述数据集市层和所述数据应用层的所述数据资产管理模型。

[0128] 本申请实施例中,通过分层、分数据表建立数据资产管理模型,可以更高效地对终端信令数据的元数据进行管理和维护,由于建立的是逻辑表,体现的是数据之间的逻辑关系,因此还可以帮助数据需求者快速理解数据。

[0129] 本申请实施例还提供一种数据资产管理模型的建立方法,所述方法包括以下步骤:

[0130] 步骤S602:在在所述数据仓库层中,根据终端信令数据的用户标识信息和用户位置信息,建立多个基础实时数据表;根据所述终端信令数据的用户标识信息、用户位置信息和时间点,建立离线事实数据表;根据所述基础实时数据表和所述离线事实数据表,建立与所述基础实时数据表、所述离线事实数据表相关的其他数据表;

[0131] 步骤S604:在数据集市层中,对所述数据仓库层中的多个基础实时数据表、所述离线事实数据表和所述其他数据表中的数据进行汇总处理,建立包括多种标签数据的汇总标

签数据表；

[0132] 步骤S606:在数据应用层中,对所述数据仓库层中的多个基础实时数据表、所述离线事实数据表和所述其他数据表中的数据,和/或所述数据集市层中的汇总标签数据表进行处理,建立包含多种粒度数据的汇总粒度数据表;

[0133] 步骤S608:根据所述终端信令数据的三要素,建立每一所述数据表对应的表级元数据和字段级元数据;

[0134] 其中,所述表级元数据包括所属层级、业务场景、数据时效和数据时延;所述字段级元数据包括类别、范围和粒度。

[0135] 步骤S610:建立包括所述数据仓库层、所述数据集市层和所述数据应用层的所述数据资产管理模型。

[0136] 本申请实施例中,通过将汇总粒度数据表放入数据应用层,将汇总标签数据表放入数据集市层,将基础实时数据表和离线事实数据表放入数据仓库层,从而可以帮助数据需求者从数据应用层更方便地调取中间层数据,减少重复开发成本,缩短了数据处理时间。

[0137] 本申请实施例还提供一种数据资产管理模型的建立方法,所述方法包括以下步骤:

[0138] 步骤S702:根据所述终端信令数据的用户标识信息、用户位置信息和时间点之间的检索关系,从预建立的各项业务场景中选取至少一种业务场景作为每一所述数据表的业务场景;

[0139] 其中,所述检索关系可以是根据用户标识信息和时间点,来检索用户位置信息,还可以是根据用户位置信息和时间点,来检索用户标识信息等,可以分别在所述数据仓库层、所述数据集市层和所述数据应用层进行数据表的建模,在进行数据表建模之前,还可以进行业务场景的建模,业务场景建模可以根据“人-地-时”三要素,枚举出六类可能的数据应用原子场景(又可以称为业务场景),即步骤S204中的场景1至场景6。

[0140] 步骤S704:根据终端信令数据的三要素,分别建立存储于数据仓库层、数据集市层和数据应用层中的数据表;

[0141] 图3为本申请实施例一种数据资产管理模型的数据表建模的示意图;参见图3,在数据表建模中,所述数据仓库层301,考虑[人+时间:地点]和[地点+时间:人]常用场景的快速检索需求,分别以用户和位置做检索生成两张基础实时表:用户轨迹序列列表3011和基站(小区)切片表3012;数据集市层302主要存放以人或地点为键值的各类轻度汇总标签表;数据应用层303存放大多数应用需求可能调用的各种粒度的汇总表以方便于应用直接调用。

[0142] 步骤S706:根据所述终端信令数据的用户标识信息的所属层级、用户位置信息的所属层级和时间点的所属层级,建立每一所述数据表的所属层级;所述所属层级为所述数据仓库层、所述数据集市层和所述数据应用层中的一种;

[0143] 步骤S708:根据所述终端信令数据的用户标识信息、用户位置信息和时间点之间的检索关系,从所述各项业务场景中选取至少一种业务场景作为每一所述数据表的业务场景;

[0144] 其中,所述业务场景可以为场景1至场景6中的一种场景或多种场景的组合。

[0145] 步骤S710:根据所述终端信令数据的时间点的所属类别,建立每一所述数据表的数据时效;所述数据时效为离线表或实时表;

[0146] 其中,所述数据时效可以为离线表或者实时表;所述时间点的所属类别包括实时和离线,用于表示数据是实时处理即时出结果还是离线处理批量出结果的;所述数据时延可以为Delay函数记录从原始数据到当前表的平均处理时间(又可以称为平均处理时长)。

[0147] 步骤S712:根据所述终端信令数据的时间点的最大值,建立每一所述数据表的数据时延;所述数据时延为从原始数据到对应数据表的平均处理时长;所述时间点的最大值可以表示数据时延允许的最大值。

[0148] 步骤S714:根据所述终端信令数据的用户标识信息的所属类别、用户位置信息的所属类别和时间点的所属类别,建立每一所述数据表的人物类别、地点类别、时间类别和标签属性;

[0149] 其中,可以根据用户标识信息的所属类别,确定所述数据表的人物类别,根据用户位置信息的所属类别,确定所述数据表的地点类别,根据时间点的所属类别,确定所述数据表的时间类别。

[0150] 步骤S716:根据所述终端信令数据的用户标识信息的所属范围、用户位置信息的所属范围和时间点的所属范围,建立每一所述数据表的人物范围、地点范围、时间范围和标签值;

[0151] 其中,所述用户标识信息的所属范围可以是不限或者重点人员名单,用户位置信息的所属范围可以是全国,XX区,YY区某商场,所述时间点的所属范围可以是2019年全年,7天等。

[0152] 步骤S718:根据所述终端信令数据的用户标识信息的最小单元、用户位置信息的最小单元和时间点的最小单元,建立每一所述数据表的人物粒度、地点粒度和时间粒度。

[0153] 其中,所述人物粒度可以是人,地点粒度可以是基站小区,区/县等,所述时间粒度可以是日、小时等。

[0154] 需要说明的是,每个表的每个字段对应一项数据元,对于数据开放的权限粒度可以控制到字段级,因此每项字段的元数据信息对于数据使用者来说非常重要。针对移动位置数据资产的业务特征,规范字段的业务元数据(又可以称为字段级元数据)可以包括:类别、范围和粒度;其中,所述类别可以是三要素关键词(例如可以是人物、地点或时间)、标签属性、其他等;所述范围可以是区域范围、时间跨度(又可以称为时间范围)、人物范围(按号码归属地、某类属性划分)、标签值(例如:职业标签,包括学生、公务员、教师等标签值);所述粒度可以是地点粒度、时间粒度或人物粒度。表2是一张江苏省区域范围内2020年6月全月所有男性用户位置信令拉链表,可配置的字段级业务元数据如表2所示:

[0155] 表2

字段名	说明	类别	范围	粒度
IMSI	SIM 卡号	人物-关键词	男性用户	单一用户
MSISDN	手机号码	人物-关键词	男性用户	单一用户
From_country	国家码	人物-标签	男性用户	单一用户
From_Source	手机号码登记地, 有效手机号码前 7 位	人物-标签	男性用户	单一用户
Longitude	纬度	地点关键词	江苏省	基站(小区)
Latitude	经度	地点关键词	江苏省	基站(小区)
TAC	当前小区 TAC	地点关键词	江苏省	基站(小区)
Cell ID	当前小区的	地点关键词	江苏省	基站(小区)

[0156]

[0157]

	ECI			
geohash		地点关键词	江苏省	基站(小区)
ProvinceCode	省编码, 分区字段	地点-标签	江苏省	基站(小区)
City	基站所在城市	地点-标签	江苏省	基站(小区)
Timestamp	信令产生时间	时间关键词	2020 年 6 月	小时
Day	日, 分区字段	时间关键词	2020 年 6 月	小时
PROCEDURE TYPE/Event ID	过程类型	其他	---	---
Gen	信令溯源 2:2/3G, 4:4G	其他	---	---

[0158] 其中, 参见表 2, IMSI (International Mobile Subscriber Identity, 国际移动用户识别码) 是 SIM (Subscriber Identity Module, 用户身份识别卡) 卡号; MSISDN 是指主叫用户为呼叫 GSM (Global System for Mobile Communications, 全球移动通信系统) PLMN (Public Land Mobile Network, 公共陆地移动网) 中的一个移动用户所需拨的号码, 作用

同于固定网PSTN(Public Switched Telephone Network,公共交换电话网络)号码,是在公共电话交换网络编号计划中,唯一能识别移动用户的号码,即被叫用户的手机号码;From_country是国家码;From_Source是手机号码的登记地,是有效手机号码的前7位;Longitude表示经度;Latitude表示纬度;TAC(Tracking Area Code,跟踪区代码)是当前小区的TAC;Cell ID(Cell Identity)又可以称为小区识别码,即当前小区的ECI(E-UTRAN Cell Identifier,E-UTRAN小区唯一标识),通过识别网络中哪一个小区传输用户呼叫并将该信息翻译成纬度和经度来确定用户位置;geohash编码就是利用geohash算法,将经纬度编码,将二维变一维,给地址位置分区;ProvinceCode表示省编码;City表示基站所在城市;Timestamp表示信令产生的时间戳;Day表示日;PROCEDURE TYPE/Event ID(Event Identity,事件代码)表示过程类型或者事件代码;Gen表示信令溯源。

[0159] 步骤S720:建立包括所述数据仓库层、所述数据集市层和所述数据应用层的所述数据资产管理模型。

[0160] 本申请实施例中,通过根据终端信令数据的三要素,建立每一所述数据表对应的表级元数据和字段级元数据,从而可以更高效地对终端信令数据的元数据进行管理和维护。

[0161] 相关技术中,关于移动位置数据的管理方法中虽然提出了一些关于移动位置类数据管理的方法,但存在如下技术缺点:

[0162] 首先,上述方法局限于数据库层面的功能优化,而非数据资产层面的数据管理。

[0163] 其中,上述方法针对时序、时空数据库设计,属于数据库层面的功能优化,侧重于解决一项或一类应用需求问题,而非解决数据资产管理者对于数据资产在多场景应用使用的价值最大化问题。

[0164] 其次,针对移动位置数据的独特性,不能解决数据使用者理解数据的困难。

[0165] 由于移动位置数据来源于通信运营领域,其字段格式具有一定的专业独特性,数据的产生的频率间隔也富含很多现实意义。上述方法侧重于通用时空数据在物理层面进行存储优化,而缺乏数据内涵在逻辑层面的知识显现,不能有效帮助数据使用者进行数据理解,并快速进行开发使用。

[0166] 再者,上述方法无法解决数据资产管理中,大量中间层数据的设计和使用问题。

[0167] 上述方法不能解决数据资产管理者对于数据分层管理的设计困难问题;不能帮助数据资产管理者决策哪些中间过程数据以逻辑表形式存在,哪些需要落盘存储;不能帮助数据使用者选择哪一层哪一份数据能够有效合规满足业务需求并最小化数据延迟。

[0168] 最后,上述方法无法根据数据使用需求导向动态调整数据资产管理策略。

[0169] 上述方法不能帮助数据资产管理者根据上层应用对数据使用的需求,动态科学的调整数据管理策略,持续提升数据价值。

[0170] 针对上述问题,本申请实施例提出了一种基于动态反馈机制的移动位置数据资产运营方法。该方法建立了一套以移动位置数据为特色的数据资产管理模型和数据应用需求收集、评估与优化反馈模型(又可以称为需求评估与数据推荐模型),侧重于面向数据资产应用价值最大化的管理。移动位置数据资产运营方法贴合移动位置信令数据“人-地-时”三要素特征,在数据资产管理模型的建设中注重“建设-服务-反馈”良性循环,初次建设依赖位置信令的基础特征,资产运营过程中强调以业务需求为导向的服务,通过不断评估新业

务需求与资产数据的匹配度,不断反馈优化数据资产管理模型,逐步达成资产价值最大化的目标,同时服务更多数据使用方,达成社会利益与资产管理者利益双赢局面。

[0171] 本申请实施例提供一种基于动态反馈机制的移动位置数据资产运营方法。这种动态反馈机制的运营方法,包含数据资产管理模型与数据应用需求(又可以称为业务需求)之间相互影响的动态机制。在该机制下,数据资产管理模型能面向不断增加的数据应用需求不断优化建设,降低管理成本,提高数据利用价值;数据需求方也能从移动位置数据特性中快速了解数据,根据导向性的需求收集过程,自助体验到数据推荐的服务。

[0172] 本申请实施例中,面向移动位置类数据场景,通过建设有效的数据资产管理模型,自动评估数据需求并推荐最佳数据共享方案;通过众多数据需求的收集,不断优化资产管理模型,形成“建设-服务-反馈”的良性循环运营机制。

[0173] 图4为本申请实施例一种数据资产的循环运营管理方法的流程示意图;参见图4,该循环运营机制包括如下步骤401至步骤403,其通过符合移动位置数据特征的“人-地-时”业务元数据紧密联系,图5为图4所示的循环运营管理方法的详细流程示意图。

[0174] 步骤401:移动位置数据的数据资产管理模型的建设;

[0175] 其中,参见图5,所述步骤401可以包括步骤5011至步骤5013:

[0176] 步骤5011:数据资产分层构建。

[0177] 其中,根据移动位置数据“人-地-时”三要素特性,构建三层数据资产管理模型:数据仓库层、数据集市层和数据应用层;设定数据仓库层存放经过预处理、规整好的移动位置明细数据,以两大基础数据结构[人+时间:地点]和[地点+时间:人],作为数据仓库层的主要组成部分,[人+时间:地点]表示以人和时间作为主键,可以得到地点,即通过人和时间去匹配地点,[地点+时间:人]表示以地点和时间作为主键,可以得到人,即通过地点和时间去匹配人;设定数据集市层的数据,由数据仓库层的数据进行加工处理生成,属于经过汇总的,可供应用直接调用的标签数据,分为人、地、时三类主题。设定数据应用层的数据,由数据仓库层和/或数据集市层的数据进行加工处理生成,数据应用层存放常用调用需求的中间计算过程表。

[0178] 对于数据仓库层,可以通过枚举方法,选取移动位置数据的三要素中任意两个要素作为主键,都可以得到另一个要素,但由于[人+地点:时间]通过人和地点去匹配时间,该种方式存储在检索和数据压缩性上都远不如另两种模型,且应用场景较少,不适合做基础数据,顾不建议放在数据仓库层。综上,所有关于[人+时间:地点],[地点+时间:人]的基础数据,无论采用何种数据库存储,都可以放在数据仓库层,所述数据库可以是Hive、HBase或Redis(Remote Dictionary Server,远程字典服务)等。

[0179] 对于数据集市层,以人为主题的集市数据可以为[人编号:人属性],其中,“人编号”为关键字段,如各类用户ID;人属性可以为多个标签字段,如性别、年龄、职业、常住地等;以地点/区域为主题的集市数据可以为[地点/区域编号:地点/区域属性],其中“地点/区域编号”为关键字段,如各类区域编码;地点/区域属性可以为多个标签字段,如舒适度指数、老年活动区、路况指数等;以时间为主题的集市数据可以为[时间段:时间段属性],多用于各类报表场景,其中“时间段”为关键字段,如节假日编码(例如春节对应的编码)、报表类型编码等,时间段属性可以为多个标签字段,如人口迁徙指数、人口区域统计。

[0180] 对于数据应用层,数据应用层根据数据资产开放的丰富程度是从无到有,不断丰

富的过程,数据应用层建设起初是为满足某些数据应用的调用而留存的一些中间过程临时表。随着数据资产的使用方逐步增多,容易发现很多需求处理的逻辑趋向一致,部分需求甚至无需调用基础数据而直接使用已经存在的中间过程表,其产生的开发和调用成本都会小很多。至此,考虑将部分中间过程临时表也开放在数据目录中供给应用使用,并将其归属在数据应用层。

[0181] 步骤5012:数据模型框架构建。

[0182] 其中,数据模型框架的构建可以包括业务场景建模和数据表建模。业务场景建模可以根据“人-地-时”三要素,枚举出六类可能的数据应用原子场景(又可以称为业务场景)。在数据表建模中,考虑[人+时间:地点]和[地点+时间:人]常用场景的快速检索需求,数据仓库层分别以用户和位置做检索生成两张基础实时表:用户轨迹序列列表和基站(小区)切片表;数据集市层主要存放以人或地点为键值的各类轻度汇总标签表;数据应用层存放大多数应用需求可能调用的各种粒度的汇总表以方便于应用直接调用。

[0183] 对于业务场景建模而言,所述业务场景可以包括:场景1:已知区域和时间范围,查找人(群);场景2:已知人(群)和时间范围,查询地点/区域;场景3:已知人(群)和区域范围,查询时间(时段专题报告);场景4:查找关于人的标签属性(性别、年龄、职业、常住地);场景5:查找关于区域(地点)的标签属性;场景6:查找关于时间段的“人地关系”标签属性。

[0184] 任意一项需求可以由一个或多个原子场景组合而满足。例如,想要知道某个商圈每小时的人流数量,场景1即可满足;想要知道某个商圈每小时青年人口的人流数量,就需要引用场景1和场景4组合。

[0185] 对于数据表建模而言,根据数据分层设计,数据仓库层包括实时事实表、离线事实表和相关维表。其中,为满足实时需求对于三元素的高效检索,分别以用户和位置做检索设计了两张基础实时表:用户轨迹序列列表和基站(小区)切片表,满足[人+时间:地点]和[地点+时间:人]场景需求的快速检索。离线事实表将[人+时间+地点]均设置检索,同时满足场景1、2、3需求。由于从原始数据进行重复数据去重和基站乒乓切换等预处理的时间窗口大小不同,离线事实表相对于实时事实表存储空间更小,可以存储更长时间的数据,参见图3,离线事实表可以是位置拉链表3013;所述相关维表可以是与所述实时事实表和所述离线事实表相关的基站工参维表3014和自定义区域基站维表315。

[0186] 数据集市层主要存放以人或地点为键值的各类轻度汇总标签表。从数据仓库层基础数据经过模型计算或关联其他类别的数据源,可得到规整的数据集市层。每张轻度汇总标签表中,都可能涉及到1个或多个属性标签。因此,数据集市层也是汇聚各类标签数据的集市。

[0187] 数据应用层存放大多数应用需求可能调用的各种粒度的汇总表以方便于应用直接调用,例如从区域层面汇总或从时间粒度上进行汇总。

[0188] 步骤5013:元数据规范构建。

[0189] 其中,元数据规范包括表级元数据模型和字段级数据元模型。针对数据模型框架中的数据表,规范并配置每个数据表的元数据项,同时规范数据元字段的业务元数据。

[0190] 需要说明的是,规范表级元数据可以包括:所属层级、业务场景、数据时效和标准时延。

[0191] 步骤402:数据需求收集与评估;

[0192] 其中,为满足最大化数据服务的愿景,需要数据管理者能准确收集各项数据需求;基于已建立的数据资产管理模型,首先收集数据需求,评估需求后给出推荐的数据。

[0193] 其中,所述步骤402可以以下步骤5021至步骤5024:

[0194] 步骤5021:选择原子场景;

[0195] 其中,可以收集业务需求所属的原子场景;

[0196] 步骤5022:判断数据时效;

[0197] 其中,可以收集业务需求所属的时效性要求;

[0198] 步骤5023:配置需求参数;

[0199] 其中,可以收集业务需求的需求参数;

[0200] 步骤5024:需求评估与数据推荐;

[0201] 其中,可以采用一种评估推荐模型,将数据需求作为输入,通过数据表筛选、数据元匹配,自动生成候选数据组,最后考虑数据时延和数据层级两个因素,给出组合排序推荐的候选开放数据(又可以称为目标数据组)。

[0202] 其中,所述步骤5024可以包括步骤50241至步骤50244:

[0203] 步骤50241:数据表筛选;

[0204] 步骤50242:数据元匹配;

[0205] 步骤50243:生成候选数据组;

[0206] 其中,可以将步骤50243中筛选出的表进行自由排列组合生成候选数据组,每个候选数据组内的所有表和字段均可满足目标需求;一个表可以重复出现在多个候选数据组中,枚举出所有可能的候选数据组,记为 G_n ,其中 n 为候选数据组的数量,对于 $a \in [1, n]$,每个候选数据组可表示为 $G_a \in \{Table_1, Table_2, Table_3, Table_4, Table_5\}$ 。至此,任意一个候选数据组提供给数据使用者均可满足目标需求,但推荐哪个还需考量。

[0207] 步骤50244:选取目标推荐数据组;

[0208] 其中,关于从候选数据组中选取目标推荐数据组,可以考虑数据时延和数据层级两个因素;数据时延越短、数据所属数据层级越接近上层应用越优先推荐。因此为PDW(数据仓库层)、MID(数据集市层)、APP(数据应用层)三个数据层级配置层级系数分别为 L_1 、 L_2 、 L_3 ,对于 G_a ,其推荐指数 I_a 为候选数据组内所有表的层级系数和数据时延之和。推荐指数 I_a 可以用公式(3)表示:

$$[0209] \quad I_a = I_{a1} + I_{a2} = \sum_1^3 L_x * (Num(Table_{L_x})) + \sum_1^i Delay \quad (3);$$

[0210] 其中, $Table_{L_x}$ 表示候选数据组中层级系数为 L_x 的表; $Num(Table_{L_x})$ 表示候选数据组中层级系数为 L_x 的表的个数, $\sum_1^3 L_x * (Num(Table_{L_x}))$ 表示候选数据组中三个数据层级中每一数据层级的表的个数乘以对应数据层级的层级系数得到的乘积的累加和; $\sum_1^i Delay$ 表示候选数据组中所有表的数据时延的累加和。

[0211] 假设候选数据组 $G_a \in \{Table_1, Table_2, Table_3, Table_4, Table_5\}$,且 $Table_1$ 和 $Table_2$ 为数据集市层的表, $Table_3$ 为数据仓库层的表, $Table_4$ 和 $Table_5$ 为数据应用层的表, $Table_1$ 至 $Table_5$ 的数据时延依次为 $Delay_1$ 至 $Delay_5$,即数据仓库层的表的个数为1,数据集市层和数据应用层的表的个数为2,则候选数据组 G_a 的推荐指数 I_a 可以用公式(4)表示:

[0212] $I_a = L_1 * 1 + L_2 * 2 + L_3 * 2 + Delay_1 + Delay_2 + Delay_3 + Delay_4 + Delay_5$ (4);

[0213] 其中,可以在 I_a 最小的情况下,将 G_a 选为目标推荐数据组;也可以将 I_a 进行从小到大排序,选择前几项为推荐数据组,供使用者参考选用,由使用者从推荐数据组中选取目标推荐数据组;为了尽可能推荐用户使用非明细数据,一般可以将 L_1 系数值调整较大,若仍需要申请 L_1 层数据,可以启动敏感数据安全审批机制。

[0214] 步骤403:反馈与数据资产管理模型优化;

[0215] 其中,数据管理者可以以数据资产的角度去管理数据,考虑到数据将面向众多业务需求而非一项。所述步骤403可以包括如下步骤5031至步骤5033:

[0216] 步骤5031:同类需求匹配度统计;

[0217] 其中,可以收集各项需求出现的频次,进行需求与资产匹配度统计。可以根据业务需求的需求参数的相似程度,对各项业务需求进行分类,并确定每一类业务需求出现的频次(即出现频次),同类业务需求的出现频次越高,建设满足此类业务需求的中间过程表越迫切;还可以确定每一项业务需求与对应的目标推荐数据组之间的匹配度,匹配度差距越大,表明现有的MID(数据集市层)或APP层(数据应用层)中的表满足业务需求的程度偏低。

[0218] 假设某条业务需求项为 D^i ,选取的目标推荐数据组为 G_a ,则可以遍历 G_a 中各表中关键数据元三要素的粒度和范围属性,和业务需求项进行对比,若有一项一致则计数1,最终将总计数 M^i 作为 D^i 与 G_a 的匹配度。当一条新的业务需求项与业务需求项为 D^i 相同或相似时,则认为两个业务需求项为同类业务需求项,也将新的业务需求项与 G_a 的匹配度累加计数在 M^i 中。

[0219] 步骤5032:根据业务需求匹配度和出现频次,排序资产优化项;

[0220] 其中,当统计到某类业务需求同时满足与现有MID(数据集市层)或APP层(数据应用层)中表的匹配度低且业务需求出现频次高,则需考虑在MID或APP层中优化已有的数据表,或补充建设直接满足该类业务需求的表。在日常数据资产建设过程中,可参考根据 M^i 的从大到小的优先级进行MID和APP层数据表的建设,建设内容包括优化MID和APP层中已有的表,更新数据元参数(如表级元数据或者字段级元数据),观察总体需求匹配度情况是否改善;或优化数据表模型,补充引入新表建设,直至数据资产达到满足所有数据需求调取的最优状态。

[0221] 步骤5033:综合考虑资产调整方案与影响。

[0222] 其中,在进行数据资产管理模型优化建设前,不仅需要评估优化调整方案对整个数据资产管理模型的影响,还需评估每一项优化调整将对所需投入的机器资源和相关配套设施的影响。

[0223] 每新增或变更一项数据表设计,都可通过表或记录级的血缘分析进行影响范围推演,考虑每个生产线上的数据都有可能影响在线业务,在变更前需先通过元数据的血缘分析,评估数据变更范围,对可能影响在线业务的变更需谨慎,或选择合适的时机进行调整。

[0224] 评估数据资产设计的调整对机器资源和相关配套设施的影响,一般需借助智能运维监控系统,对数据资产日常的资源消耗等监控数据进行统计分析,大致估算出标准定量数据规模的计算和存储空间。基于此,推算出对于数据管理的调整,可能带来的资源扩容或扩容。新增资源扩容的项目需做出合理的预算申请,对于扩容数量较少,但调整后效益增加显著的项目容易得到企业财务的支持。

[0225] 本申请实施例中,通过根据业务需求匹配度和出现频次排序资产优化项,可以实现最大化满足的需求范围、最便捷的数据定位和使用、最贴合的数据元匹配、最小的数据延迟和最高的数据质量等。

[0226] 在不断的需求收集与反馈的过程中,数据资产管理模型的建设将逐步发挥最大价值,服务于资产管理方与更多的数据使用者。

[0227] 本申请实施例提供一种符合移动位置数据特色的数据分层设计、数据模型框架和元数据规范组合而成的数据资产管理模型。

[0228] 其中,数据分层设计的关键包括:以两大基础数据结构[人+时间:地点]、[地点+时间:人],作为数据仓库层的主要组成部分;以移动位置数据“人-地-时”三要素特性,构建的围绕人、地、时三类主题标签数据作为数据集市层;存放数据应用层设计用来存放常用调用需求留存的中间计算过程表。

[0229] 数据模型框架的关键包括:以移动位置类数据“人-地-时”三要素,设计的六类数据应用原子场景作为业务场景模型。数据表模型以用户轨迹序列列表和基站(小区)切片表作为中心;周边围绕以人或地点为键值的各类轻度汇总标签表;外围存放大多数应用需求可能调用的各种粒度的汇总表以方便于应用直接调用。

[0230] 元数据规范的关键包括:表级元数据配置项设置所属层级、业务场景、数据时效和标准时延。字段级元数据配置项包括类别(人物、地点、时间三要素关键词,标签属性,其他)、范围(区域范围、时间跨度、人物范围)、粒度(地点粒度、时间粒度、人物粒度)。

[0231] 本申请实施例提供一种符合动态反馈机制的需求收集与评估模型。其中,通过业务需求的原子场景和数据时效中的时效类别参数可以快速缩小数据推荐范围,通过数据元匹配筛选满足条件的表组合生成候选数据组,最后依据需求数据时效和数据层级为候选数据组进行排序。

[0232] 本申请实施例提供一种针对移动位置数据的资产管理模型优化方法。其中,将每一条需求参数与关键数据元三要素的粒度和范围属性进行匹配,建立数据使用需求(即业务需求)与资产管理现状的匹配度对照表,针对匹配度低和出现频率高的某一类需求,综合考虑调整方案和影响后,可针对这类业务需求对数据资产管理模型进行优先优化。

[0233] 需要说明的是,本申请实施例是对数据资产层面的数据管理,而非数据库层面的功能优化。本申请实施例可以解决数据资产管理者对于数据资产在多场景应用使用的价值最大化问题,而非侧重于解决一项或一类应用需求问题。

[0234] 本申请实施例可以针对移动位置数据的独特性,帮助数据需求者高效理解、使用数据。由于移动位置数据来源于通信运营领域,其字段格式具有一定的专业独特性,数据的产生的频率间隔也富含很多现实意义。通过数据资产管理模型对于元数据的设计帮助数据需求者能快速理解数据;提供数据需求收集和评估,根据需求自动推荐候选数据,帮助数据需求者能快速找到合适使用的数据。

[0235] 本申请实施例中可以通过中间层数据留存价值显现,数据价值转化率提升。本申请实施例解决了数据资产管理者对于数据分层管理的设计困难问题;可以帮助数据资产管理决策哪些中间过程数据以逻辑表形式存在,哪些需要落盘存储;可以帮助同类数据需求者有直接可调取的公共中间层数据,减少重复开发成本,缩短了数据处理时间。

[0236] 本申请实施例可以根据数据使用需求导向动态调整数据资产管理策略。本申请实

施例通过“建设-服务-反馈”的良性循环运营机制,帮助数据资产管理者根据上层应用对数据使用的需求,动态科学的调整数据管理策略,持续提升数据价值。

[0237] 基于前述的实施例,本申请实施例提供一种数据资产的运营装置,该装置包括所包括的各模块,可以通过电子设备中的处理器来实现;当然也可通过具体的逻辑电路实现;在实施的过程中,处理器可以为中央处理器(CPU, Central Processing Unit)、微处理器(MPU, Microprocessor Unit)、数字信号处理器(DSP, Digital Signal Processing)或现场可编程门阵列(FPGA, Field Programmable Gate Array)等。

[0238] 图6为本申请实施例数据资产的运营装置的组成结构示意图,如图6所示,所述装置600包括获取模块601、筛选模块602、第一确定模块603、第二确定模块604和优化模块605,其中:

[0239] 获取模块601,用于获取多个业务需求;

[0240] 筛选模块602,用于针对每一业务需求,从数据资产管理模型中筛选出与所述业务需求对应的目标数据组;所述目标数据组中包括所述数据资产管理模型中的第一表级元数据和第一字段级元数据;

[0241] 第一确定模块603,用于确定每一业务需求与所对应的目标数据组之间的匹配度;

[0242] 第二确定模块604,用于将各个业务需求中,每一业务需求在多个业务需求中的出现频次和对应的匹配度满足特定条件的业务需求确定为资产优化需求;

[0243] 优化模块605,用于基于所述资产优化需求,对所述数据资产管理模型进行优化。

[0244] 在一个实施例中,所述数据资产管理模型包括数据仓库层、数据集市层和数据应用层;所述数据仓库层、所述数据集市层和所述数据应用层中包括多个所述数据表;每一所述数据表中包括表级元数据和字段级元数据。

[0245] 在一个实施例中,所述业务需求包括目标表级元数据和目标字段级元数据;所述筛选模块602包括:第一筛选子模块,用于根据每一所述业务需求的目标表级元数据,从所述数据资产管理模型中筛选出第一候选数据表;第二筛选单子模块,用于根据每一所述业务需求的目标字段级元数据,从所述第一候选数据表中筛选出第二候选数据表,并从所述第二候选数据表中筛选出候选元数据;组合子模块,用于对所述第二候选数据表进行排列组合,得到多个候选数据组;第三筛选子模块,用于根据所述多个候选数据组中每一所述候选数据组的各第二候选数据表的数据时延和数据层级,从所述多个候选数据组中确定目标数据组。

[0246] 在一个实施例中,所述第三筛选子模块,包括:第一确定单元,用于确定每一所述候选数据组中各第二候选数据表对应的数据层级;第二确定单元,用于根据每一所述数据层级中第二候选数据表的个数和对应数据层级的层级系数,确定第一推荐指数;第三确定单元,根据每一所述候选数据组中各第二候选数据表的数据时延,确定第二推荐指数;筛选单元,用于根据每一所述第一推荐指数和对应的第二推荐指数,从所述多个候选数据组中确定目标数据组。

[0247] 在一个实施例中,所述筛选单元,包括:

[0248] 第一确定子单元,用于将所述第一推荐指数和所述第二推荐指数之和最小的候选数据组,确定为目标数据组。

[0249] 在一个实施例中,所述筛选单元,包括:排序子单元,用于根据每一所述第一推荐

指数和对应的第二推荐指数,对所述多个候选数据组进行排序;第一筛选子单元,用于根据排序结果,从所述多个候选数据组中确定多个第一数据组;第二筛选子单元,用于响应于接收到的选取指令,从所述多个第一数据组中确定目标数据组。

[0250] 在一个实施例中,所述第二确定模块604,包括:第一确定子模块,用于将所述多个业务需求中出现频次最高且匹配度最低的业务需求,确定为资产优化需求。

[0251] 在一个实施例中,所述第二确定模块604,包括:第二确定子模块,用于根据所述多个业务需求中每一所述业务需求的出现频次和对应的匹配度,确定对应业务需求的优先级;第三确定子模块,用于将优先级满足特定条件的业务需求确定为资产优化需求。

[0252] 在一个实施例中,所述业务需求包括目标表级元数据和目标字段级元数据;所述优化模块605,包括:建立子模块,用于基于所述至少一个资产优化需求中每一所述资产优化需求,建立对应的第一数据表;添加子模块,用于将至少一个第一数据表加入所述数据资产管理模型中;

[0253] 在一个实施例中,所述优化模块605,包括:更新子模块,用于基于所述至少一个资产优化需求中每一所述资产优化需求的目标表级元数据和目标字段级元数据,对所述数据资产管理模型中的对应目标数据表的表级元数据和字段级元数据进行更新。

[0254] 在一个实施例中,所述装置还包括:第一建立模块,用于分别建立存储于数据仓库层、数据集市层和数据应用层中的数据表;第二建立模块,用于建立每一所述数据表对应的表级元数据和字段级元数据;第三建立模块,用于建立包括所述数据仓库层、所述数据集市层和所述数据应用层的所述数据资产管理模型。

[0255] 图7为本申请实施例数据资产管理模型的建立装置的组成结构示意图,如图7所示,所述装置700包括第一建立模块701、第二建立模块702和第三建立模块703,其中:

[0256] 第一建立模块701,用于根据终端信令数据的三要素,分别生成存储于数据仓库层、数据集市层和数据应用层中的数据表;所述三要素包括用户标识信息、用户位置信息和时间点;

[0257] 第二建立模块702,用于根据所述终端信令数据的三要素,建立每一所述数据表对应的表级元数据和字段级元数据,其中,所述数据仓库层用于存放经过预处理的所述终端信令数据的元数据中的明细数据;所述明细数据以用户标识信息和时间点匹配用户位置信息,或者以用户位置信息和时间点匹配用户标识信息;所述数据集市层用于存放经过所述数据仓库层处理后的所述终端信令数据的元数据中的标签数据;所述标签数据按主题分为用户标识信息、用户位置信息和时间点;所述数据应用层用于存放经过所述数据仓库层或者所述数据集市层处理后的所述终端信令数据的元数据中的常用的中间计算过程表;

[0258] 第三建立模块703,用于建立包括所述数据仓库层、所述数据集市层和所述数据应用层的所述数据资产管理模型。

[0259] 在一个实施例中,所述第一建立模块701,包括:第一建立子模块,用于在在所述数据仓库层中,根据终端信令数据的用户标识信息和用户位置信息,建立多个基础实时数据表;根据所述终端信令数据的用户标识信息、用户位置信息和时间点,建立离线事实数据表;根据所述基础实时数据表和所述离线事实数据表,建立与所述基础实时数据表、所述离线事实数据表相关的其他数据表;第二建立子模块,用于在所述数据集市层中,对所述数据仓库层中的多个基础实时数据表、所述离线事实数据表和所述其他数据表中的数据进行汇

总处理,建立包括多种标签数据的汇总标签数据表;第三建立子模块,用于在所述数据应用层中,对所述数据仓库层中的多个基础实时数据表、所述离线事实数据表和所述其他数据表中的数据,和/或所述数据集市层中的汇总标签数据表进行处理,建立包含多种粒度数据的汇总粒度数据表。

[0260] 在一个实施例中,所述表级元数据包括所属层级、业务场景、数据时效和数据时延;所述字段级元数据包括类别、范围和粒度;所述第二建立模块702,包括:第四建立子模块,用于根据所述终端信令数据的用户标识信息的所属层级、用户位置信息的所属层级和时间点的所属层级,建立每一所述数据表的所属层级;所述所属层级为所述数据仓库层、所述数据集市层和所述数据应用层中的一种;第五建立子模块,用于根据所述终端信令数据的用户标识信息、用户位置信息和时间点之间的检索关系,从预建立的各项业务场景中选取至少一种业务场景作为每一所述数据表的业务场景;第六建立子模块,用于根据所述终端信令数据的时间点的所属类别,建立每一所述数据表的数据时效;所述数据时效为离线表或实时表;第七建立子模块,用于根据所述终端信令数据的时间点的最大值,建立每一所述数据表的数据时延;所述数据时延为从原始数据到对应数据表的平均处理时长;第八建立子模块,用于根据所述终端信令数据的用户标识信息的所属类别、用户位置信息的所属类别和时间点的所属类别,建立每一所述数据表的人物类别、地点类别、时间类别和标签属性;第九建立子模块,用于根据所述终端信令数据的用户标识信息的所属范围、用户位置信息的所属范围和时间点的所属范围,建立每一所述数据表的人物范围、地点范围、时间范围和标签值;第十建立子模块,用于根据所述终端信令数据的用户标识信息的最小单元、用户位置信息的最小单元和时间点的最小单元,建立每一所述数据表的人物粒度、地点粒度和时间粒度。

[0261] 需要说明的是,本申请实施例中,如果以软件功能模块的形式实现上述的数据资产的运营方法,并作为独立的产品销售或使用,也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,本申请实施例的技术方案本质上或者说对相关技术做出贡献的部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质中,包括若干指令用以使得电子设备(可以是手机、平板电脑、台式机、个人数字助理、导航仪、数字电话、视频电话、电视机、传感设备等)执行本申请各个实施例所述方法的全部或部分。而前述的存储介质包括:U盘、移动硬盘、只读存储器(Read Only Memory,ROM)、磁碟或者光盘等各种可以存储程序代码的介质。这样,本申请实施例不局限于任何特定的硬件和软件结合。

[0262] 以上装置实施例的描述,与上述方法实施例的描述是类似的,具有同方法实施例相似的有益效果。对于本申请装置实施例中未披露的技术细节,请参照本申请方法实施例的描述而理解。

[0263] 对应地,本申请实施例提供一种电子设备,图8为本申请实施例电子设备的一种硬件实体示意图,如图8所示,该电子设备800的硬件实体包括:包括存储器801和处理器802,所述存储器801存储有可在处理器802上运行的计算机程序,所述处理器802执行所述程序时实现上述实施例数据资产的运营方法或者数据资产管理模型的建立方法中的步骤。

[0264] 存储器801配置为存储由处理器802可执行的指令和应用,还可以缓存待处理器802以及电子设备800中各模块待处理或已经处理的数据(例如,图像数据、音频数据、语音通信数据和视频通信数据),可以通过闪存(FLASH)或随机访问存储器(Random Access

Memory, RAM) 实现。

[0265] 对应地,本申请实施例提供一种计算机可读存储介质,其上存储有计算机程序,该计算机程序被处理器执行时实现上述实施例中提供的数据资产的运营方法或者数据资产管理模型的建立方法中的步骤。

[0266] 这里需要指出的是:以上存储介质和设备实施例的描述,与上述方法实施例的描述是类似的,具有同设备实施例相似的有益效果。对于本申请存储介质和方法实施例中未披露的技术细节,请参照本申请设备实施例的描述而理解。

[0267] 应理解,说明书通篇中提到的“一个实施例”或“一实施例”意味着与实施例有关的特定特征、结构或特性包括在本申请的至少一个实施例中。因此,在整个说明书各处出现的“在一个实施例中”或“在一实施例中”未必一定指相同的实施例。此外,这些特定的特征、结构或特性可以任意适合的方式结合在一个或多个实施例中。应理解,在本申请的各种实施例中,上述各过程的序号的大小并不意味着执行顺序的先后,各过程的执行顺序应以其功能和内在逻辑确定,而不对本申请实施例的实施过程构成任何限定。上述本申请实施例序号仅仅为了描述,不代表实施例的优劣。

[0268] 需要说明的是,在本文中,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者装置不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者装置所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括该要素的过程、方法、物品或者装置中还存在另外的相同要素。

[0269] 在本申请所提供的几个实施例中,应该理解到,所揭露的设备和方法,可以通过其它的方式实现。以上所描述的设备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,如:多个单元或组件可以结合,或可以集成到另一个系统,或一些特征可以忽略,或不执行。另外,所显示或讨论的各组成部分相互之间的耦合、或直接耦合、或通信连接可以通过一些接口,设备或单元的间接耦合或通信连接,可以是电性的、机械的或其它形式的。

[0270] 上述作为分离部件说明的单元可以是、或也可以不是物理上分开的,作为单元显示的部件可以是、或也可以不是物理单元;既可以位于一个地方,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全部单元来实现本实施例方案的目的。另外,在本申请各实施例中的各功能单元可以全部集成在一个处理单元中,也可以是各单元分别单独作为一个单元,也可以两个或两个以上单元集成在一个单元中;上述集成的单元既可以采用硬件的形式实现,也可以采用硬件加软件功能单元的形式实现。

[0271] 本领域普通技术人员可以理解:实现上述方法实施例的全部或部分步骤可以通过程序指令相关的硬件来完成,前述的程序可以存储于计算机可读取存储介质中,该程序在执行时,执行包括上述方法实施例的步骤;而前述的存储介质包括:移动存储设备、只读存储器(Read Only Memory, ROM)、磁碟或者光盘等各种可以存储程序代码的介质。或者,本申请上述集成的单元如果以软件功能模块的形式实现并作为独立的产品销售或使用,也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,本申请实施例的技术方案本质上或者说对相关技术做出贡献的部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质中,包括若干指令用以使得计算机设备(可以是手机、平板电脑、台式

机、个人数字助理、导航仪、数字电话、视频电话、电视机、传感设备等)执行本申请各个实施例所述方法的全部或部分。而前述的存储介质包括:移动存储设备、ROM、磁碟或者光盘等各种可以存储程序代码的介质。

[0272] 本申请所提供的几个方法实施例中所揭露的方法,在不冲突的情况下可以任意组合,得到新的方法实施例。本申请所提供的几个产品实施例中所揭露的特征,在不冲突的情况下可以任意组合,得到新的产品实施例。本申请所提供的几个方法或设备实施例中所揭露的特征,在不冲突的情况下可以任意组合,得到新的方法实施例或设备实施例。

[0273] 以上所述,仅为本申请的实施方式,但本申请的保护范围并不局限于此,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,可轻易想到变化或替换,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所述权利要求的保护范围为准。

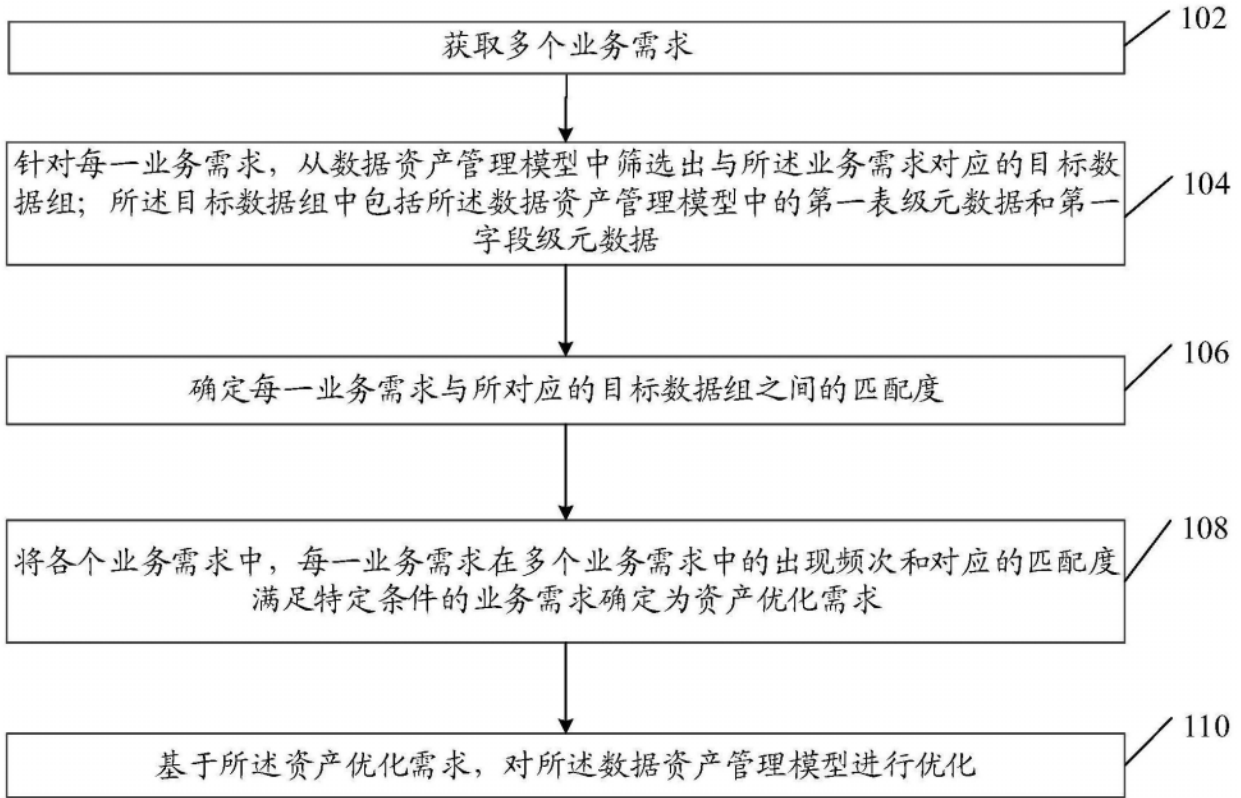


图1

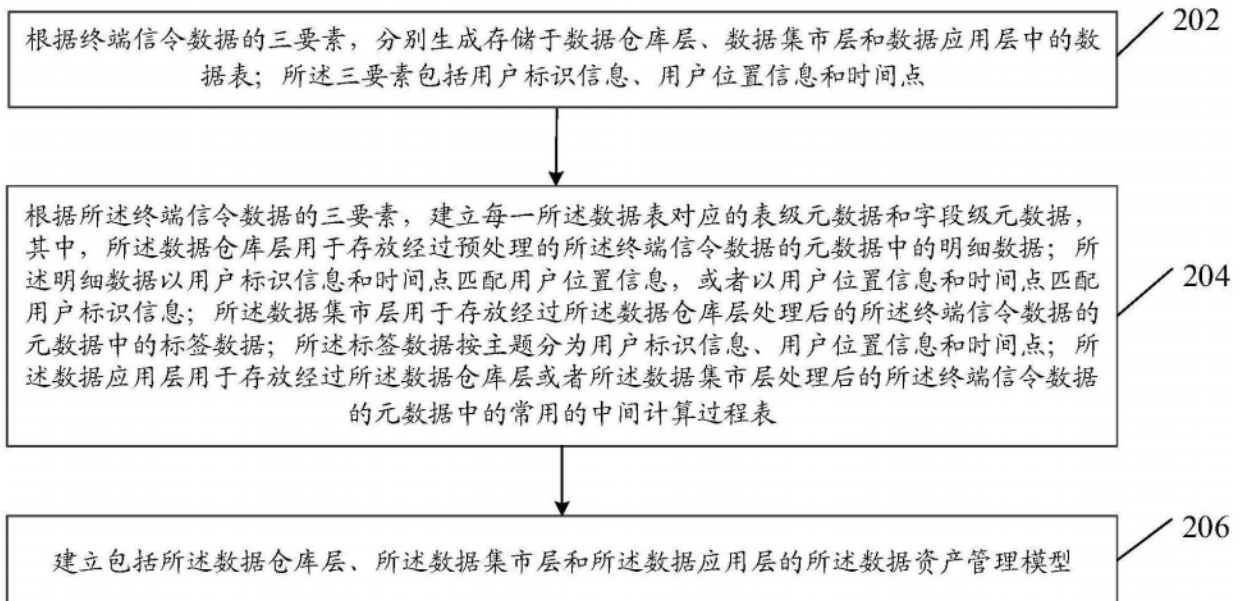


图2

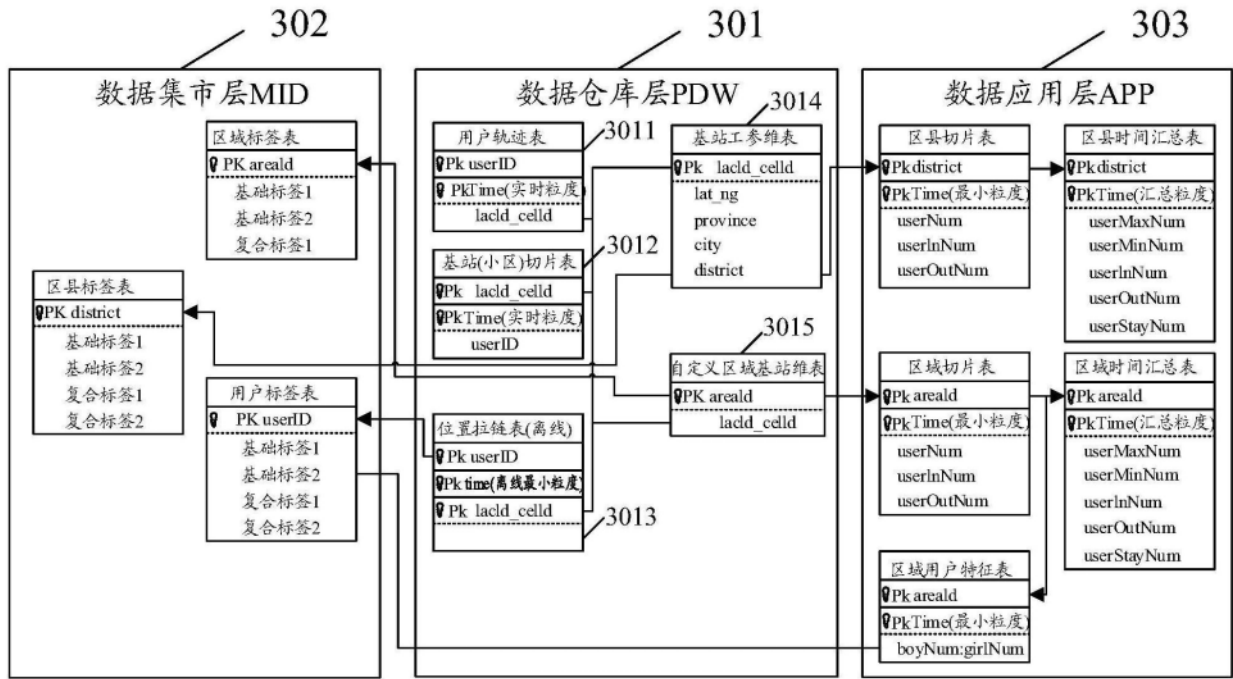


图3

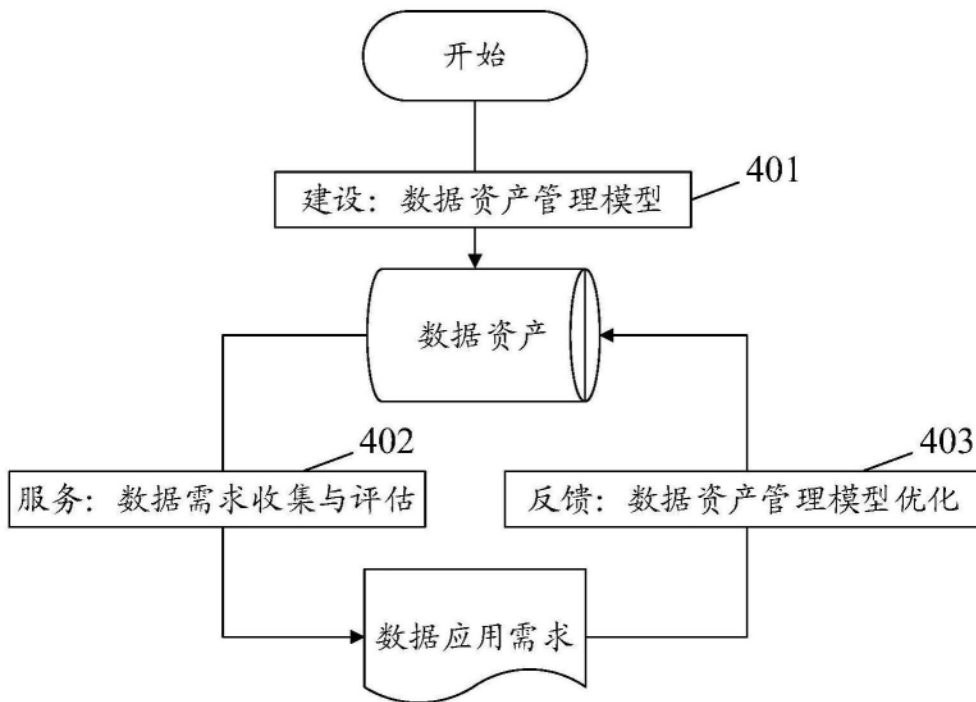


图4

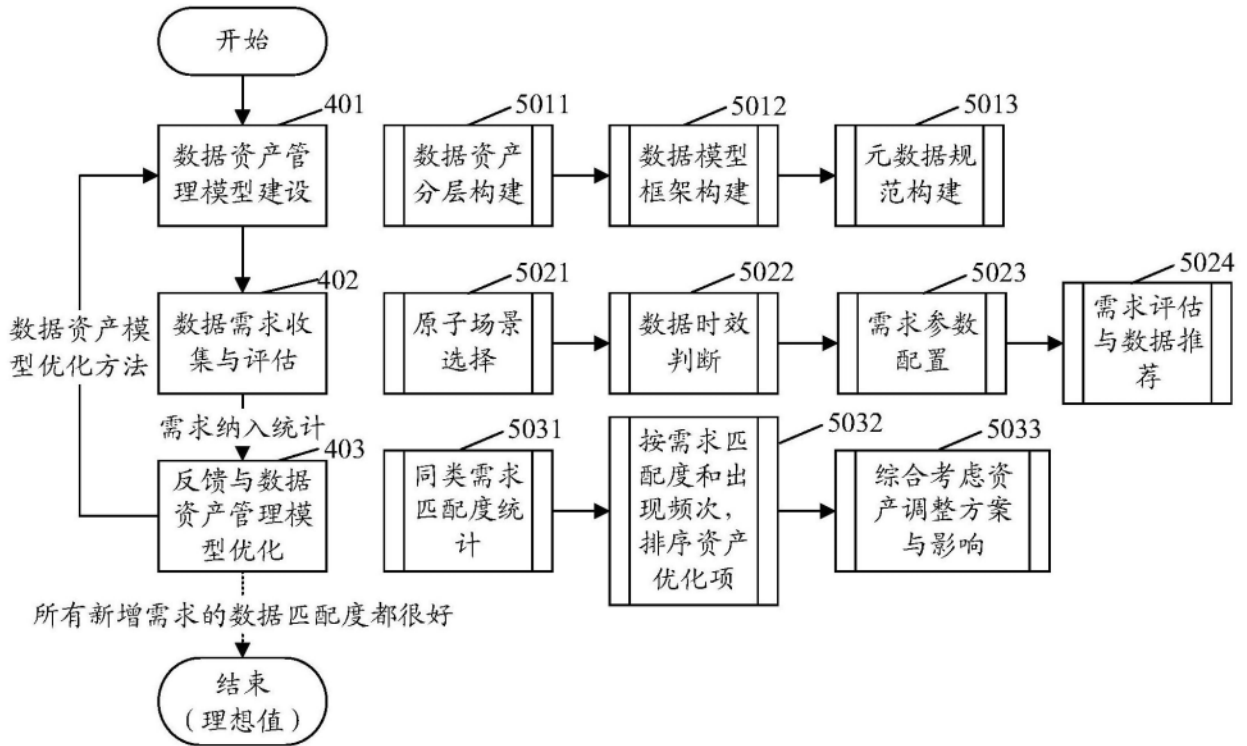


图5

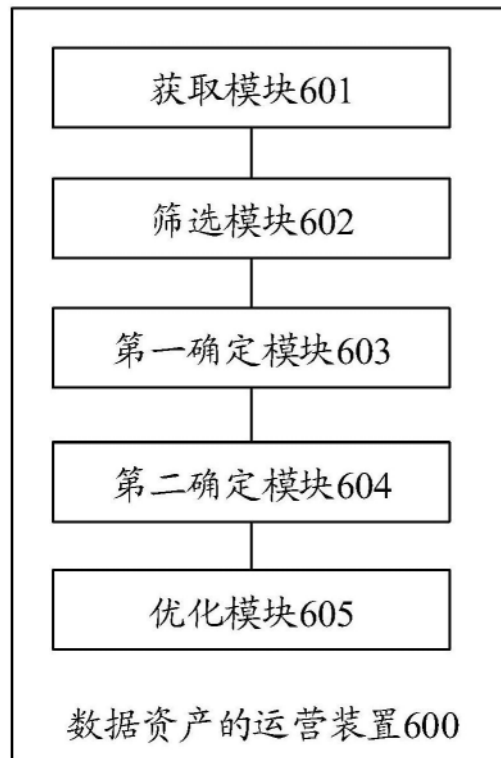


图6

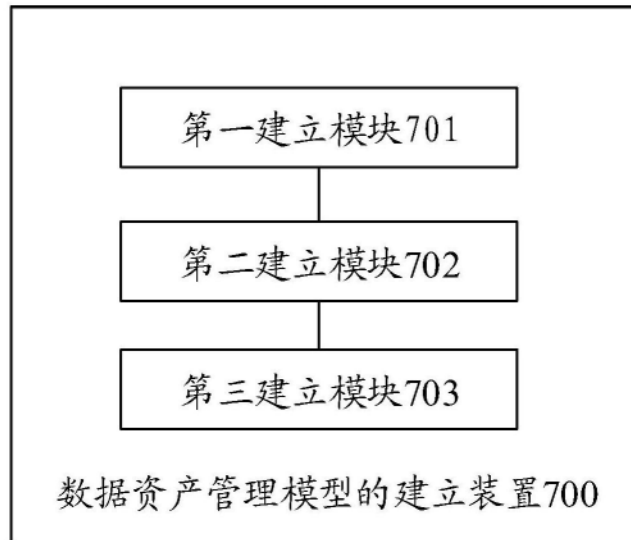


图7

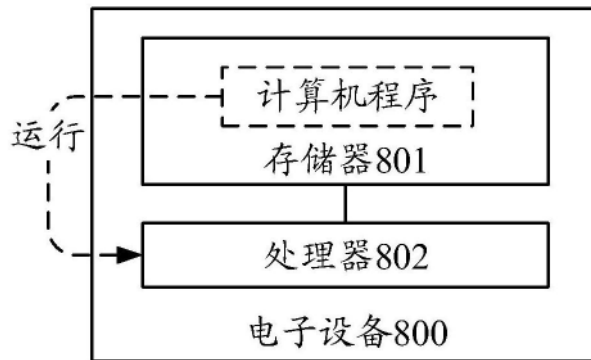


图8