[19] Patents Registry
The Hong Kong Special Administrative Region
香港特別行政區
專利註冊處

[11] 40075673 B
CN 115292353 B

[12] **STANDARD PATENT (R) SPECIFICATION**
轉錄標準專利說明書

[21] Application no. 申請編號
42022065263.0

[22] Date of filing 提交日期
07.12.2022

[51] Int. Cl.
G06F 16/242 (2019.01) G06F 16/2455 (2019.01)

[54] DATA QUERY METHOD AND APPARATUS, COMPUTER DEVICE, AND STORAGE MEDIUM
數據查詢方法、裝置、計算機設備和存儲介質

[43] Date of publication of application 申請發表日期
27.01.2023

[45] Date of publication of grant of patent 批予專利的發表日期
31.03.2023

CN Application no. & date 中國專利申請編號及日期
CN 202211227039.8 09.10.2022

CN Publication no. & date 中國專利申請發表編號及日期
CN 115292353 04.11.2022

Date of grant in designated patent office 指定專利當局批予專利日期
27.12.2022

[73] Proprietor 專利所有人
TENCENT TECHNOLOGY (SHENZHEN) COMPANY LIMITED
腾讯科技(深圳)有限公司
35/F,Tencent Building
Kejizhongyi Road, Midwest District of Hi-tech Park,
Nanshan District,Shenzhen,Guangdong 518057
中國
广东省深圳市南山区高新区科技中一路
腾讯大厦 35 层

[72] Inventor 發明人
HUANG,CHENYU 黃晨宇
JIANG,JIE 蔣杰
LIU,YUHONG 刘煜宏
CHEN,PENG 陈鹏
CHENG,YONG 程勇
FAN,XIAOLIANG 范晓亮

[74] Agent and / or address for service 代理人及/或送達地址
集佳知識產權有限公司
香港
九龍旺角洗衣街 39-55 號
金雞廣場 12 層 1201 室

(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 115292353 B
(45)授权公告日 2022.12.27

(21)申请号 202211227039.8
(22)申请日 2022.10.09
(65)同一申请的已公布的文献号
申请公布号 CN 115292353 A
(43)申请公布日 2022.11.04
(73)专利权人 腾讯科技(深圳)有限公司
地址 518057 广东省深圳市南山区高新区科技中一路腾讯大厦35层
(72)发明人 黄晨宇 蒋杰 刘煜宏 陈鹏 程勇 范晓亮
(74)专利代理机构 华进联合专利商标代理有限公司 44224
专利代理师 毛丹
(51)Int.Cl.
G06F 16/242(2019.01)
G06F 16/2455(2019.01)
(56)对比文件
CN 113468208 A,2021.10.01
US 2015074083 A1,2015.03.12
US 2019147085 A1,2019.05.16
审查员 夏雪

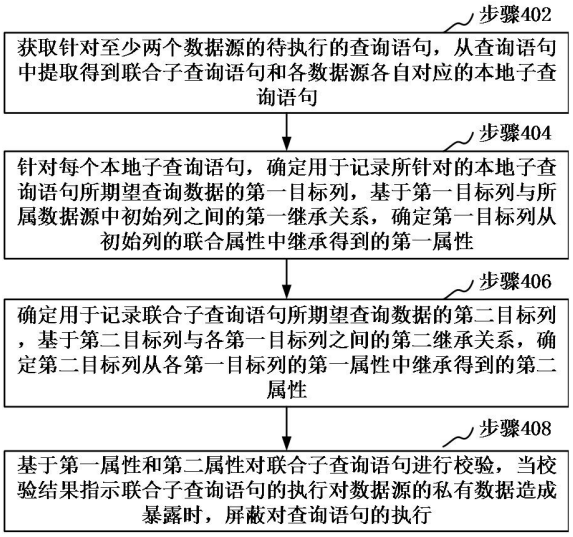
权利要求书7页 说明书20页 附图5页

(54)发明名称

数据查询方法、装置、计算机设备和存储介质

(57)摘要

本申请涉及一种数据查询方法、装置、计算机设备和存储介质，可应用于云技术、人工智能、智慧交通、辅助驾驶等各种场景。其中方法包括：获取查询语句，从查询语句中提取得到联合子查询语句和本地子查询语句；确定本地子查询语句对应的第一目标列，基于第一目标列与所属数据源中初始列之间的第一继承关系，确定第一目标列继承得到的第一属性；确定联合子查询语句对应的第二目标列，基于第二目标列与第一目标列之间的第二继承关系，确定第二目标列继承得到的第二属性；基于第一属性和第二属性对联合子查询语句进行校验，当校验结果指示不合规，屏蔽对查询语句的执行。采用本申请的方法可以提高联合分析过程中隐私数据的安全性。



CN 115292353 B

1.一种数据查询方法,其特征在于,所述方法包括:

获取针对至少两个数据源的待执行的查询语句,对所述查询语句进行拆分,得到联合子查询语句和各数据源各自对应的本地子查询语句;所述至少两个数据源分别属于不同的数据拥有方;

针对每个本地子查询语句,确定用于记录所针对的本地子查询语句所期望查询数据的第一目标列,从所述第一目标列所属数据源的初始列中确定目标初始列;所述目标初始列包括所述所针对的本地子查询语句中列选择算子所作用的列或者分组算子所作用的列中的至少一种;

基于所述第一目标列与所述目标初始列之间的第一继承关系,确定所述第一目标列从所述目标初始列的联合属性中继承得到的第一属性;所述联合属性用于对所述初始列进行针对各数据源的联合描述;

确定用于记录所述联合子查询语句所期望查询数据的第二目标列,基于所述第二目标列与各所述第一目标列之间的第二继承关系,确定所述第二目标列从各所述第一目标列的第一属性中继承得到的第二属性;

基于所述第一属性和所述第二属性对所述联合子查询语句进行校验,当校验结果指示所述联合子查询语句的执行对数据源的私有数据造成暴露时,屏蔽对所述查询语句的执行。

2.根据权利要求1所述的方法,其特征在于,所述基于所述第二目标列与各所述第一目标列之间的第二继承关系,确定所述第二目标列从各所述第一目标列的第一属性中继承得到的第二属性包括:

从各所述第一目标列中确定所述联合子查询语句中的列选择算子所作用的列,得到目标选择列;

基于所述第二目标列与所述目标选择列之间的第二继承关系,确定所述第二目标列从所述目标选择列的第一属性中继承得到的第二属性。

3.根据权利要求2所述的方法,其特征在于,所述基于所述第二目标列与所述目标选择列之间的第二继承关系,确定所述第二目标列从所述目标选择列的第一属性中继承得到的第二属性,包括:

当所述第二继承关系包括来源继承关系时,确定所述第二目标列从所述目标选择列的第一属性中继承得到的第二属性包括数据来源信息;

当所述第二继承关系包括权限继承关系时,确定所述第二目标列从所述目标选择列的第一属性中继承得到的第二属性包括联合操作权限信息;

当所述第二继承关系包括主键继承关系时,确定所述第二目标列从所述目标选择列的第一属性中继承得到的第二属性包括主键信息。

4.根据权利要求1所述的方法,其特征在于,所述基于所述第二目标列与各所述第一目标列之间的第二继承关系,确定所述第二目标列从各所述第一目标列的第一属性中继承得到的第二属性包括:

从各所述第一目标列中确定所述联合子查询语句中的分组算子所作用的列,得到目标分组列;

基于所述第二目标列与所述目标分组列之间的第二继承关系,确定所述第二目标列从

所述目标分组列的第一属性中继承得到的第二属性。

5.根据权利要求4所述的方法，其特征在于，所述基于所述第二目标列与所述目标分组列之间的第二继承关系，确定所述第二目标列从所述目标分组列的第一属性中继承得到的第二属性包括：

当所述第二继承关系包括来源继承关系时，确定所述第二目标列从所述目标分组列的第一属性中继承得到的第二属性包括数据来源信息；

当所述第二继承关系包括权限继承关系时，确定所述第二目标列从所述目标分组列的第一属性中继承得到的第二属性包括联合操作权限信息；

当所述第二继承关系包括主键继承关系且所述第二目标列的列名与所述目标分组列的列名一致时，确定所述第二目标列从所述目标分组列的第一属性中继承得到的第二属性包括主键信息。

6.根据权利要求1所述的方法，其特征在于，所述从所述第一目标列所属数据源的初始列中确定目标初始列，包括：

从所述初始列中确定所述所针对的本地子查询语句包括的列选择算子所作用的列，得到目标选择列；

所述基于所述第一目标列与所述目标初始列之间的第一继承关系，确定所述第一目标列从所述目标初始列的联合属性中继承得到的第一属性，包括：

基于所述第一目标列与所述目标选择列之间的第一继承关系，确定所述第一目标列从所述目标选择列的联合属性中继承得到的第一属性。

7.根据权利要求6所述的方法，其特征在于，所述基于所述第一目标列与所述目标选择列之间的第一继承关系，确定所述第一目标列从所述目标选择列的第一属性中继承得到的第一属性，包括：

当所述第一继承关系包括来源继承关系时，确定所述第一目标列从所述目标选择列的第一属性中继承得到的第一属性包括数据来源信息；

当所述第一继承关系包括权限继承关系时，确定所述第一目标列从所述目标选择列的第一属性中继承得到的第一属性包括联合操作权限信息；

当所述第一继承关系包括主键继承关系时，确定所述第一目标列从所述目标选择列的第一属性中继承得到的第一属性包括主键信息。

8.根据权利要求6所述的方法，其特征在于，所述从所述第一目标列所属数据源的初始列中确定目标初始列，包括：

从所述初始列中确定所述分组算子所作用的列，得到目标分组列；

所述基于所述第一目标列与所述目标初始列之间的第一继承关系，确定所述第一目标列从所述目标初始列的联合属性中继承得到的第一属性，包括：

基于所述第一目标列与所述目标分组列之间的第一继承关系，确定所述第一目标列从所述目标分组列的联合属性中继承得到的第一属性。

9.根据权利要求8所述的方法，其特征在于，所述基于所述第一目标列与所述目标分组列之间的第一继承关系，确定所述第一目标列从所述目标分组列的第一属性中继承得到的第一属性包括：

当所述第一继承关系包括来源继承关系时，确定所述第一目标列从所述目标分组列的

第一属性中继承得到的第一属性包括数据来源信息；

当所述第一继承关系包括权限继承关系时，确定所述第一目标列从所述目标分组列的第一属性中继承得到的第一属性包括联合操作权限信息；

当所述第一继承关系包括主键继承关系且所述第一目标列的列名与所述目标分组列的列名一致时，确定所述第一目标列从所述目标分组列的第一属性中继承得到的第一属性包括主键信息。

10.根据权利要求1所述的方法，其特征在于，所述基于所述第一属性和所述第二属性对所述联合子查询语句进行校验包括：

从所述联合子查询语句中确定所包含的算子；

针对每个算子，确定所针对的算子所作用的列，基于所述第一属性和所述第二属性确定所述所针对的算子所作用的列的目标属性；

基于所述所针对的算子对应的校验规则对所述目标属性进行校验，得到校验结果。

11.根据权利要求10所述的方法，其特征在于，所述针对每个算子，确定所针对的算子所作用的列，基于所述第一属性和所述第二属性确定所述所针对的算子所作用的列的目标属性，包括：

针对列选择算子，确定所述列选择算子所作用的列，基于所述第一属性和所述第二属性确定所述列选择算子所作用的列的联合操作权限信息；

所述基于所述所针对的算子对应的校验规则对所述目标属性进行校验，得到校验结果，包括：

当所述列选择算子所作用的列的联合操作权限信息指示不具备联合操作权限，得到第一校验结果，所述第一校验结果指示所述列选择算子的执行对数据源的私有数据造成暴露。

12.根据权利要求10所述的方法，其特征在于，所述针对每个算子，确定所针对的算子所作用的列，基于所述第一属性和所述第二属性确定所述所针对的算子所作用的列的目标属性，包括：

针对分组算子，确定所述分组算子所作用的列，基于所述第一属性和所述第二属性确定所述分组算子所作用的列的主键信息；

所述基于所述所针对的算子对应的校验规则对所述目标属性进行校验，得到校验结果，包括：

当所述分组算子所作用的列的主键信息指示包含所属数据源的联合主键，得到第二校验结果，所述第二校验结果指示所述联合子查询语句中分组算子的执行对数据源的私有数据造成暴露。

13.根据权利要求10所述的方法，其特征在于，所述针对每个算子，确定所针对的算子所作用的列，基于所述第一属性和所述第二属性确定所述所针对的算子所作用的列的目标属性，包括：

针对条件过滤算子，确定所述条件过滤算子所作用的列，基于所述第一属性和所述第二属性确定所述条件过滤算子所作用的列的联合操作权限信息；

所述基于所述所针对的算子对应的校验规则对所述目标属性进行校验，得到校验结果，包括：

当所述条件过滤算子所作用的列的联合操作权限信息指示不具备联合操作权限，得到第三校验结果，所述第三校验结果指示所述条件过滤算子的执行对数据源的私有数据造成暴露。

14.根据权利要求1至9任意一项所述的方法，其特征在于，所述方法还包括：针对各数据源中初始列生成对应的初始节点，将各所述初始列的联合属性记录在各自对应的初始节点中；

所述基于所述第一目标列与所述目标初始列之间的第一继承关系，确定所述第一目标列从所述目标初始列的联合属性中继承得到的第一属性包括：

针对所述第一目标列生成对应的第一节点，基于所述第一目标列与所述目标初始列之间的第一继承关系，建立第一节点和初始节点之间的第一连接关系，基于所述第一连接关系确定第一节点从初始节点所记录的联合属性中继承得到的第一属性，并记录在对应的第一节点中；

所述基于所述第二目标列与各所述第一目标列之间的第二继承关系，确定所述第二目标列从各所述第一目标列的第一属性中继承得到的第二属性，包括：

针对所述第二目标列生成对应的第二节点，基于所述第二目标列与各所述第一目标列之间的第二继承关系，建立第一节点和第二节点之间的第二连接关系，以构建数据血缘森林，基于所述第二连接关系确定第二节点从各所述第一节点所记录的第一属性中继承得到的第二属性，并记录在对应的第二节点中。

15.根据权利要求14所述的方法，其特征在于，所述数据血缘森林中各个节点存在对应的全局唯一标识，所述基于所述第一属性和所述第二属性对所述联合子查询语句进行校验包括：

从所述联合子查询语句中确定所包含的算子；

针对每个算子，确定所针对的算子所作用的列，将所述所针对的算子所作用的列的全局唯一标识，和数据血缘森林中各个节点的全局唯一标识进行匹配；

获取匹配成功的节点所记录的目标属性，基于所述所针对的算子对应的校验规则对所述目标属性进行校验，得到校验结果。

16.一种数据查询装置，其特征在于，所述装置包括：

查询语句获取模块，用于获取针对至少两个数据源的查询语句，对所述查询语句进行拆分，得到联合子查询语句和各数据源各自对应的本地子查询语句；所述至少两个数据源分别属于不同的数据拥有方；

第一属性确定模块，用于针对每个本地子查询语句，确定用于记录所针对的本地子查询语句所期望查询数据的第一目标列，从所述第一目标列所属数据源的初始列中确定目标初始列；基于所述第一目标列与所述目标初始列之间的第一继承关系，确定所述第一目标列从所述目标初始列的联合属性中继承得到的第一属性；所述联合属性用于对所述初始列进行针对各数据源的联合描述；所述目标初始列包括所述所针对的本地子查询语句中列选择算子所作用的列或者分组算子所作用的列中的至少一种；

第二属性确定模块，用于确定用于记录所述联合子查询语句所期望查询数据的第二目标列，基于所述第二目标列与各所述第一目标列之间的第二继承关系，确定所述第二目标列从各所述第一目标列的第一属性中继承得到的第二属性；

校验模块，用于基于所述第一属性和所述第二属性对所述联合子查询语句进行校验，当校验结果指示所述联合子查询语句的执行对数据源的私有数据造成暴露时，屏蔽对所述查询语句的执行。

17.根据权利要求16所述的装置，其特征在于，所述第二属性确定模块，还用于：

从各所述第一目标列中确定所述联合子查询语句中的列选择算子所作用的列，得到目标选择列；

基于所述第二目标列与所述目标选择列之间的第二继承关系，确定所述第二目标列从所述目标选择列的第一属性中继承得到的第二属性。

18.根据权利要求17所述的装置，其特征在于，所述第二属性确定模块，还用于：

当所述第二继承关系包括来源继承关系时，确定所述第二目标列从所述目标选择列的第一属性中继承得到的第二属性包括数据来源信息；

当所述第二继承关系包括权限继承关系时，确定所述第二目标列从所述目标选择列的第一属性中继承得到的第二属性包括联合操作权限信息；

当所述第二继承关系包括主键继承关系时，确定所述第二目标列从所述目标选择列的第一属性中继承得到的第二属性包括主键信息。

19.根据权利要求16所述的装置，其特征在于，所述第二属性确定模块，还用于：

从各所述第一目标列中确定所述联合子查询语句中的分组算子所作用的列，得到目标分组列；

基于所述第二目标列与所述目标分组列之间的第二继承关系，确定所述第二目标列从所述目标分组列的第一属性中继承得到的第二属性。

20.根据权利要求19所述的装置，其特征在于，所述第二属性确定模块，还用于：

当所述第二继承关系包括来源继承关系时，确定所述第二目标列从所述目标分组列的第一属性中继承得到的第二属性包括数据来源信息；

当所述第二继承关系包括权限继承关系时，确定所述第二目标列从所述目标分组列的第一属性中继承得到的第二属性包括联合操作权限信息；

当所述第二继承关系包括主键继承关系且所述第二目标列的列名与所述目标分组列的列名一致时，确定所述第二目标列从所述目标分组列的第一属性中继承得到的第二属性包括主键信息。

21.根据权利要求16所述的装置，其特征在于，所述第一属性确定模块，还用于：

从所述初始列中确定所述所针对的本地子查询语句包括的列选择算子所作用的列，得到目标选择列；

基于所述第一目标列与所述目标选择列之间的第一继承关系，确定所述第一目标列从所述目标选择列的联合属性中继承得到的第一属性。

22.根据权利要求21所述的装置，其特征在于，所述第一属性确定模块，还用于：

从所述初始列中确定所述分组算子所作用的列，得到目标分组列；

基于所述第一目标列与所述目标分组列之间的第一继承关系，确定所述第一目标列从所述目标分组列的联合属性中继承得到的第一属性。

23.根据权利要求16所述的装置，其特征在于，所述校验模块，还用于

从所述联合子查询语句中确定所包含的算子；

针对每个算子，确定所针对的算子所作用的列，基于所述第一属性和所述第二属性确定所述所针对的算子所作用的列的目标属性；

基于所述所针对的算子对应的校验规则对所述目标属性进行校验，得到校验结果。

24.根据权利要求23所述的装置，其特征在于，所述校验模块，还用于：

针对列选择算子，确定所述列选择算子所作用的列，基于所述第一属性和所述第二属性确定所述列选择算子所作用的列的联合操作权限信息；

当所述列选择算子所作用的列的联合操作权限信息指示不具备联合操作权限，得到第一校验结果，所述第一校验结果指示所述列选择算子的执行对数据源的私有数据造成暴露。

25.根据权利要求23所述的装置，其特征在于，所述校验模块，还用于：

针对分组算子，确定所述分组算子所作用的列，基于所述第一属性和所述第二属性确定所述分组算子所作用的列的主键信息；

当所述分组算子所作用的列的主键信息指示包含所属数据源的联合主键，得到第二校验结果，所述第二校验结果指示所述联合子查询语句中分组算子的执行对数据源的私有数据造成暴露。

26.根据权利要求23所述的装置，其特征在于，所述校验模块，还用于：

针对条件过滤算子，确定所述条件过滤算子所作用的列，基于所述第一属性和所述第二属性确定所述条件过滤算子所作用的列的联合操作权限信息；

当所述条件过滤算子所作用的列的联合操作权限信息指示不具备联合操作权限，得到第三校验结果，所述第三校验结果指示所述条件过滤算子的执行对数据源的私有数据造成暴露。

27.根据权利要求16至22任意一项所述的装置，其特征在于，所述装置还用于：针对各数据源中初始列生成对应的初始节点，将各所述初始列的联合属性记录在各自对应的初始节点中；

所述第一属性确定模块，还用于：

针对所述第一目标列生成对应的第一节点，基于所述第一目标列与所述目标初始列之间的第一继承关系，建立第一节点和初始节点之间的第一连接关系，基于所述第一连接关系确定第一节点从初始节点所记录的联合属性中继承得到的第一属性，并记录在对应的第一节点中；

所述第二属性确定模块，还用于：

针对所述第二目标列生成对应的第二节点，基于所述第二目标列与各所述第一目标列之间的第二继承关系，建立第一节点和第二节点之间的第二连接关系，以构建数据血缘森林，基于所述第二连接关系确定第二节点从各所述第一节点所记录的第一属性中继承得到的第二属性，并记录在对应的第二节点中。

28.根据权利要求27所述的装置，其特征在于，所述校验模块，还用于：

从所述联合子查询语句中确定所包含的算子；

针对每个算子，确定所针对的算子所作用的列，将所述所针对的算子所作用的列的全局唯一标识，和数据血缘森林中各个节点的全局唯一标识进行匹配；

获取匹配成功的节点所记录的目标属性，基于所述所针对的算子对应的校验规则对所

述目标属性进行校验,得到校验结果。

29.一种计算机设备,包括存储器和处理器,所述存储器存储有计算机程序,其特征在于,所述处理器执行所述计算机程序时实现权利要求1至15中任一项所述的方法的步骤。

30.一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,所述计算机程序被处理器执行时实现权利要求1至15中任一项所述的方法的步骤。

数据查询方法、装置、计算机设备和存储介质

技术领域

[0001] 本申请涉及隐私计算技术领域，特别是涉及一种数据查询方法、装置、计算机设备、存储介质和计算机程序产品。

背景技术

[0002] 随着隐私计算的发展，出现了联合分析(Federated Analysis)技术，联合分析是一种分布式OLAP(on-Line Analytic Processing，联机分析处理)，该方法分散了分析数据的过程，从而无需将数据发送到集中式的服务器就可以进行数据分析，联合分析采用密码学方法等方法，在打破数据孤岛的同时保护用户数据隐私。在整个计算过程中，参与方/攻击者无法通过中间数据或是结果推导对方用户的原始数据。例如参与双方各有一张表想要执行Join操作，即求两张表的交集，双方可以执行PSI(Private Set Intersection)协议在获得交集的同时不会暴露己方非交集的数据给对方。

[0003] 然而目前的联合分析是在数据的计算等方面进行隐私保护，经常存在用户输入的分析指令虽然是正确的，但是执行指令可能会造成隐私数据的暴露的问题，导致联合分析过程中隐私数据的安全性低。

发明内容

[0004] 基于此，有必要针对上述技术问题，提供一种数据查询方法、装置、计算机设备、计算机可读存储介质和计算机程序产品，以提高联合分析过程中隐私数据的安全性。

[0005] 一方面，本申请提供了一种数据查询方法。所述方法包括：获取针对至少两个数据源的待执行的查询语句，从所述查询语句中提取得到联合子查询语句和各数据源各自对应的本地子查询语句；针对每个本地子查询语句，确定用于记录所述所针对的本地子查询语句所期望查询数据的第一目标列，基于所述第一目标列与所属数据源中初始列之间的第一继承关系，确定所述第一目标列从所述初始列的联合属性中继承得到的第一属性；所述联合属性用于对所述初始列进行针对各数据源的联合描述；确定用于记录所述联合子查询语句所期望查询数据的第二目标列，基于所述第二目标列与各所述第一目标列之间的第二继承关系，确定所述第二目标列从各所述第一目标列的第一属性中继承得到的第二属性；基于所述第一属性和所述第二属性对所述联合子查询语句进行校验，当校验结果指示所述联合子查询语句的执行对数据源的私有数据造成暴露时，屏蔽对所述查询语句的执行。

[0006] 在其中一个实施例中，所述基于所述第一目标列与所述目标选择列之间的第一继承关系，确定所述第一目标列从所述目标选择列的第一属性中继承得到的第一属性，包括：当所述第一继承关系包括来源继承关系时，确定所述第一目标列从所述目标选择列的第一属性中继承得到的第一属性包括数据来源信息；当所述第一继承关系包括权限继承关系时，确定所述第一目标列从所述目标选择列的第一属性中继承得到的第一属性包括联合操作权限信息；当所述第一继承关系包括主键继承关系时，确定所述第一目标列从所述目标选择列的第一属性中继承得到的第一属性包括主键信息。

[0007] 在其中一个实施例中，所述基于所述第一目标列与所述目标分组列之间的第一继承关系，确定所述第一目标列从所述目标分组列的第一属性中继承得到的第一属性包括：当所述第一继承关系包括来源继承关系时，确定所述第一目标列从所述目标分组列的第一属性中继承得到的第一属性包括数据来源信息；当所述第一继承关系包括权限继承关系时，确定所述第一目标列从所述目标分组列的第一属性中继承得到的第一属性包括联合操作权限信息；当所述第一继承关系包括主键继承关系且所述第一目标列的列名与所述目标分组列的列名一致时，确定所述第一目标列从所述目标分组列的第一属性中继承得到的第一属性包括主键信息。

[0008] 另一方面，本申请还提供了一种数据查询装置。所述装置包括：查询语句获取模块，用于获取针对至少两个数据源的查询语句，从所述查询语句中提取得到联合子查询语句和各数据源各自对应的本地子查询语句；第一属性确定模块，用于针对每个本地子查询语句，确定用于记录所述所针对的本地子查询语句所期望查询数据的第一目标列，基于所述第一目标列与所属数据源中初始列之间的第一继承关系，确定所述第一目标列从所述初始列的联合属性中继承得到的第一属性；所述联合属性用于对所述初始列进行针对各数据源的联合描述；第二属性确定模块，用于确定用于记录所述联合子查询语句所期望查询数据的第二目标列，基于所述第二目标列与各所述第一目标列之间的第二继承关系，确定所述第二目标列从各所述第一目标列的第一属性中继承得到的第二属性；

[0009] 校验模块，用于基于所述第一属性和所述第二属性对所述联合子查询语句进行校验，当校验结果指示所述联合子查询语句的执行对数据源的私有数据造成暴露时，屏蔽对所述查询语句的执行。

[0010] 另一方面，本申请还提供了一种计算机设备。所述计算机设备包括存储器和处理器，所述存储器存储有计算机程序，所述处理器执行所述计算机程序时实现上述数据查询方法的步骤。

[0011] 另一方面，本申请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，其上存储有计算机程序，所述计算机程序被处理器执行时实现数据查询方法的步骤。

[0012] 另一方面，本申请还提供了一种计算机程序产品。所述计算机程序产品，包括计算机程序，该计算机程序被处理器执行时实现数据查询方法的步骤。

[0013] 上述数据查询方法、装置、计算机设备、存储介质和计算机程序产品，在获取到针对至少两个数据源的待执行的查询语句，从查询语句中提取得到联合子查询语句和各数据源各自对应的本地子查询语句，针对每个本地子查询语句，确定用于记录所针对的本地子查询语句所期望查询数据的第一目标列，基于第一目标列与所属数据源中初始列之间的第一继承关系，确定第一目标列从初始列的联合属性中继承得到的第一属性，确定用于记录联合子查询语句所期望查询数据的第二目标列，基于第二目标列与各第一目标列之间的第二继承关系，确定第二目标列从各第一目标列的第一属性中继承得到的第二属性，基于第一属性和第二属性对联合子查询语句进行校验，当校验结果指示联合子查询语句的执行对数据源的私有数据造成暴露时，屏蔽对查询语句的执行，从而可以实现通过继承关系的分析，在SQL指令执行之前，对SQL指令是否合规进行校验，对不合规的SQL指令屏蔽执行，提高了联合分析过程中的数据安全性。

附图说明

[0014] 图1为一个实施例中数据源的数据示例；

[0015] 图2为一个实施例中SQL指令的举例示意；

[0016] 图3为一个实施例中数据查询方法的应用环境图；

[0017] 图4为一个实施例中数据查询方法的流程示意图；

[0018] 图5为一个实施例中继承关系的示例；

[0019] 图6为一个实施例中数据血缘森林的示例；

[0020] 图7为一个实施例中数据查询方法的整体流程示意图；

[0021] 图8为另一个实施例中数据查询方法的流程示意图；

[0022] 图9为一个实施例中子查询的DAG示意图；

[0023] 图10为一个实施例中数据查询装置的结构框图；

[0024] 图11为一个实施例中计算机设备的内部结构图。

具体实施方式

[0025] 为了使本申请的目的、技术方案及优点更加清楚明白，以下结合附图及实施例，对本申请进行进一步详细说明。应当理解，此处描述的具体实施例仅仅用以解释本申请，并不用于限定本申请。

[0026] 人工智能(Artificial Intelligence, AI)是利用数字计算机或者数字计算机控制的机器模拟、延伸和扩展人的智能，感知环境、获取知识并使用知识获得最佳结果的理论、方法、技术及应用系统。换句话说，人工智能是计算机科学的一个综合技术，它企图了解智能的实质，并生产出一种新的能以人类智能相似的方式做出反应的智能机器。人工智能也就是研究各种智能机器的设计原理与实现方法，使机器具有感知、推理与决策的功能。

[0027] 人工智能技术是一门综合学科，涉及领域广泛，既有硬件层面的技术也有软件层面的技术。人工智能基础技术一般包括如传感器、专用人工智能芯片、云计算、分布式存储、大数据处理技术、操作/交互系统、机电一体化等技术。人工智能软件技术主要包括计算机视觉技术、语音处理技术、自然语言处理技术以及机器学习/深度学习等几大方向。

[0028] 随着人工智能技术研究和进步，人工智能技术在多个领域展开研究和应用，例如常见的智能家居、智能穿戴设备、虚拟助理、智能音箱、智能营销、无人驾驶、自动驾驶、无人机、机器人、智能医疗、智能客服等，相信随着技术的发展，人工智能技术将在更多的领域得到应用，并发挥越来越重要的价值。

[0029] 本申请实施例提供的方案涉及人工智能的分布式存储、大数据处理技术等技术，具体通过下文实施例进行说明。

[0030] 随着各个企业和机构采集存储维护不同的数据，有越来越多的需求希望打通企业和机构间的数据墙实现更全面和准确地数据分析，例如不同的医院的病例分析、不同的交易平台的交易数据分析。而传统的OLAP需要将数据集中进行分析，这可能会导致明文数据的泄露，对企业机构的数据安全构成了威胁，此外随着《个人数据保护法》等一系列法规法律的出台、用户隐私意识的上升、企业机构间的商业利益竞争的出现，传统的OLAP已经越来越不能适应现在大数据分析的“数据孤岛”场景。在这种情况下，联合分析通过分散了分析数据的过程，由各个本地分布式的分析本地数据，并通过密码学等方法进行安全联合分析，

达到了打破数据孤岛的同时不破坏用户隐私的目的。但是目前的联合分析是在数据的计算等方面进行隐私保护，而未考虑到用户的分析指令本身可能是合法(能正确被执行)但是不合规(结果会暴露数据的隐私)。

[0031] 举例说明，如图1所示，两方分别拥有如图1中(a)图所示的客户表(Customers)和如图1中(b)图所示的订单表(Orders)，其中customer_id和order_id分别为Customers和Orders的主键。在进行安全联合分析时，用户提交了如图2中的代码所示的用于进行数据分析的查询语句，该查询语句包含了本地计算的部分和联合计算的部分。该查询语句在语法上均合法且能被正确执行，但是由于customer_id是主键，因此GROUP BY后的分组导致每一项数据都为一组，AVG()的函数的输入的每一项数据量也均为1，最后customer_id，age和amount的全量数据都被暴露在结果中，这与安全联合分析的要求相违背，导致隐私数据的安全性降低，因此这条指令是不合规的。

[0032] 本申请基于此提供了一种数据查询方法，旨在在联合分析过程中，对用户输入的查询语句中的数据列进行血缘继承分析，得到各个数据列的联合属性，进而对查询语句进行隐私校验，防止用户输入合法但不合规的查询语句，同时由于整体校验先于执行进行，可以在无需运行查询语句，即不接触数据源的原始数据的情况下高效完成，提高了联合分析过程中的数据安全性。

[0033] 本申请实施例提供的数据查询方法，可以应用于如图3所示的应用环境中。其中，第一数据源对应的终端302通过网络与第一数据源对应的服务器304进行通信，第一数据源对应的服务器304可以通过网络与一个或者多个第二数据源对应的服务器306进行通信。其中，终端302可以但不限于是各种台式计算机、笔记本电脑、智能手机、平板电脑、物联网设备和便携式可穿戴设备，物联网设备可为智能语音交互设备、智能家电、智能车载设备、飞行器等。便携式可穿戴设备可为智能手表、智能手环、头戴设备等。服务器可以是独立的物理服务器，也可以是多个物理服务器构成的服务器集群或者分布式系统，还可以是提供云服务、云数据库、云计算、云函数、云存储、网络服务、云通信、中间件服务、域名服务、安全服务、CDN、以及大数据和人工智能平台等基础云计算服务的云服务器。用户可以通过终端302向服务器304发送查询语句，指示服务器304按照该查询语句对第一数据源和第二数据源的数据进行联合分析，并返回分析结果。

[0034] 本发明实施例可应用于各种场景，包括但不限于云技术、人工智能、智慧交通、辅助驾驶等。

[0035] 在一个实施例中，如图4所示，提供了一种数据查询方法，以该方法应用于图1中的服务器304为例进行说明，包括以下步骤：

[0036] 步骤402，获取针对至少两个数据源的待执行的查询语句，从查询语句中提取得到联合子查询语句和各数据源各自对应的本地子查询语句。

[0037] 其中，至少两个数据源指的是两个或者两个以上的数据源，不同的数据源对应不同的数据拥有方。针对至少两个数据源的查询语句指的是用于对至少两个数据源的数据进行联合分析的SQL(Structured Query Language，结构化查询语言)指令。子查询语句指的是一个查询块，在SQL语言中，一个SELECT-FROM语句称为一个查询块。联合子查询语句指的是针对各个数据源的，且包含联邦化算子的查询语句。这里的联邦化算子可以包括涉及表连接的算子，例如UDAF on Join、UDAF on Union(Union All)、Group By、隐式表连接等。

UDAF是用户定义聚合函数,是一次作用于多个行的用户可编程例程,它返回单个聚合值作为结果。本地子查询语句指的是在数据源的本地端执行,且只包含本地执行的算子查询语句。

[0038] 具体地,服务器在获取到用户输入的查询语句后,可以对查询语句进行语法树(Syntax tree)分析,进而将查询语句拆分成子查询语句。这里的语法树,也可以称为抽象语法树,是源代码语法结构的一种抽象表示,它以树状的形式表现编程语言的语法结构,树上的每个节点都表示源代码中的一种结构。例如,对于图2所示的SQL指令,可以通过语法树分析拆分成两个本地子查询语句:(SELECT * FROM Customers) CusTab和(SELECT * FROM Orders)OrdTab,以及一个联合子查询语句SELECT CusTab.customer_id, Avg(CusTab.age), AVG(OrdeTab.amount) FROM CusTab FULL JOIN OrdTab ON CustTab.customer_id = OrdTab.customer_id GROUP BY CusTab.customer_id。

[0039] 在一个实施例中,联合子查询语句可以包括多个。各个联合子查询语句可以按照所包括的select算子在SQL指令中的执行顺序形成层级关系,其中,在SQL指令中越先执行的联合子查询语句层级越高。例如,假设某个SQL指令的联合执行部分包括select COLA from (select COLB from TABA)TMPTAB,则可以得到两个联合子查询语句(select COLB from TABA)TMPTAB,select COLA from TMPTAB,其中,(select COLB from TABA)TMPTAB的层级高于select COLA from TMPTAB,即(select COLB from TABA)TMPTAB先于select COLA from TMPTAB执行。

[0040] 在一个实施例中,每个数据源对应的本地子查询语句也可以包括多个。每个数据源对应的多个本地子查询语句,可以按照所包括的select算子在SQL指令中的执行顺序形成层级关系。

[0041] 步骤404,针对每个本地子查询语句,确定用于记录所针对的本地子查询语句所期望查询数据的第一目标列,基于第一目标列与所属数据源中初始列之间的第一继承关系,确定第一目标列从初始列的联合属性中继承得到的第一属性;联合属性用于对初始列进行针对各数据源的联合描述。

[0042] 其中,本地子查询语句所期望查询数据指的是本地子查询语句执行后所期望得到的数据,例如图2中的SQL所提取到的本地子查询语句(SELECT * FROM Customers) CusTab的期望查询数据为Customers表中的所有数据。第一目标列用于记录所针对的本地子查询语句所期望查询数据。

[0043] 第一继承关系为第一目标列与所属数据源中初始列之间的数据血缘继承关系,这里的初始列主要考虑以下两种中的至少一种:被列选择算子即select所作用的初始列,以及被分组算子GROUP BY所作用的初始列。第一继承关系可以是直接继承关系或者间接继承关系,当某个数据源对应的本地子查询语句只有一层时,该本地子查询语句对应的第一目标列与所属数据源中初始列之间的第一继承关系为直接继承关系;当某个数据源对应的本地子查询语句包括多个层级时,只有层级最高的本地子查询语句对应的第一目标列与所属数据源中初始列之间的第一继承关系为直接继承关系,其他的本地子查询语句对应的第一目标列与所属数据源中初始列之间的第一继承关系为间接继承关系。

[0044] 第一继承关系包括来源继承关系、权限继承关系、主键继承关系中的至少一种。在不同的继承关系下,第一目标列可以从所属数据源中初始列的联合属性中继承得到该继承

关系对应的属性。这里的初始列是用于记录数据源中数据的列，例如图1中数据源Customers包括三个初始列customers_id,first_name,age。联合属性用于对初始列进行针对各数据源的联合描述，包括数据来源、联合操作权限信息、主键信息指示信息中的至少一种，例如，对于初始列customers_id,其联合属性包括:数据来源于Customers中customers_id列，具有联合操作权限，包含主键。

[0045] 具体地，针对每个本地子查询语句，服务器可以通过语法树分析确定该本地子查询语句新生成列所属数据表的表名以及一个或多个列名，每一个表名和一个列名可以确定一个第一目标列，从而可以确定一个或者多个第一目标列。对于每一个第一目标列，可以对该第一目标列进行数据血缘分析，确定该第一目标列与所属数据源中哪些初始列之间有继承关系，基于继承关系可以确定该第一目标列从这些有继承关系的初始列的联合属性中继承得到的第一属性。可以理解的是，由于第一属性是从联合属性中继承得到的，因此第一属性也是联合属性。

[0046] 其中，对于一个本地子查询语句:如果被SELECT的列AS成新列，那么AS成的新列即为第一目标列的列名，例如，假设一个本地子查询语句为SELECT colA AS colB FROM tabC WHERE expressions on colD GROUP BY colE,则colB即为第一目标列的列名;如果被SELECT的列没有AS成新列，则默认新列的列名和原名一样，例如，假设一个本地子查询语句为SELECT colA FROM tabC WHERE expressions on colD GROUP BY colE,则colA即第一目标列的列名;如果有AS成新表的操作，则AS后接的即为新表的表名，新表的表名即第一目标列所属数据表的表名，这里的AS在有些查询语句中可以省略，例如图2所示的SQL指令提取得到的本地子查询语句(SELECT * FROM Customers) CusTab中，CusTab即为第一目标列所属数据表的表名;如果没有AS成新表的操作，则构建临时表作为第一目标列所属数据表的表名，例如可以构建TmpTab作为表名。

[0047] 以图2所示的SQL指令提取得到的本地子查询语句(SELECT * FROM Customers) CusTab为例，可以确定3个第一目标列，即CusTab.customer_id,CusTab. first_name,CusTab. age。需要说明的是，本实施例是在SQL指令执行之前对SQL进行的分析，因此，这里的3个第一目标列并没有真正生成，只是对这3个列的一种表示。

[0048] 在一个实施例中，若某个数据源对应多个具有层级关系的本地子查询语句，则，则首先确定层级最高的本地子查询语句对应的第一目标列，基于该第一目标列与所属数据源中初始列之间的第一继承关系，确定该第一目标列从初始列的联合属性中继承得到的第一属性;确定下一层级的本地子查询语句对应的第一目标列，基于该第一目标列与上一层级的本地子查询语句对应的第一目标列之间的第一继承关系，确定该第一目标列从上一层级的本地子查询语句对应的第一目标列的第一属性中继承得到的第一属性，重复该过程，直至获得最后一个层级的本地子查询语句对应的第一目标列的第一属性。

[0049] 步骤406，确定用于记录联合子查询语句所期望查询数据的第二目标列，基于第二目标列与各第一目标列之间的第二继承关系，确定第二目标列从各第一目标列的第一属性中继承得到的第二属性。

[0050] 其中，联合子查询语句所期望查询数据指的是联合子查询语句执行后所期望得到的数据，例如图2中的SQL所提取到的联合子查询语句SELECT CusTab.customer_id, Avg (CusTab.age), AVG(OrdeTab.amount) FROM CusTab FULL JOIN OrdTab ON

CustTab.customer_id = OrdTab.customer_id GROUP BY CusTab.customer_id的期望查询数据为CusTab.customer_id, Avg(CusTab.age), AVG(OrdeTab.amount)这三个列中符合ON和GROUP BY条件的数据。第二目标列用于记录联合子查询语句所期望查询数据。

[0051] 第二继承关系为第二目标列与第一目标列之间的数据血缘继承关系,这里的第一目标列主要考虑以下两种中的至少一种:被列选择算子即select所作用的第一目标列,以及被分组算子GROUP BY所作用的第一目标列。第二继承关系包括来源继承关系、权限继承关系、主键继承关系中的至少一种。在不同的继承关系下,第二目标列可以从第一目标列的第一属性中继承得到该继承关系对应的属性。第二继承关系可以是直接继承关系或者间接继承关系,当联合子查询语句只有一层时,该联合子查询语句对应的第二目标列与第一目标列之间的第二继承关系为直接继承关系;当联合子查询语句包括多个层级时,只有层级最高的联合子查询语句对应的第二目标列与第一目标列之间的第二继承关系为直接继承关系,其他的联合子查询语句对应的第二目标列与第一目标列之间的第二继承关系为间接继承关系。

[0052] 具体地,服务器可以通过语法树分析确定联合子查询语句新生成列所属数据表的表名以及一个或多个列名,每一个表名和一个列名可以确定一个第一目标列,从而可以确定一个或者多个第二目标列。对于每一个第二目标列,可以对该第二目标列进行数据血缘分析,确定该第二目标列与第一目标列中哪些列之间有继承关系,基于继承关系可以确定该第二目标列从这些有继承关系的第一目标列中的第一属性中继承得到的第二属性。可以理解的是,这里的第二属性也是联合属性。

[0053] 其中,对于一个联合子查询语句:如果被SELECT的列AS成新列,那么AS成的新列即为第二目标列的列名;如果被SELECT的列没有AS成新列,则默认新列的列名和原名一样;如果有AS成新表的操作,则AS后接的即为新表的表名,新表的表名即第二目标列所属数据表的表名;如果没有AS成新表的操作,则构建临时表作为第一目标列所属数据表的表名。以图2所示的SQL指令提取得到的联合子查询语句SELECT CusTab.customer_id, Avg(CusTab.age), AVG(OrdeTab.amount) FROM CusTab FULL JOIN OrdTab ON CustTab.customer_id = OrdTab.customer_id GROUP BY CusTab.customer_id为例,该子查询语句中没有AS成新表的操作,则可以构建临时表TmpTab作为第二目标列的表名,从而可以确定3个第二目标列分别为:TmpTab.customer_id,TmpTab. amount,TmpTab. age。

[0054] 在一个实施例中,若联合子查询语句包括具有层级关系的联合子查询语句,则首先确定层级最高的联合子查询语句对应的第二目标列,基于该第二目标列与第一目标列之间的第二继承关系,确定该第二目标列从第一目标列的第一属性中继承得到的第二属性;确定下一层级的联合子查询语句对应的第二目标列,基于该第二目标列与上一层级的联合子查询语句对应的第二目标列之间的第二继承关系,确定该第二目标列从上一层级的联合子查询语句对应的第二目标列的第二属性中继承得到的第二属性,重复该过程,直至获得最后一个层级的联合子查询语句对应的第二目标列的第二属性。

[0055] 步骤408,基于第一属性和第二属性对联合子查询语句进行校验,当校验结果指示联合子查询语句的执行对数据源的私有数据造成暴露时,屏蔽对查询语句的执行。

[0056] 其中,数据源的私有数据即数据源的隐私数据。数据源的私有数据可以是当前数据源有的而其他数据源没有的数据,例如,图1中对于数据源Costomers,其私有数据包括

first_name中的数据,age中的数据以及costomer_id中与Orders中不相同的数据。数据源的私有数据还可以是数据源的数据拥有方不希望暴露给其他数据拥有方的数据。

[0057] 具体地,由于第二目标列的第二属性是继承自第一目标列的第一属性的,而第一目标列的第一属性是继承自初始列的联合属性的,因此第二目标列的第二属性也是联合属性,可以对第二目标列进行针对各数据源的联合描述。基于此,服务器可以通过第一目标列的第一属性和第二目标列的第二属性对联合子查询语句进行校验,当校验结果指示联合子查询语句的执行对任意一个数据源的私有数据造成暴露时,表示待执行的查询语句是不合规的,服务器可以拒绝执行该查询语句,并向用户返回查询语句不合规的提示信息。

[0058] 可以理解的是,这里对数据源的私有数据进行暴露可以是对数据源的一部分私有数据进行暴露。还可以理解的是,当校验结果指示联合子查询语句的执行未对任何数据源的私有数据造成暴露时,服务器可以对查询语句进行执行。

[0059] 上述数据查询方法中,在获取到针对至少两个数据源的待执行的查询语句,从查询语句中提取得到联合子查询语句和各数据源各自对应的本地子查询语句,针对每个本地子查询语句,确定用于记录所针对的本地子查询语句所期望查询数据的第一目标列,基于第一目标列与所属数据源中初始列之间的第一继承关系,确定第一目标列从初始列的联合属性中继承得到的第一属性,确定用于记录联合子查询语句所期望查询数据的第二目标列,基于第二目标列与各第一目标列之间的第二继承关系,确定第二目标列从各第一目标列的第一属性中继承得到的第二属性,基于第一属性和第二属性对联合子查询语句进行校验,当校验结果指示联合子查询语句的执行对数据源的私有数据造成暴露时,屏蔽对查询语句的执行,从而可以实现通过继承关系的分析,在SQL指令执行之前,对SQL指令是否合规进行校验,对不合规的SQL指令屏蔽执行,提高了联合分析过程中的数据安全性。

[0060] 在一个实施例中,基于第二目标列与各第一目标列之间的第二继承关系,确定第二目标列从各第一目标列的第一属性中继承得到的第二属性包括:从各第一目标列中确定联合子查询语句中的列选择算子所作用的列,得到目标选择列;基于第二目标列与目标选择列之间的第二继承关系,确定第二目标列从目标选择列的第一属性中继承得到的第二属性。

[0061] 其中,列选择算子指的是select算子,列选择算子所作用的列指的是直接接在select算子后面的列,即在SELECT colA AS colB FROM tabC中,colA为select算子所作用的列。SELECT算子后AS成的新列(即colB)与select算子所作用的列之间有继承关系,即colB可以继承colA的属性。需要说明的是,此处的继承映射为一一对应,每个colB中的列对应继承一个colA中的列,以图中的SQL指令为例:colA可以为{CusTab.customer_id,OrdTab. amount},colB为{TmpTab.customer_id,TmpTab.amount},其中,TmpTab.customer_id继承CusTab.customer_id的属性,TmpTab.amount继承OrdTab.amount的属性。

[0062] 具体地,联合子查询语句的执行顺序在本地子查询语句之后,是在本地子查询语句所生成的第一目标列的基础上执行的,因此服务器可以从第一目标列中确定联合子查询语句中的列选择算子所作用的列,得到目标选择列,进而基于第二目标列与目标选择列之间的第二继承关系,确定第二目标列从目标选择列的第一属性中继承得到的第二属性。

[0063] 上述实施例中,通过确定列选择算子所作用的列,得到目标选择列,基于与目标选

择列之间的继承关系，确定从目标选择列的第一属性中继承得到的属性，确保了继承得到的属性的准确性。

[0064] 在一个实施例中，基于第二目标列与目标选择列之间的第二继承关系，确定第二目标列从目标选择列的第一属性中继承得到的第二属性，包括：当第二继承关系包括来源继承关系时，确定第二目标列从目标选择列的第一属性中继承得到的第二属性包括数据来源信息；当第二继承关系包括权限继承关系时，确定第二目标列从目标选择列的第一属性中继承得到的第二属性包括联合操作权限信息；当第二继承关系包括主键继承关系时，确定第二目标列从目标选择列的第一属性中继承得到的第二属性包括主键信息。

[0065] 其中，第二继承关系包括来源继承关系、权限继承关系、主键继承关系中的至少一种。具体地，对形如SELECT colA AS colB FROM tabC WHERE expressions on colD GROUPBY colE的子查询语句，这里expressions on colD为colD中所有列的表达式：当第二继承关系包括来源继承关系时，colB继承得到的第二属性中包括colA的数据来源；当第二继承关系包括权限继承关系时，colB继承得到的第二属性中包括colA的联合操作权限信息，联合操作权限信息用于指示colA是否具有联合操作权限；当第二继承关系包括主键继承关系时，colB继承得到的第二属性中包括所有colA继承的主键集合，若colA同时也为主键，则也加入到colB的主键集合中。

[0066] 上述实施例中，由于第二继承关系可以包括多种类型，因此，第二目标列可以从目标选择列的第一属性中继承得到丰富的属性信息，使得得到的第二属性更加准确全面。

[0067] 在一个实施例中，基于第二目标列与各第一目标列之间的第二继承关系，确定第二目标列从各第一目标列的第一属性中继承得到的第二属性包括：从各第一目标列中确定联合子查询语句中的分组算子所作用的列，得到目标分组列；基于第二目标列与目标分组列之间的第二继承关系，确定第二目标列从目标分组列的第一属性中继承得到的第二属性。

[0068] 其中，分组算子指的是GROUP BY算子，分组算子所作用的列指的是直接接在GROUP BY算子后面的列，即在SELECT colA AS colB FROM tabC WHERE expressions on colD GROUPBY colE中，colE为GROUP BY算子所作用的列。SELECT算子后AS成的新列与GROUPBY算子所作用的列之间有继承关系，即colB可以继承colE的属性。需要说明的是，colB继承自colE，此处为笛卡尔积关系，每一个colB的列与colE中所有列有继承关系。例如colB为{TmpTab.customer_id, TmpTab.amount}，colE为{ CusTab.customer_id}，则TmpTab.customer_id ，TmpTab.amount都有CusTab.customer_id 有继承关系。

[0069] 具体地，联合子查询语句的执行顺序在本地子查询语句之后，是在本地子查询语句所生成的第一目标列的基础上执行的，因此服务器可以从第一目标列中确定联合子查询语句中的分组算子所作用的列，得到目标分组列，进而基于第二目标列与目标分组列之间的第二继承关系，确定第二目标列从目标分组列的第一属性中继承得到的第二属性。

[0070] 上述实施例中，通过确定分组算子所作用的列，得到目标分组列，基于与目标分组列之间的继承关系，确定从目标分组列的第一属性中继承得到的属性，确保了继承得到的属性的准确性。

[0071] 在一个实施例中，基于第二目标列与目标分组列之间的第二继承关系，确定第二目标列从目标分组列的第一属性中继承得到的第二属性包括：当第二继承关系包括来源继

承关系时，确定第二目标列从目标分组列的第一属性中继承得到的第二属性包括数据来源信息；当第二继承关系包括权限继承关系时，确定第二目标列从目标分组列的第一属性中继承得到的第二属性包括联合操作权限信息；当第二继承关系包括主键继承关系且第二目标列的列名与目标分组列的列名一致时，确定第二目标列从目标分组列的第一属性中继承得到的第二属性包括主键信息。

[0072] 具体地，对形如SELECT colA AS colB FROM tabC WHERE expressions on colD GROUPBY colE的子查询语句：当第二继承关系包括来源继承关系时，colB继承得到的第二属性中包括colE的数据来源；当第二继承关系包括权限继承关系时，colB继承得到的第二属性中包括colE的联合操作权限信息，联合操作权限信息用于指示colB是否具有联合操作权限；当第二继承关系包括主键继承关系且colB的列名与所colE的列名一致时，colB继承得到的第二属性中包括所有colE继承的主键集合，若colE同时也为主键，则也加入到colB的主键集合中。

[0073] 举例说明，参考图5，假设子查询语句为SELECT CusTab.customer_id, OrdTab.amount FROM CusTab, OrdTab GROUP BY CusTab.customer_id) TmpTable，则继承关系可以参考图5所示。其中，箭头表示来源继承，锁表示权限继承，*表示主键继承。由图5可以看出，TmpTable. customer_id、TmpTable. amount均继承了CusTab.customer_id的数据来源，TmpTable. customer_id、TmpTable. amount均继承了CusTab.customer_id的联合操作权限，只有TmpTable. customer_id继承了CusTab.customer_id的主键信息。

[0074] 上述实施例中，由于第二继承关系可以包括多种类型，因此，第二目标列可以从目标选择列的第一属性中继承得到丰富的属性信息，使得得到的第二属性更加准确全面。

[0075] 在一个实施例中，基于第一目标列与所属数据源中初始列之间的第一继承关系，确定第一目标列从初始列的联合属性中继承得到的第一属性包括：从初始列中确定所针对的本地子查询语句包括的列选择算子所作用的列，得到目标选择列；基于第一目标列与目标选择列之间的第一继承关系，确定第一目标列从目标选择列的联合属性中继承得到的第一属性。

[0076] 具体地，服务器可以从初始列中确定所针对的本地子查询语句中的SELECT算子所作用的列，得到目标选择列，进而基于第一目标列与目标选择列之间的第一继承关系，确定第一目标列从目标选择列的联合属性中继承得到的第一属性。

[0077] 在一个实施例中，基于第一目标列与目标选择列之间的第一继承关系，确定第一目标列从目标选择列的联合属性中继承得到的第一属性，包括：当第一继承关系包括来源继承关系时，确定第一目标列从目标选择列的第一属性中继承得到的第一属性包括数据来源信息；当第一继承关系包括权限继承关系时，确定第一目标列从目标选择列的第一属性中继承得到的第一属性包括联合操作权限信息；当第一继承关系包括主键继承关系时，确定第一目标列从目标选择列的第一属性中继承得到的第一属性包括主键信息。

[0078] 本实施例中，第一继承关系包括来源继承关系、权限继承关系、主键继承关系中的至少一种。具体地，对形如SELECT colA AS colB FROM tabC WHERE expressions on colD GROUPBY colE的本地子查询语句：当第一继承关系包括来源继承关系时，colB继承得到的第一属性中包括colA的数据来源；当第一继承关系包括权限继承关系时，colB继承得到的第一属性中包括colA的联合操作权限信息，联合操作权限信息用于指示colA是否具有联合

操作权限;当第一继承关系包括主键继承关系时,colB继承得到的第一属性中包括所有colA继承的主键集合,若colA同时也为主键,则也加入到colB的主键集合中。

[0079] 上述实施例中,通过确定列选择算子所作用的列,得到目标选择列,基于与目标选择列之间的继承关系,确定从目标选择列的联合属性中继承得到的属性,确保了继承得到的属性的准确性。

[0080] 在一个实施例中,所针对的本地子查询语句还包括分组算子,基于第一目标列与所属数据源中初始列之间的第一继承关系,确定第一目标列从初始列的联合属性中继承得到的第一属性还包括:从初始列中确定分组算子所作用的列,得到目标分组列;基于第一目标列与目标分组列之间的第一继承关系,确定第一目标列从目标分组列的联合属性中继承得到的第一属性。

[0081] 具体地,服务器可以从初始列中确定本地子查询语句中的GROUP BY算子所作用的列,得到目标分组列,进而基于第一目标列与目标分组列之间的第一继承关系,确定第一目标列从目标分组列的联合属性中继承得到的第一属性。

[0082] 在一个实施例中,基于第一目标列与目标分组列之间的第一继承关系,确定第一目标列从目标分组列的联合属性中继承得到的第一属性包括:当第一继承关系包括来源继承关系时,确定第一目标列从目标分组列的第一属性中继承得到的第一属性包括数据来源信息;当第一继承关系包括权限继承关系时,确定第一目标列从目标分组列的第一属性中继承得到的第一属性包括联合操作权限信息;当第一继承关系包括主键继承关系且第一目标列的列名与目标分组列的列名一致时,确定第一目标列从目标分组列的第一属性中继承得到的第一属性包括主键信息。

[0083] 本实施例中,第一继承关系包括来源继承关系、权限继承关系、主键继承关系中的至少一种。对形如SELECT colA AS colB FROM tabC WHERE expressions on colD GROUPBY colE的子查询语句:当第一继承关系包括来源继承关系时,colB继承得到的第一属性中包括colE的数据来源;当第一继承关系包括权限继承关系时,colB继承得到的第一属性中包括colE的联合操作权限信息,联合操作权限信息用于指示colB是否具有联合操作权限;当第一继承关系包括主键继承关系且colB的列名与所colE的列名一致时,colB继承得到的第一属性中包括所有colE继承的主键集合,若colE同时也为主键,则也加入到colB的主键集合中。

[0084] 上述实施例中,通过确定分组算子所作用的列,得到目标分组列,基于与目标分组列之间的继承关系,确定从目标分组列的第一属性中继承得到的属性,确保了继承得到的属性的准确性。

[0085] 在一个实施例中,基于第一属性和第二属性对联合子查询语句进行校验包括:从联合子查询语句中确定所包含的算子;针对每个算子,确定所针对的算子所作用的列,基于第一属性和第二属性确定所针对的算子所作用的列的目标属性;基于所针对的算子对应的校验规则对目标属性进行校验,得到校验结果。

[0086] 具体地,考虑到联合子查询语句的执行即为联合子查询语句所包含的算子的执行,因此对联合子查询语句的校验即对联合子查询语句中各个算子进行校验,联合子查询语句中各个算子所作用的列要么是第一目标列要么是第二目标列,从而在获得到了各个第一目标列的第一属性以及各个第二目标列的第二属性后,服务器即可以确定联合子查询语

句中各个算子所作用的列的目标属性，本实施例中，预先对各种可能涉及的算子设置了校验规则，从而对于联合子查询语句中确定所包含的各个算子，可以基于各个算子各自对应的校验规则对各个算子各自所作用的列的目标属性进行校验，得到校验结果。

[0087] 在具体应用中，联合子查询语句中所包含的算子可以包括列选择算子、分组算子以及条件过滤算子中的至少一种。以下实施例将针对各种类型算子的校验进行说明。

[0088] 在一个实施例中，针对每个算子，确定所针对的算子所作用的列，基于第一属性和第二属性确定所针对的算子所作用的列的目标属性，包括：针对列选择算子，确定列选择算子所作用的列，基于第一属性和第二属性确定列选择算子所作用的列的联合操作权限信息；基于所针对的算子对应的校验规则对目标属性进行校验，得到校验结果，包括：当列选择算子所作用的列的联合操作权限信息指示不具备联合操作权限，得到第一校验结果，第一校验结果指示列选择算子的执行对数据源的私有数据造成暴露。

[0089] 具体地，对于列选择算子，设置的校验规则为联合子查询语句中的列选择算子所作用的列必须具备联合操作权限，若联合子查询语句中的列选择算子所作用的列不具备联合操作权限，则该联合子查询语句的执行必然会导致数据源的私有数据的暴露，因此该列选择算子是不合规的。这里，某个列具备联合操作权限指的是可以该列可以被联合子查询语句中的算子所作用，也就是说该列的数据可以在联合分析的过程中进行公开。列选择算子所作用的列可以是列选择算子直接作用的列，同时考虑到列选择算子所得到的新列与列选择算子所作用的列之间具有一对一的继承关系，这里列选择算子所作用的列还可以是列选择算子所得到的新列。即对于形如SELECT colA AS colB FROM tabC WHERE expressions on colD GROUP BY colE的联合子查询语句，列选择算子所作用的列可以是colA，也可以是colB。

[0090] 本实施例中，服务器可以基于第一属性和第二属性确定列选择算子所作用的列的联合操作权限信息，联合操作权限信息可以指示该列选择算子所作用的列具备联合操作权限或者不具备联合操作权限，当列选择算子所作用的列的联合操作权限信息指示不具备联合操作权限，则得到第一校验结果，该第一校验结果是指示该列选择算子不合规的，即列选择算子的执行会对数据源的私有数据造成暴露。可以理解的是，当列选择算子所作用的列的联合操作权限信息指示具备联合操作权限，则得到的校验结果是指示该列选择算子合规的，即列选择算子的执行不会对数据源的私有数据造成暴露。

[0091] 在一个实施例中，针对每个算子，确定所针对的算子所作用的列，基于第一属性和第二属性确定所针对的算子所作用的列的目标属性，包括：针对分组算子，确定分组算子所作用的列，基于第一属性和第二属性确定分组算子所作用的列的主键信息；基于所针对的算子对应的校验规则对目标属性进行校验，得到校验结果，包括：当分组算子所作用的列的主键信息指示包含所属数据源的联合主键，得到第二校验结果，第二校验结果指示联合子查询语句中分组算子的执行对数据源的私有数据造成暴露。

[0092] 具体地，对于分组算子，设置地校验规则为联合子查询语句中的分组算子所作用的列不能包含所属数据源中的联合主键，若联合子查询语句中的分组算子所作用的列包含所属数据源中的联合主键，则该联合子查询语句的执行必然会导致数据源的私有数据的暴露，因此该分组算子是不合规的。这里，联合主键指的是分组算子所作用的列所属数据源中的联合主键。

[0093] 本实施例中,服务器可以基于第一属性和第二属性确定分组算子所作用的列的主键信息,联合操作权限信息可以指示该分组算子所作用的列是否包含所属数据源的所有联合主键,当分组算子所作用的列的包含所属数据源的所有联合主键,则得到第二校验结果,该第一校验结果是指示该分组算子不合规的,即分组算子的执行会对数据源的私有数据造成暴露。可以理解的是,当分组算子所作用的列的主键信息指示不包含所属数据源的所有联合主键,则得到的校验结果是指示该分组算子合规的,即分组算子的执行不会对数据源的私有数据造成暴露。

[0094] 在一个实施例中,针对每个算子,确定所针对的算子所作用的列,基于第一属性和第二属性确定所针对的算子所作用的列的目标属性,包括:针对条件过滤算子,确定条件过滤算子所作用的列,基于第一属性和第二属性确定条件过滤算子所作用的列的联合操作权限信息;基于所针对的算子对应的校验规则对目标属性进行校验,得到校验结果,包括:当条件过滤算子所作用的列的联合操作权限信息指示不具备联合操作权限,得到第三校验结果,第三校验结果指示条件过滤算子的执行对数据源的私有数据造成暴露。

[0095] 其中,条件过滤算子是用于对列选择算子所作用的列进行数据过滤的,列选择算子可以包括where算子和on算子中的其中一种。条件过滤算子所作用的列指的是直接跟在条件过滤算子后的列,例如对于图2的SQL指令提取得到的联合子查询语句SELECT CusTab.customer_id, Avg(CusTab.age), AVG(OrdeTab.amount) FROM CusTab FULL JOIN OrdTab ON CustTab.customer_id = OrdTab.customer_id GROUP BY CusTab.customer_id,其中on算子所作用的列包括CustTab.customer_id和OrdTab.customer_id。

[0096] 具体地,对于条件过滤算子,设置的校验规则为联合子查询语句中的条件过滤算子所作用的列必须具备联合操作权限,若联合子查询语句中的条件过滤算子所作用的列不具备联合操作权限,则该联合子查询语句的执行必然会导致数据源的私有数据的暴露,因此该条件过滤算子是不合规的。

[0097] 本实施例中,服务器可以基于第一属性和第二属性确定条件过滤算子所作用的列的联合操作权限信息,联合操作权限信息可以指示该条件过滤算子所作用的列具备联合操作权限或者不具备联合操作权限,当条件过滤算子所作用的列的联合操作权限信息指示不具备联合操作权限,则得到第一校验结果,该第一校验结果是指示该条件过滤算子不合规的,即条件过滤算子的执行会对数据源的私有数据造成暴露。可以理解的是,当条件过滤算子所作用的列的联合操作权限信息指示具备联合操作权限,则得到的校验结果是指示该条件过滤算子合规的,即条件过滤算子的执行不会对数据源的私有数据造成暴露。

[0098] 上述实施例中,通过设置不同的校验规则,对联合子查询语句中各个算子进行校验,可以准确快速地判断出联合子查询语句的执行是否对数据源的私有数据造成暴露。

[0099] 在一个实施例中,其特征在于,方法还包括:针对各数据源中初始列生成对应的初始节点,将各初始列的联合属性记录在各自对应的初始节点中;基于第一目标列与所属数据源中初始列之间的第一继承关系,确定第一目标列从初始列的联合属性中继承得到的第一属性包括:针对第一目标列生成对应的第一节点, 基于第一目标列与初始列之间的第一继承关系,建立第一节点和初始节点之间的第一连接关系,基于第一连接关系确定第一节点从初始节点所记录的联合属性中继承得到的第一属性,并记录在第一节点中;基于第二

目标列与各第一目标列之间的第二继承关系，确定第二目标列从各第一目标列的第一属性中继承得到的第二属性，包括：针对第二目标列生成对应的第二节点，基于第二目标列与各第一目标列之间的第二继承关系，建立第一节点和第二节点之间的第二连接关系，以构建数据血缘森林，基于第二连接关系确定第二节点从各第一节点所记录的第一属性中继承得到的第二属性，并记录在第二节点中。

[0100] 具体地，由前文实施例可知，本申请是在SQL指令执行之前对SQL指令进行数据血缘分析，由于SQL指令还未执行，上文提到的第一目标列以及第二目标列并不是实际创建的列，仅仅是用于表征这些列，本实施例中，为了更好地表征这些列，可以针对每一个第一目标列和每一个第二目标列生成对应的节点，以节点的形式表示这些列，节点的标识即为这些列的全局唯一标识，例如，列的全局唯一标识可以是表名加列名的组合。各个列对应的节点中可以记录各自对应的联合属性。

[0101] 举例说明，假设待执行的查询语句为如图2所示的SQL指令，则可以生成如图6所示的数据血缘森林。参考图6，其中，Customers.customer_id、Customers.fisrst_name、Customers.age为数据源Customers中各个初始列对应的节点，Orders.order_id、Orders.item 、Orders.amount、Orders. customer_id为数据源Orders中各个初始列对应的节点，CusTab.customer_id、CusTab.fisrst_name、 CusTab.age为数据源Customers对应的本地子查询语句对应的第一目标列，OrdTab.order_id、OrdTab.item 、OrdTab.amount、OrdTab. customer_id为数据源Orders对应的本地子查询语句对应的第一目标列，TmpTab.customer_id、TmpTab.AVG_age、TmpTab.AVG_amount为联合子查询语句对应的第二目标列，第二目标列的列名中将UDAF操作作为列名的一部分以区分操作前和操作后的区别。继续参考图6，图6中的连接线表示继承关系，连接线的箭头指向表示数据来源，打开的锁代表具备联合操作权限，未打开的锁代表不具备联合操作权限，*代表包含主键。由图6可以看出，第一目标列和初始列之间存在继承关系，第二目标列和第一目标列之间存在继承关系。

[0102] 需要说明的是，由图6可以看出，节点TmpTab.AVG_age继承了两种不同的联合操作权限信息，其中，从CusTab.customer_id中继承得到的联合操作权限信息是指示具备联合操作权限的，而从CusTab.age中继承得到的联合操作权限信息是指示不具备联合操作权限，最终TmpTab.AVG_age的联合操作权限信息是指示不具备联合操作权限的，也就是说当某个列继承的联合属性中包括指示不具备联合操作权限的联合操作权限信息时，该列的联合操作权限信息即为不具备联合操作权限。

[0103] 在一个实施例中，数据血缘森林中各个节点存在对应的全局唯一标识，基于第一属性和第二属性对联合子查询语句进行校验包括：从联合子查询语句中确定所包含的算子；针对每个算子，确定所针对的算子所作用的列，将列的全局唯一标识和数据血缘森林中各个节点的全局唯一标识进行匹配；获取匹配成功的节点所记录的目标属性，基于所针对的算子对应的校验规则对目标属性进行校验，得到校验结果。

[0104] 具体地，数据血缘森林各个节点存在对应的全局唯一标识，进而可以将各个算子所作用的列的全局唯一标识和数据血缘森林中各个节点各自对应的全局唯一标识进行匹配，以确定各个列的目标属性，从而可以根据校验规则对目标属性进行校验。举例说明，继续参考图6，根据图6对图2中的SQL指令进行校验，有两处不合规：1、TmpTable.AVG_age不具

备联合操作权限，违反了SELECT对应的校验规则；2、Customers.customer_id为数据源Customers中的主键，违反了GROUP BY的校验规则。

[0105] 上述实施例中，通过构建数据血缘森林，可以更加快速地确定各个算子所需要校验的目标属性，提高了校验效率。

[0106] 在一个实施例中，如图7所示，为一个具体的实施例中，数据查询方法的整体流程图。参考图7，该数据查询方法包括步骤702至步骤712，具体地：在步骤702中，服务器从配置数据中读取得到各数据源各自的初始列信息以及各个初始列对应的联合属性；在步骤704中，服务器通过对SQL指令进行数据血缘分析，确定SQL指令提取得到的各个子查询新生成的列继承得到的联合属性，在步骤706中，服务器基于步骤704中得到的联合属性对SQL指令中提取得到的联合子查询语句进行校验，在步骤708中，服务器根据校验结果判断SQL指令是否合规，在SQL指令的执行会导致数据源的私有数据暴露时，判定SQL指令不合规，进入步骤710，不执行该SQL指令，并向用户返回SQL指令不合规的提示，在SQL指令的执行不会导致数据源的私有数据暴露时，判定SQL指令合规，进入步骤712，执行该SQL指令。

[0107] 在一个实施例中，如图8所示，为一个具体的实施例中，数据查询方法的整体流程图。参考图8，该数据查询方法具体包括以下步骤：

[0108] 步骤802，初始化。

[0109] 具体地，服务器可以从数据源中读入所有列，并建立对应的列的数据血缘节点，每一个节点需要一个全局唯一标识来表征，这里可以用“表名.列名”作为全局唯一标识的一种，也可以采用别的形式。节点中记录自身的联合操作权限信息(是否具备联合操作权限)、主键信息(是否为主键)、数据源信息。

[0110] 步骤804，分析SQL，建立子查询DAG(Directed Acyclic Graph，有向无环图)。

[0111] 具体地，服务器可以通过语法树对SQL指令进行分析，提取其中的联合子查询语句和本地子查询语句，建立有向无环图。在有向无环图，各个子查询语句分别作为节点，子查询语句之间的数据依赖关系作为边，边的方向整体遵循从本地子查询语句指向联合子查询语句的原则，若是某个数据源提取得到多个具有层级关系的本地子查询语句，由于相邻层级的本地子查询语句之间是有数据依赖关系的，因此可以将各个本地子查询语句按照层级关系生成首尾相连的节点，节点之间的连接边从层级高的节点指向层级低的节点，若是联合子查询语句包括多个具有层级关系的本地子查询语句，则可以将本地子查询语句按照层级关系生成首尾相连的节点，节点之间的连接边从层级高的节点指向层级低的节点。举例说明，假设某个SQL指令如下所示，需要说明的是，该SQL指令仅作为示例，并不考虑语法正确性。

[0112] SELECT colA FROM(SELECT colB FROM

[0113] (SELECT colC FROM (SELECT * FROM tabA)TmpTab1 WHERE XXX)TmpTab2 FULL JOIN

[0114] (SELECT colD FROM (SELECT * FROM tabB) TmpTab3 WHERE XXX)TmpTab4 ON XXX)TmpTab5 WHERE XX

[0115] 以上SQL指令可以提取得到6个查询，其中数据源tabA 对应两个本地子查询语句(以下简称子查询)，分别为：子查询1：(SELECT * FROM tabA)TmpTab1，子查询2：(SELECT colC FROM TmpTab1 WHERE XXX)TmpTab2；其中数据源tabB 对应两个本地子查询语句(以

下简称子查询),分别为:子查询3:(SELECT * FROM tabB)TmpTab3,子查询4:(SELECT colD FROM TmpTab3 WHERE XXX)TmpTab4;联合子查询语句(以下简称子查询)包括两个,分别为:子查询5:(SELECT colB FROM TmpTab2 FULL JOIN TmpTab4 ON XXX)TmpTab5;子查询6:SELECT colA FROM TmpTab5。

[0116] 基于上述6个子查询可以构建如图9所示的有向无环图。

[0117] 步骤806、判断是否完成遍历对所有子查询的遍历,若否,则从DAG中遍历当前节点,针对当前节点所表示的子查询进入步骤808A,若是,则进入步骤810。

[0118] 具体地,服务器从DAG中的初始节点出发对DAG中的各个节点进行遍历,针对每一个遍历到的节点所表征的子查询,执行步骤808A-808D,然后判断是否完成对DAG的遍历,若否,则该当前节点的下一个节点作为当前节点,重复上述步骤,直至完成对所有节点的遍历。

[0119] 举例说明,如图9所示的DAG,可以按照以下顺序进行遍历,子查询1、子查询2、子查询3、子查询4、子查询5、子查询6,需要说明的是,图9中子查询1和子查询3均为初始节点,由图9可以看出,遍历的顺序是先遍历本地子查询,再遍历联合子查询;在遍历本地子查询的过程中,先遍历层级高的本地子查询,再遍历层级低的本地子查询,即按照层级从高至低进行遍历;在遍历联合子查询的过程中,先遍历层级高的联合子查询,再遍历层级低的联合子查询,即按照层级从高至低进行遍历。

[0120] 步骤808A、提取数据全局ID(即上文的全局唯一标识)。针对每一个新生成的列,从SQL指令对应的AST语法树中抽出列名和对应的表名构成全局ID。

[0121] 步骤808B、建立数据血缘节点。针对每一个新生成的列,建立新的对应的数据血缘节点。

[0122] 步骤808C、建立SELECT-FROM继承。每一个子查询中,生成的新列(以下用b指代)和SELECT所作用的列(以下用a指代)之间存在以下三种继承关系:1、来源继承:b为a的父亲节点,a为b的孩子节点;2、权限继承关系:b继承a的联合操作权限,若b的父节点中有不具备联合操作权限的节点,则b也不具备联合操作权限;3、主键继承:b继承所有a继承的主键集合,同时若a为主键,则也加入到b的主键集合中。根据这三种继承关系建立数据血缘节点之间的连接关系,并在节点中记录联合属性,其中连接关系根据来源继承确定。具体可以参考上文实施例中的描述。

[0123] 需要说明的是,此处的继承是一对一映射关系的继承,即对于每一个生成的新列,在SELECT所作用的列中存在一个与之对应的存在继承关系的列。具体可以参考上文实施例中的描述。

[0124] 步骤808D、建立GROUP BY继承。每一个子查询中,生成的新列(以下用b指代)和GROUP BY所作用的列(以下用a指代)之间存在以下三种继承关系:1、来源继承,b为a的父亲节点,a为b的孩子节点;2、权限继承关系:b继承a的联合操作权限,若b的父节点中有不具备联合操作权限的节点,则b也不具备联合操作权限;3、主键继承:b中与a列名相同的列,继承a的主键信息。根据这三种继承关系建立数据血缘节点之间的连接关系,并在节点中记录联合属性,其中连接关系根据来源继承确定。具体可以参考上文实施例中的描述。

[0125] 需要说明的是,此处的继承是笛卡尔积关系的继承,即对于每一个生成的新列,与GROUP BY所作用的列的每一个列之间存在继承关系。具体可以参考上文实施例中的描述。

[0126] 步骤810、数据隐私校验。

[0127] 通过上述步骤可以建立得到数据血缘森林，对于SQL指令提取得到的每一个联合子查询语句：从联合子查询语句中确定所包含的算子；针对每个算子，确定所针对的算子所作用的列，将列的全局唯一标识和数据血缘森林中各个节点的全局唯一标识进行匹配；获取匹配成功的节点所记录的目标属性，基于所针对的算子对应的校验规则对目标属性进行校验，得到校验结果。当校验结果指示SQL指令不合规时，不执行该SQL指令，并向用户返回SQL指令不合规的提示，当校验结果指示SQL指令合规，执行该SQL指令。其中，关于各个算子对应的校验规则，可以参考上文实施例的描述。

[0128] 上述实施例中，对用户输入的SQL指令进行数据隐私校验，在一个SQL指令被切分成多个子查询情况下，针对每个子查询构建数据血缘森林，记录包括来源的继承、权限的继承、主键的继承，最终校验子查询的语句是否合规，达到防止违反隐私规定的指令被运行的目标，本申请实施例在SQL指令执行之前进行，可以高效的完成合规检查，同时校验在不接触原始数据的情况下进行，保证了校验过程的隐私性，提高联合分析过程中隐私数据的安全性。

[0129] 在一个具体的实施例中，本申请还提供一种应用场景，在该应用场景中，上述数据查询方法可以配置于大数据计算平台，用户可以通过命令行输入SQL指令，该大数据计算平台通过执行本申请实施例提供的数据查询方法对用户输入的SQL指令进行校验。在另一个具体的实施例中，上述数据查询方法可以配置于大数据计算平台的SQL执行引擎中，用户可以通过前端页面提交SQL指令，该大数据计算平台通过SQL执行引擎对用户提交的SQL指令进行校验。

[0130] 应该理解的是，虽然如上的各实施例所涉及的流程图中的各个步骤按照箭头的指示依次显示，但是这些步骤并不是必然按照箭头指示的顺序依次执行。除非本文中有明确的说明，这些步骤的执行并没有严格的顺序限制，这些步骤可以以其它的顺序执行。而且，如上的各实施例所涉及的流程图中的至少一部分步骤可以包括多个步骤或者多个阶段，这些步骤或者阶段并不必然是在同一时刻执行完成，而是可以在不同的时刻执行，这些步骤或者阶段的执行顺序也不必然是依次进行，而是可以与其它步骤或者其它步骤中的步骤或者阶段的至少一部分轮流或者交替地执行。

[0131] 基于同样的发明构思，本申请实施例还提供了一种用于实现上述所涉及的数据查询方法的数据查询装置。该装置所提供的解决问题的实现方案与上述方法中所记载的实现方案相似，故下面所提供的一个或多个数据查询装置实施例中的具体限定可以参见上文中对于数据查询方法的限定，在此不再赘述。

[0132] 在一个实施例中，如图10所示，提供了一种数据查询装置1000，包括：

[0133] 查询语句获取模块1002，用于获取针对至少两个数据源的查询语句，从查询语句中提取得到联合子查询语句和各数据源各自对应的本地子查询语句；

[0134] 第一属性确定模块1004，用于针对每个本地子查询语句，确定用于记录所针对的本地子查询语句所期望查询数据的第一目标列，基于第一目标列与所属数据源中初始列之间的第一继承关系，确定第一目标列从初始列的联合属性中继承得到的第一属性；联合属性用于对初始列进行针对各数据源的联合描述；

[0135] 第二属性确定模块1006，用于确定用于记录联合子查询语句所期望查询数据的第

二目标列，基于第二目标列与各第一目标列之间的第二继承关系，确定第二目标列从各第一目标列的第一属性中继承得到的第二属性；

[0136] 校验模块1008，用于基于第一属性和第二属性对联合子查询语句进行校验，当校验结果指示联合子查询语句的执行对数据源的私有数据造成暴露时，屏蔽对查询语句的执行。

[0137] 上述数据查询装置，在获取到针对至少两个数据源的待执行的查询语句，从查询语句中提取得到联合子查询语句和各数据源各自对应的本地子查询语句，针对每个本地子查询语句，确定用于记录所针对的本地子查询语句所期望查询数据的第一目标列，基于第一目标列与所属数据源中初始列之间的第一继承关系，确定第一目标列从初始列的联合属性中继承得到的第一属性，确定用于记录联合子查询语句所期望查询数据的第二目标列，基于第二目标列与各第一目标列之间的第二继承关系，确定第二目标列从各第一目标列的第一属性中继承得到的第二属性，基于第一属性和第二属性对联合子查询语句进行校验，当校验结果指示联合子查询语句的执行对数据源的私有数据造成暴露时，屏蔽对查询语句的执行，从而可以实现通过继承关系的分析，在SQL指令执行之前，对SQL指令是否合规进行校验，对不合规的SQL指令屏蔽执行，提高了联合分析过程中的数据安全性。

[0138] 在一个实施例中，第二属性确定模块，用于从各第一目标列中确定联合子查询语句中的列选择算子所作用的列，得到目标选择列；基于第二目标列与目标选择列之间的第二继承关系，确定第二目标列从目标选择列的第一属性中继承得到的第二属性。

[0139] 在一个实施例中，第二属性确定模块，还用于当第二继承关系包括来源继承关系时，确定第二目标列从目标选择列的第一属性中继承得到的第二属性包括数据来源信息；当第二继承关系包括权限继承关系时，确定第二目标列从目标选择列的第一属性中继承得到的第二属性包括联合操作权限信息；当第二继承关系包括主键继承关系时，确定第二目标列从目标选择列的第一属性中继承得到的第二属性包括主键信息。

[0140] 在一个实施例中，第二属性确定模块，还用于从各第一目标列中确定联合子查询语句中的分组算子所作用的列，得到目标分组列；基于第二目标列与目标分组列之间的第二继承关系，确定第二目标列从目标分组列的第一属性中继承得到的第二属性。

[0141] 在一个实施例中，第二属性确定模块，还用于当第二继承关系包括来源继承关系时，确定第二目标列从目标分组列的第一属性中继承得到的第二属性包括数据来源信息；当第二继承关系包括权限继承关系时，确定第二目标列从目标分组列的第一属性中继承得到的第二属性包括联合操作权限信息；当第二继承关系包括主键继承关系且第二目标列的列名与目标分组列的列名一致时，确定第二目标列从目标分组列的第一属性中继承得到的第二属性包括主键信息。

[0142] 在一个实施例中，第一属性确定模块，用于从初始列中确定所针对的本地子查询语句包括的列选择算子所作用的列，得到目标选择列；基于第一目标列与目标选择列之间的第一继承关系，确定第一目标列从目标选择列的联合属性中继承得到的第一属性。

[0143] 在一个实施例中，第一属性确定模块，用于从初始列中确定分组算子所作用的列，得到目标分组列；基于第一目标列与目标分组列之间的第一继承关系，确定第一目标列从目标分组列的联合属性中继承得到的第一属性。

[0144] 在一个实施例中，校验模块，还用于从联合子查询语句中确定所包含的算子针对

每个算子，确定所针对的算子所作用的列，基于第一属性和第二属性确定所针对的算子所作用的列的目标属性；基于所针对的算子对应的校验规则对目标属性进行校验，得到校验结果。

[0145] 在一个实施例中，校验模块，还用于针对列选择算子，确定列选择算子所作用的列，基于第一属性和第二属性确定列选择算子所作用的列的联合操作权限信息；当列选择算子所作用的列的联合操作权限信息指示不具备联合操作权限，得到第一校验结果，第一校验结果指示列选择算子的执行对数据源的私有数据造成暴露。

[0146] 在一个实施例中，校验模块，还用于针对分组算子，确定分组算子所作用的列，基于第一属性和第二属性确定分组算子所作用的列的主键信息；当分组算子所作用的列的主键信息指示包含所属数据源的联合主键，得到第二校验结果，第二校验结果指示联合子查询语句中分组算子的执行对数据源的私有数据造成暴露。

[0147] 在一个实施例中，校验模块，还用于针对条件过滤算子，确定条件过滤算子所作用的列，基于第一属性和第二属性确定条件过滤算子所作用的列的联合操作权限信息；当条件过滤算子所作用的列的联合操作权限信息指示不具备联合操作权限，得到第三校验结果，第三校验结果指示条件过滤算子的执行对数据源的私有数据造成暴露。

[0148] 在一个实施例中，上述装置还用于：针对各数据源中初始列生成对应的初始节点，将各初始列的联合属性记录在各自对应的初始节点中；第一属性确定模块，还用于针对第一目标列生成对应的第一节点，基于第一目标列与初始列之间的第一继承关系，建立第一节点和初始节点之间的第一连接关系，基于第一连接关系确定第一节点从初始节点所记录的联合属性中继承得到的第一属性，并记录在对应的第一节点中；第二属性确定模块，还用于针对第二目标列生成对应的第二节点，基于第二目标列与各第一目标列之间的第二继承关系，建立第一节点和第二节点之间的第二连接关系，以构建数据血缘森林，基于第二连接关系确定第二节点从各第一节点所记录的第一属性中继承得到的第二属性，并记录在对应的第二节点中。

[0149] 在一个实施例中，校验模块，还用于从联合子查询语句中确定所包含的算子；针对每个算子，确定所针对的算子所作用的列，将列的全局唯一标识和数据血缘森林中各个节点的全局唯一标识进行匹配；获取匹配成功的节点所记录的目标属性，基于所针对的算子对应的校验规则对目标属性进行校验，得到校验结果。

[0150] 上述数据查询装置中的各个模块可全部或部分通过软件、硬件及其组合来实现。上述各模块可以硬件形式内嵌于或独立于计算机设备中的处理器中，也可以以软件形式存储于计算机设备中的存储器中，以便于处理器调用执行以上各个模块对应的操作。

[0151] 在一个实施例中，提供了一种计算机设备，该计算机设备可以是服务器，其内部结构图可以如图11所示。该计算机设备包括处理器、存储器、输入/输出接口(Input/Output，简称I/O)和通信接口。其中，处理器、存储器和输入/输出接口通过系统总线连接，通信接口通过输入/输出接口连接到系统总线。其中，该计算机设备的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质和内存储器。该非易失性存储介质存储有操作系统、计算机程序。该内存储器为非易失性存储介质中的操作系统和计算机程序的运行提供环境。该计算机设备的输入/输出接口用于处理器与外部设备之间交换信息。该计算机设备的通信接口用于与外部的终端通过网络连接通信。该计算机程序被处理器执行时

以实现一种数据查询方法。

[0152] 本领域技术人员可以理解,图11中示出的结构,仅仅是与本申请方案相关的部分结构的框图,并不构成对本申请方案所应用于其上的计算机设备的限定,具体的计算机设备可以包括比图中所示更多或更少的部件,或者组合某些部件,或者具有不同的部件布置。

[0153] 在一个实施例中,提供了一种计算机设备,包括存储器和处理器,存储器中存储有计算机程序,该处理器执行计算机程序时实现上述数据查询方法的步骤。

[0154] 在一个实施例中,提供了一种计算机可读存储介质,其上存储有计算机程序,计算机程序被处理器执行时实现上述数据查询方法的步骤。

[0155] 在一个实施例中,提供了一种计算机程序产品,包括计算机程序,该计算机程序被处理器执行时实现上述数据查询方法的步骤。

[0156] 需要说明的是,本申请所涉及的用户信息(包括但不限于用户设备信息、用户个人信息等)和数据(包括但不限于用于分析的数据、存储的数据、展示的数据等),均为经用户授权或者经过各方充分授权的信息和数据,且相关数据的收集、使用和处理需要遵守相关国家和地区的相关法律法规和标准。

[0157] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程,是可以通过计算机程序来指令相关的硬件来完成,所述的计算机程序可存储于一非易失性计算机可读取存储介质中,该计算机程序在执行时,可包括如上述各方法的实施例的流程。其中,本申请所提供的各实施例中所使用的对存储器、数据库或其它介质的任何引用,均可包括非易失性和易失性存储器中的至少一种。非易失性存储器可包括只读存储器(Read-Only Memory,ROM)、磁带、软盘、闪存、光存储器、高密度嵌入式非易失性存储器、阻变存储器(ReRAM)、磁变存储器(Magnetoresistive Random Access Memory,MRAM)、铁电存储器(Ferroelectric Random Access Memory,FRAM)、相变存储器(Phase Change Memory,PCM)、石墨烯存储器等。易失性存储器可包括随机存取存储器(Random Access Memory,RAM)或外部高速缓冲存储器等。作为说明而非局限,RAM可以是多种形式,比如静态随机存取存储器(Static Random Access Memory,SRAM)或动态随机存取存储器(Dynamic Random Access Memory,DRAM)等。本申请所提供的各实施例中所涉及的数据库可包括关系型数据库和非关系型数据库中至少一种。非关系型数据库可包括基于区块链的分布式数据库等,不限于此。本申请所提供的各实施例中所涉及的处理器可为通用处理器、中央处理器、图形处理器、数字信号处理器、可编程逻辑器、基于量子计算的数据处理逻辑器等,不限于此。

[0158] 以上实施例的各技术特征可以进行任意的组合,为使描述简洁,未对上述实施例中的各个技术特征所有可能的组合都进行描述,然而,只要这些技术特征的组合不存在矛盾,都应当认为是本说明书记载的范围。

[0159] 以上所述实施例仅表达了本申请的几种实施方式,其描述较为具体和详细,但并不能因此而理解为对本申请专利范围的限制。应当指出的是,对于本领域的普通技术人员来说,在不脱离本申请构思的前提下,还可以做出若干变形和改进,这些都属于本申请的保护范围。因此,本申请的保护范围应以所附权利要求为准。

Customers

customer_id	first_name	age
1	John	31
2	Robert	32
3	David	22
4	John	25
5	Betty	28

(a)

Orders

order_id	item	amount	customer_id
1	Keyboard	400	4
2	Mouse	300	4
3	Monitor	12000	3
4	Keyboard	400	1
5	Mousepad	250	2

(b)

图1

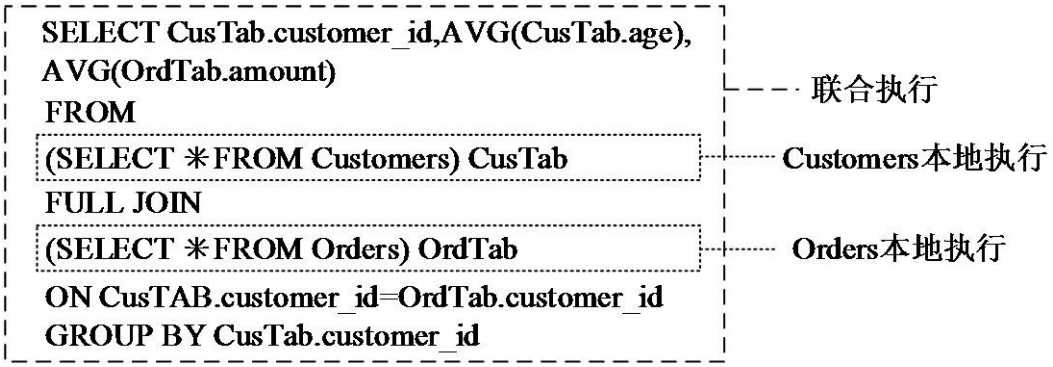


图2

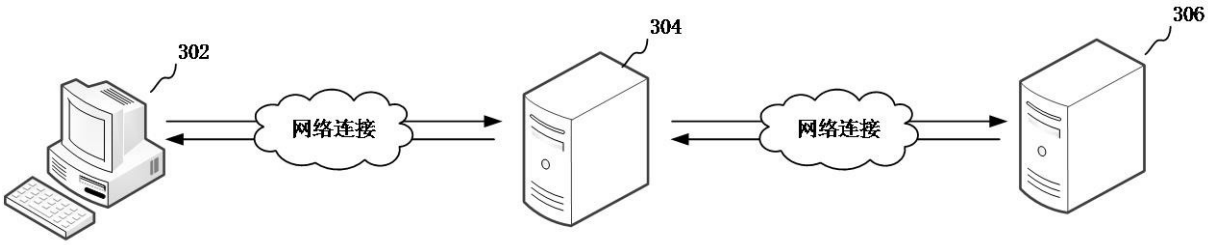


图3

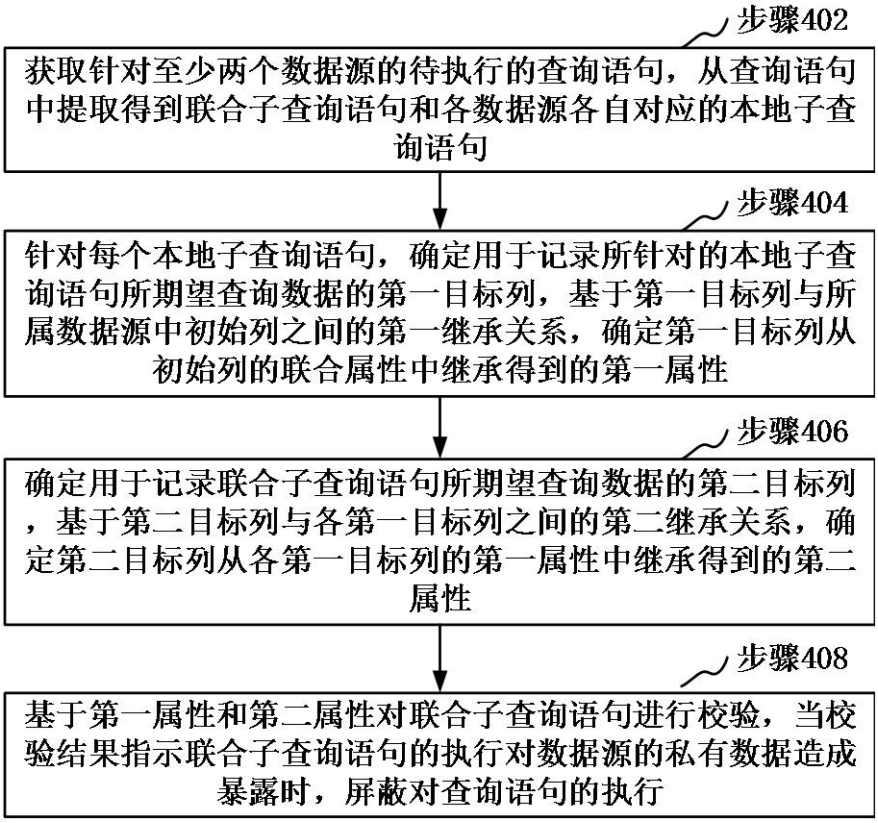

步骤402
获取针对至少两个数据源的待执行的查询语句，从查询语句中提取得到联合子查询语句和各数据源各自对应的本地子查询语句
步骤404
针对每个本地子查询语句，确定用于记录所针对的本地子查询语句所期望查询数据的第一目标列，基于第一目标列与所属数据源中初始列之间的第一继承关系，确定第一目标列从初始列的联合属性中继承得到的第一属性
步骤406
确定用于记录联合子查询语句所期望查询数据的第二目标列，基于第二目标列与各第一目标列之间的第二继承关系，确定第二目标列从各第一目标列的第一属性中继承得到的第二属性
步骤408
基于第一属性和第二属性对联合子查询语句进行校验，当校验结果指示联合子查询语句的执行对数据源的私有数据造成暴露时，屏蔽对查询语句的执行

图4

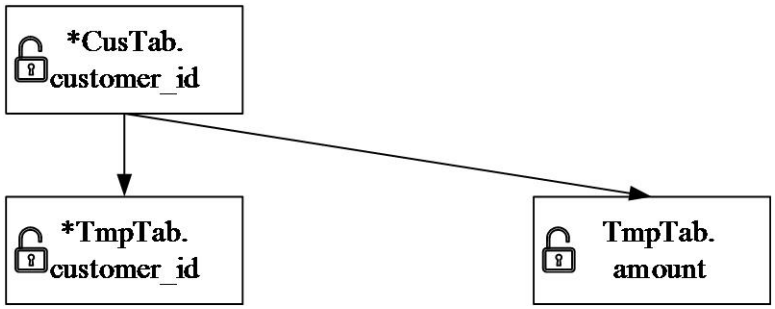

*CusTab.
customer_id
*TmpTab.
customer_id
TmpTab.
amount

图5

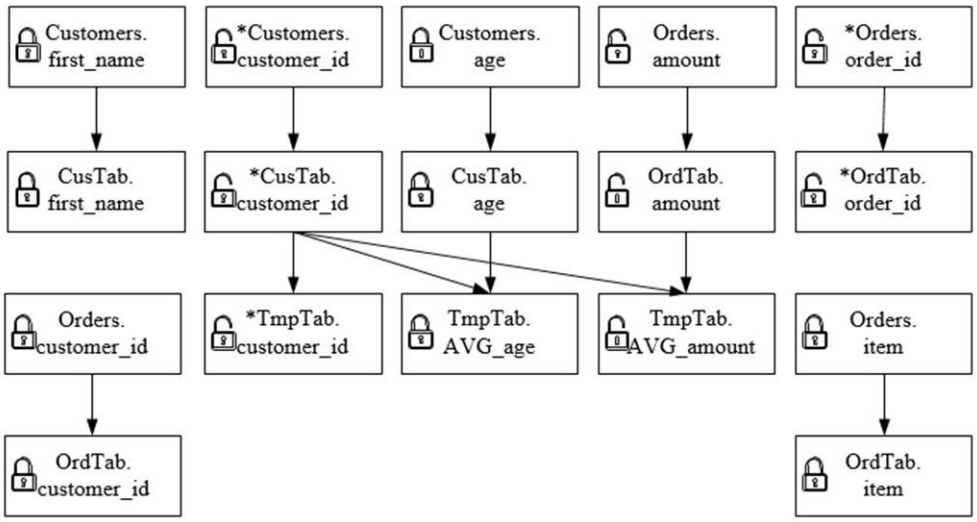

Customers.
first_name
*Customers.
customer_id
Customers.
age
Orders.
amount
*Orders.
order_id
CusTab.
first_name
*CusTab.
customer_id
CusTab.
age
OrdTab.
amount
*OrdTab.
order_id
Orders.
customer_id
*TmpTab.
customer_id
TmpTab.
AVG_age
TmpTab.
AVG_amount
Orders.
item
OrdTab.
customer_id
OrdTab.
item

图6

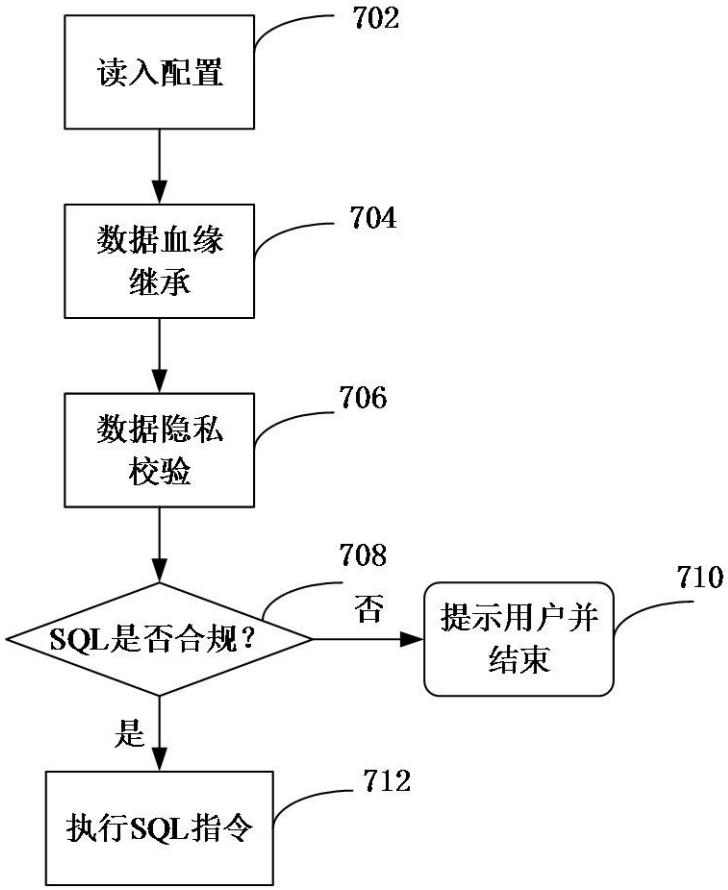

读入配置
702
数据血缘
继承
704
数据隐私
校验
706
SQL是否合规?
708
否
提示用户并
结束
710
是
执行SQL指令
712

图7

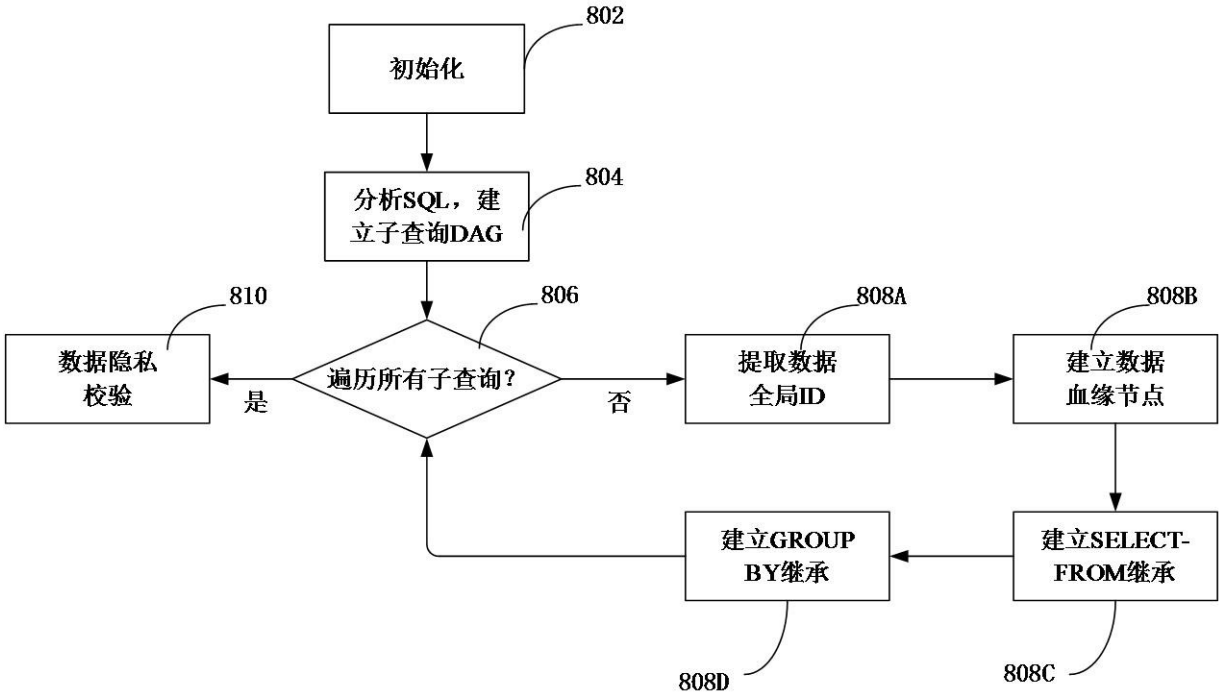

802
初始化
804
分析SQL，建立子查询DAG
806
遍历所有子查询？
是
否
810
数据隐私校验
808A
提取数据全局ID
808B
建立数据血缘节点
808C
建立SELECT-FROM继承
808D
建立GROUP BY继承

图8

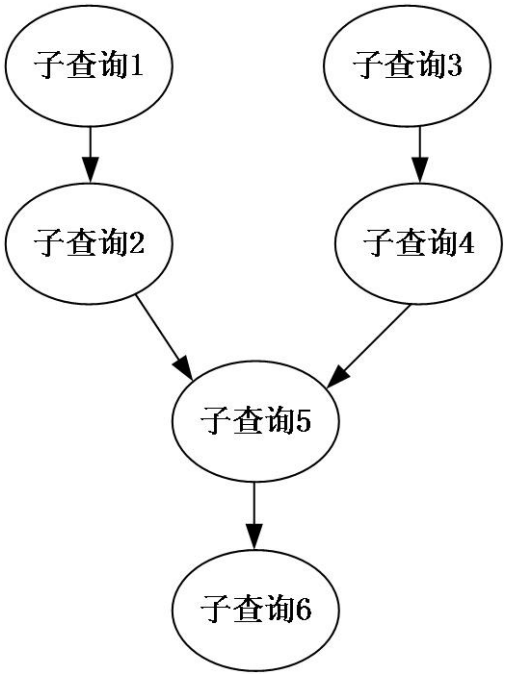

子查询1
子查询3
子查询2
子查询4
子查询5
子查询6

图9

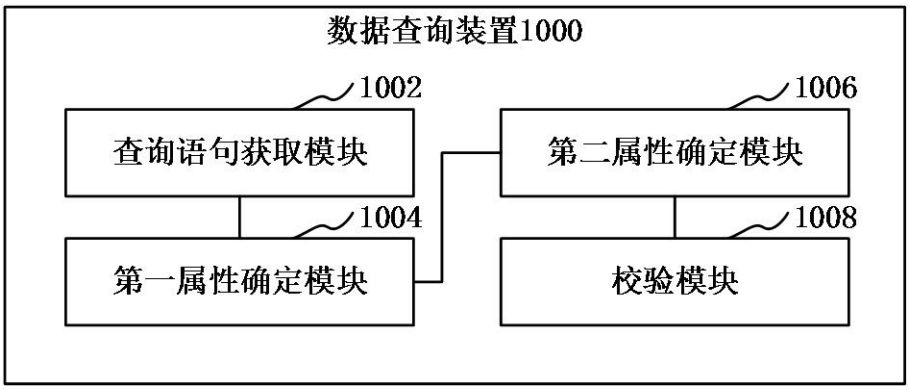

数据查询装置1000
1002
查询语句获取模块
1004
第一属性确定模块
1006
第二属性确定模块
1008
校验模块

图10

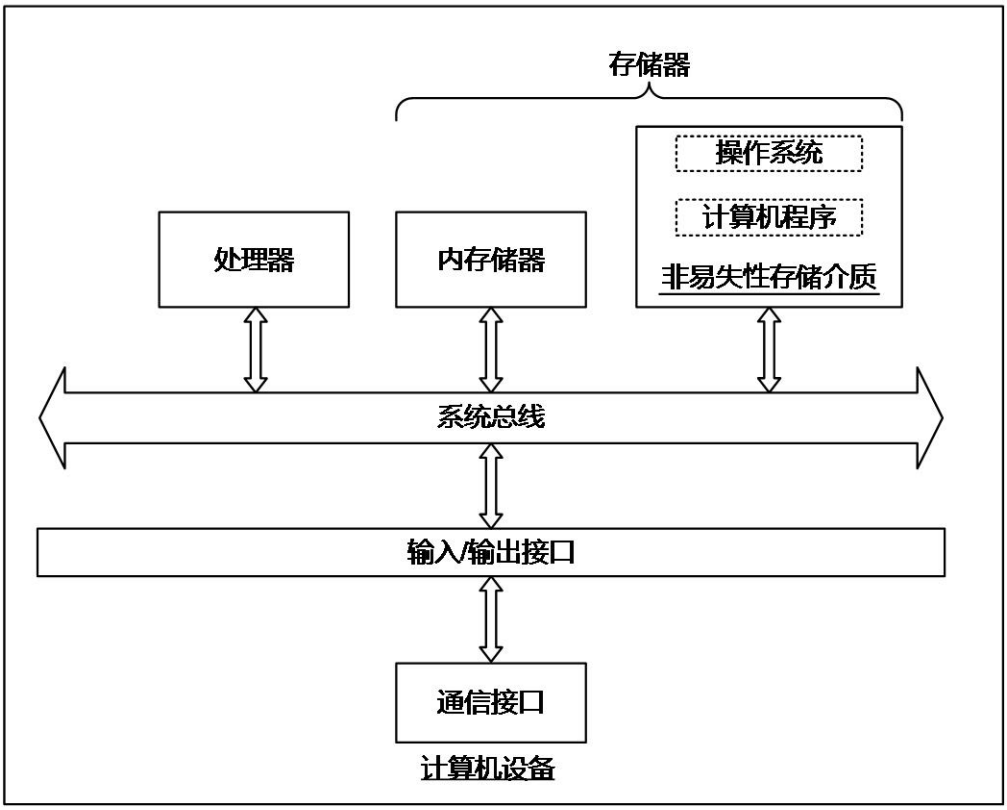

存储器
操作系统
计算机程序
非易失性存储介质
处理器
内存储器
系统总线
输入/输出接口
通信接口
计算机设备

图11