



(12) 发明专利申请

(10) 申请公布号 CN 115017182 A

(43) 申请公布日 2022. 09. 06

(21) 申请号 202210760354.0

(22) 申请日 2022.06.29

(71) 申请人 京东方科技集团股份有限公司
地址 100015 北京市朝阳区酒仙桥路10号

(72) 发明人 王莉 李卫华 李昂

(74) 专利代理机构 北京同达信恒知识产权代理有限公司 11291
专利代理师 金俊姬

(51) Int. Cl.

G06F 16/242 (2019.01)

G06F 16/2458 (2019.01)

G06F 16/248 (2019.01)

G06F 16/25 (2019.01)

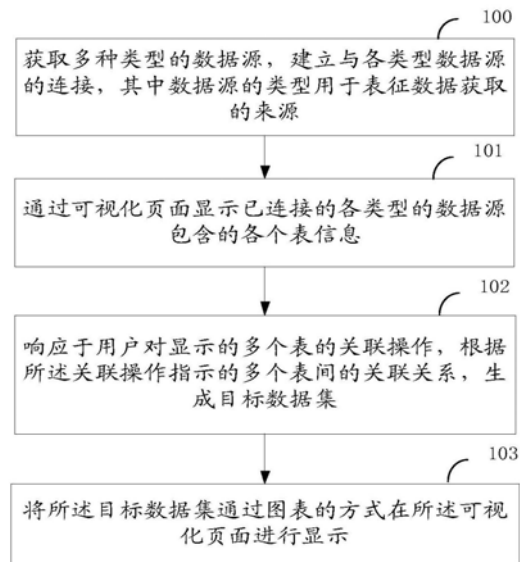
权利要求书3页 说明书39页 附图14页

(54) 发明名称

一种可视化的数据分析方法及设备

(57) 摘要

本公开提供了一种可视化的数据分析方法及设备,用于对多种类型的数据源进行可视化分析,通过建立与各类型数据源的连接关系,使得能够实时获取到多种类型的数据源,并进行各类数据源实时的组合分析。该方法包括:获取多种类型的数据源,建立与各类型数据源的连接,其中数据源的类型用于表征数据获取的来源;通过可视化页面显示已连接的各类型的数据源包含的各个表信息;响应于用户对显示的多个表的关联操作,根据所述关联操作指示的多个表间的关联关系,生成目标数据集;将所述目标数据集通过图表的方式在所述可视化页面进行显示。



1. 一种可视化的数据分析方法,其中,该方法包括:
 - 获取多种类型的数据源,建立与各类型数据源的连接,其中数据源的类型用于表征数据获取的来源;
 - 通过可视化页面显示已连接的各类型的数据源包含的各个表信息;
 - 响应于用户对显示的多个表的关联操作,根据所述关联操作指示的多个表间的关联关系,生成目标数据集;
 - 将所述目标数据集通过图表的方式在所述可视化页面进行显示。
2. 根据权利要求1所述的方法,其中,通过如下任一或任多种方式获取多种类型的数据源:
 - 接收用户输入的参数信息,根据所述参数信息获取对应类型的数据源;
 - 通过文件传送协议获取对应类型的数据源;
 - 将执行的SQL语句作为获取的对应类型的数据源。
3. 根据权利要求2所述的方法,其中,通过如下任一或任多种方式根据所述参数信息获取对应类型的数据源:
 - 接收用户输入的数据库参数,根据所述数据库参数获取数据库类型的数据源;或,
 - 接收用户输入的接口参数,根据所述接口参数获取接口类型的数据源;或,
 - 获取用户上传的文本数据,将用户命名的所述文本数据确定为文本类型的数据源;或,
 - 接收用户输入的Redis参数,根据所述Redis参数获取Redis缓存类型的数据源;或,
 - 接收用户输入的SQL语句,将输入的SQL语句确定为SQL语句类型的数据源。
4. 根据权利要求2所述的方法,其中,所述通过文件传送协议获取对应类型的数据源,包括:
 - 通过SFTP的方式获取FTP服务器中的文件,将获取的文件确定为FTP类型的数据源。
5. 根据权利要求2所述的方法,其中,所述将执行的SQL语句作为获取的对应类型的数据源,包括:
 - 接收用户对已连接的数据源执行的SQL语句,将执行的SQL语句确定为SQL语句类型的数据源。
6. 根据权利要求1~5任一所述的方法,其中,所述建立与各类型数据源的连接,包括:
 - 根据各类型的数据源的连接信息,分别建立与各类型的数据源的连接。
7. 根据权利要求6所述的方法,其中,所述根据各类型的数据源的连接信息,分别建立与各类型的数据源的连接,包括:
 - 将各类型的数据源的连接信息写入分布式查询引擎的配置文件中;
 - 当启动分布式查询引擎时,根据配置文件中各类型的数据源的连接信息,分别建立与各类型的数据源的连接。
8. 根据权利要求6所述的方法,其中,当所述数据源为数据库类型的数据源时,所述根据各类型的数据源的连接信息,分别建立与各类型的数据源的连接,包括:
 - 根据数据库参数建立与所述数据库类型的数据源的连接,其中所述数据库参数表征连接数据库所需的参数。
9. 根据权利要求6所述的方法,其中,当所述数据源为接口类型的数据源时,所述根据各类型的数据源的连接信息,分别建立与各类型的数据源的连接,包括:

根据接口参数运行接口得到JSON数据,对JSON数据进行解析,得到数据源参数;
根据解析出的数据源参数和所述接口参数,建立与接口类型的数据源的连接。

10. 根据权利要求6所述的方法,其中,当所述数据源为文本类型的数据源时,所述根据各类型的数据源的连接信息,分别建立与各类型的数据源的连接,包括:

根据文件存储服务器存储的数据源,确定数据源参数;

根据文件存储服务器的服务器参数和所述数据源参数,建立与接口类型的数据源的连接。

11. 根据权利要求9或10所述的方法,其中,所述数据源参数包括所述数据源标识、数据源的类型、库字段、表字段、列字段、列字段的字段类型中的至少一种。

12. 根据权利要求6所述的方法,其中,当所述数据源为SQL语句类型的数据源时,所述根据各类型的数据源的连接信息,分别建立与各类型的数据源的连接,包括:

对SQL语句进行语法校验,确定语法校验通过后,对所述SQL语句进行解析,得到所述SQL语句中的表信息;

根据所述SQL语句和所述SQL语句中的表信息,建立与SQL语句类型的数据源的连接。

13. 根据权利要求12所述的方法,其中,所述对所述SQL语句进行解析,得到所述SQL语句中的表信息之后,还包括:

将所述SQL语句和所述SQL语句中的表信息存储到本地数据库;

利用存储的SQL语句和用户输入的SQL语句生成嵌套的SQL语句,将生成的嵌套的SQL语句确定为获取的SQL语句类型的数据源。

14. 根据权利要求1~5任一所述的方法,其中,所述建立与各类型数据源的连接,包括:

根据各类型的数据源包含的每个数据源的连接池,构建共享数据源应用;

通过所述共享数据源应用建立各业务系统与各类型数据源的连接,其中所述共享数据源应用通过整合各类型数据源连接的能力,为各业务系统提供与各类型数据源连接的服务。

15. 根据权利要求14所述的方法,其中,所述通过所述共享数据源应用建立各业务系统与各类型数据源的连接,包括:

根据元数据描述的各类型数据源中每个数据源的连接信息,建立共享数据源应用与各类型数据源的连接;

通过所述共享数据源应用,将与共享数据源应用建立连接的各类型数据源,与各业务系统建立连接。

16. 根据权利要求14所述的方法,其中,所述通过所述共享数据源应用建立各业务系统与各类型数据源的连接,包括:

通过所述共享数据源应用接收各业务系统的访问需求;

根据各业务系统的访问需求以及各数据源的连接池的连接数量,确定各业务系统对应的目标数据源的连接池;

通过所述目标数据源的连接池,建立各业务系统与对应的目标数据源的连接。

17. 根据权利要求14所述的方法,其中,所述通过所述共享数据源应用建立各业务系统与各类型数据源的连接之后,还包括:

通过共享数据源应用,接收业务系统以元数据形式发送的操作指令;

对所述操作指令对应的数据源执行聚合、过滤、查询中的至少一种操作。

18. 根据权利要求1所述的方法,其中,所述响应于用户对显示的多个表的关联操作,根据所述关联操作指示的多个表间的关联关系,生成目标数据集,包括:

响应于用户对显示的多个表的拖拽指令,确定所述拖拽指令对应的各个目标表的表信息;

接收用户输入的多个目标表间的关联关系,根据各个目标表的表信息和所述关联关系,生成目标数据集。

19. 根据权利要求18所述的方法,其中,所述根据各个目标表的表信息和所述关联关系,生成目标数据集,包括:

根据所述关联关系确定多个目标表间相同的第一字段和多个目标表关联后保留的第二字段;

根据各个目标表的表信息、所述第一字段以及所述第二字段,生成SQL语句,执行所述SQL语句得到所述目标数据集。

20. 根据权利要求18所述的方法,其中,所述根据各个目标表的表信息和所述关联关系,生成目标数据集,还包括:

接收用户输入的过滤条件,其中所述过滤条件用于对多个目标表中的数据进行筛选;

根据所述过滤条件、多个目标表的表信息以及多个目标表间的关联关系,生成目标数据集。

21. 根据权利要求1所述的方法,其中,所述将所述目标数据集通过图表的方式在所述可视化页面进行显示,包括:

确定用户指定的图表类型以及目标数据集中的目标数据列;

将所述目标数据列作为所述图表类型对应的图表数据,利用图表组件绘制所述图表类型对应的图表;

将绘制的图表在可视化页面进行显示。

22. 一种可视化的数据分析系统,其中,该系统包括显示器和控制器:

所述显示器被配置为通过交互界面实现与用户的人机交互操作,并进行可视化页面的显示;

所述控制器被配置为基于人机交互操作执行如权利要求1~21任一所述方法的步骤。

23. 一种可视化的数据分析设备,其中,该设备包括处理器和存储器,所述存储器用于存储所述处理器可执行的程序,所述处理器用于读取所述存储器中的程序并执行权利要求1~21任一所述方法的步骤。

24. 一种计算机存储介质,其上存储有计算机程序,其中,该程序被处理器执行时实现如权利要求1~21任一所述方法的步骤。

一种可视化的数据分析方法及设备

技术领域

[0001] 本公开涉及数据分析技术领域,特别涉及一种可视化的数据分析方法及设备。

背景技术

[0002] 近些年来,各公司都在构建可视化数据分析系统,目前搭建的可视化平台大多数都是针对某一具体数据源实现的。大数据的发展带来了数据的多元化,数据的来源不仅仅是从数据库中获取,还可以从外界开放的接口、一些产品运作时的临时缓存数据等,这些数据都可以通过一定的方式固化到数据库中,从而再通过数据库可视化系统进行可视化展示。

[0003] 但是,从开放接口获取数据或从临时缓存获取数据,并固化到数据库的方式,不仅会占用可视化系统自身的存储资源,而且不利于云平台的海量数据分析。

发明内容

[0004] 本公开提供一种可视化的数据分析方法及设备,用于对多种类型的数据源进行可视化分析,通过建立与各类型数据源的连接关系,使得能够实时获取到多种类型的数据源,并进行各类数据源实时的组合分析。

[0005] 第一方面,本公开实施例提供的一种可视化的数据分析方法,该方法包括:

[0006] 获取多种类型的数据源,建立与各类型数据源的连接,其中数据源的类型用于表征数据获取的来源;

[0007] 通过可视化页面显示已连接的各类型的数据源包含的各个表信息;

[0008] 响应于用户对显示的多个表的关联操作,根据所述关联操作指示的多个表间的关联关系,生成目标数据集;

[0009] 将所述目标数据集通过图表的方式在所述可视化页面进行显示。

[0010] 作为一种可选的实施方式,通过如下任一或任多种方式获取多种类型的数据源:

[0011] 接收用户输入的参数信息,根据所述参数信息获取对应类型的数据源;

[0012] 通过文件传送协议获取对应类型的数据源;

[0013] 将执行的SQL语句作为获取的对应类型的数据源。

[0014] 作为一种可选的实施方式,通过如下任一或任多种方式根据所述参数信息获取对应类型的数据源:

[0015] 接收用户输入的数据库参数,根据所述数据库参数获取数据库类型的数据源;或,

[0016] 接收用户输入的接口参数,根据所述接口参数获取接口类型的数据源;或,

[0017] 获取用户上传的文本数据,将用户命名的所述文本数据确定为文本类型的数据源;或,

[0018] 接收用户输入的Redis参数,根据所述Redis参数获取Redis缓存类型的数据源;

或,

[0019] 接收用户输入的SQL语句,将输入的SQL语句确定为SQL语句类型的数据源。

- [0020] 作为一种可选的实施方式,所述通过文件传送协议获取对应类型的数据源,包括:
- [0021] 通过SFTP的方式获取FTP服务器中的文件,将获取的文件确定为FTP类型的数据源。
- [0022] 作为一种可选的实施方式,所述将执行的SQL语句作为获取的对应类型的数据源,包括:
- [0023] 接收用户对已连接的数据源执行的SQL语句,将执行的SQL语句确定为SQL语句类型的数据源。
- [0024] 作为一种可选的实施方式,所述建立与各类型数据源的连接,包括:
- [0025] 根据各类型的数据源的连接信息,分别建立与各类型的数据源的连接。
- [0026] 作为一种可选的实施方式,所述根据各类型的数据源的连接信息,分别建立与各类型的数据源的连接,包括:
- [0027] 将各类型的数据源的连接信息写入分布式查询引擎的配置文件中;
- [0028] 当启动分布式查询引擎时,根据配置文件中各类型的数据源的连接信息,分别建立与各类型的数据源的连接。
- [0029] 作为一种可选的实施方式,当所述数据源为数据库类型的数据源时,所述根据各类型的数据源的连接信息,分别建立与各类型的数据源的连接,包括:
- [0030] 根据数据库参数建立与所述数据库类型的数据源的连接,其中所述数据库参数表征连接数据库所需的参数。
- [0031] 作为一种可选的实施方式,当所述数据源为接口类型的数据源时,所述根据各类型的数据源的连接信息,分别建立与各类型的数据源的连接,包括:
- [0032] 根据接口参数运行接口得到JSON数据,对JSON数据进行解析,得到数据源参数;
- [0033] 根据解析出的数据源参数和所述接口参数,建立与接口类型的数据源的连接。
- [0034] 作为一种可选的实施方式,当所述数据源为文本类型的数据源时,所述根据各类型的数据源的连接信息,分别建立与各类型的数据源的连接,包括:
- [0035] 根据文件存储服务器存储的数据源,确定数据源参数;
- [0036] 根据文件存储服务器的服务器参数和所述数据源参数,建立与接口类型的数据源的连接。
- [0037] 作为一种可选的实施方式,所述数据源参数包括所述数据源标识、数据源的类型、库字段、表字段、列字段、列字段的字段类型中的至少一种。
- [0038] 作为一种可选的实施方式,当所述数据源为SQL语句类型的数据源时,所述根据各类型的数据源的连接信息,分别建立与各类型的数据源的连接,包括:
- [0039] 对SQL语句进行语法校验,确定语法校验通过后,对所述SQL语句进行解析,得到所述SQL语句中的表信息;
- [0040] 根据所述SQL语句和所述SQL语句中的表信息,建立与SQL语句类型的数据源的连接。
- [0041] 作为一种可选的实施方式,所述对所述SQL语句进行解析,得到所述SQL语句中的表信息之后,还包括:
- [0042] 将所述SQL语句和所述SQL语句中的表信息存储到本地数据库;
- [0043] 利用存储的SQL语句和用户输入的SQL语句生成嵌套的SQL语句,将生成的嵌套的

SQL语句确定为获取的SQL语句类型的数据源。

[0044] 作为一种可选的实施方式,所述建立与各类型数据源的连接,包括:

[0045] 根据各类型的数据源包含的每个数据源的连接池,构建共享数据源应用;

[0046] 通过所述共享数据源应用建立各业务系统与各类型数据源的连接,其中所述共享数据源应用通过整合各类型数据源连接的能力,为各业务系统提供与各类型数据源连接的服务。

[0047] 作为一种可选的实施方式,所述通过所述共享数据源应用建立各业务系统与各类型数据源的连接,包括:

[0048] 根据元数据描述的各类型数据源中每个数据源的连接信息,建立共享数据源应用与各类型数据源的连接;

[0049] 通过所述共享数据源应用,将与共享数据源应用建立连接的各类型数据源,与各业务系统建立连接。

[0050] 作为一种可选的实施方式,所述通过所述共享数据源应用建立各业务系统与各类型数据源的连接,包括:

[0051] 通过所述共享数据源应用接收各业务系统的访问需求;

[0052] 根据各业务系统的访问需求以及各数据源的连接池的连接数量,确定各业务系统对应的目标数据源的连接池;

[0053] 通过所述目标数据源的连接池,建立各业务系统与对应的目标数据源的连接。

[0054] 作为一种可选的实施方式,所述通过所述共享数据源应用建立各业务系统与各类型数据源的连接之后,还包括:

[0055] 通过共享数据源应用,接收业务系统以元数据形式发送的操作指令;

[0056] 对所述操作指令对应的数据源执行聚合、过滤、查询中的至少一种操作。

[0057] 作为一种可选的实施方式,所述响应于用户对显示的多个表的关联操作,根据所述关联操作指示的多个表间的关联关系,生成目标数据集,包括:

[0058] 响应于用户对显示的多个表的拖拽指令,确定所述拖拽指令对应的各个目标表的表信息;

[0059] 接收用户输入的多个目标表间的关联关系,根据各个目标表的表信息和所述关联关系,生成目标数据集。

[0060] 作为一种可选的实施方式,所述根据各个目标表的表信息和所述关联关系,生成目标数据集,包括:

[0061] 根据所述关联关系确定多个目标表间相同的第一字段和多个目标表关联后保留的第二字段;

[0062] 根据各个目标表的表信息、所述第一字段以及所述第二字段,生成SQL语句,执行所述SQL语句得到所述目标数据集。

[0063] 作为一种可选的实施方式,所述根据各个目标表的表信息和所述关联关系,生成目标数据集,还包括:

[0064] 接收用户输入的过滤条件,其中所述过滤条件用于对多个目标表中的数据进行筛选;

[0065] 根据所述过滤条件、多个目标表的表信息以及多个目标表间的关联关系,生成目

标数据集。

[0066] 作为一种可选的实施方式,所述将所述目标数据集通过图表的方式在所述可视化页面进行显示,包括:

[0067] 确定用户指定的图表类型以及目标数据集中的目标数据列;

[0068] 将所述目标数据列作为所述图表类型对应的图表数据,利用图表组件绘制所述图表类型对应的图表;

[0069] 将绘制的图表在可视化页面进行显示。

[0070] 第二方面,本公开实施例提供的一种可视化的数据分析系统,其中,该系统包括显示器和控制器:

[0071] 所述显示器被配置为通过交互界面实现与用户的人机交互操作,并进行可视化页面的显示;

[0072] 所述控制器被配置为基于人机交互操作执行如下步骤:

[0073] 获取多种类型的数据源,建立与各类型数据源的连接,其中数据源的类型用于表征数据获取的来源;

[0074] 通过可视化页面显示已连接的各类型的数据源包含的各个表信息;

[0075] 响应于用户对显示的多个表的关联操作,根据所述关联操作指示的多个表间的关联关系,生成目标数据集;

[0076] 将所述目标数据集通过图表的方式在所述可视化页面进行显示。

[0077] 作为一种可选的实施方式,所述控制器具体被配置为通过如下任一或任多种方式获取多种类型的数据源:

[0078] 接收用户输入的参数信息,根据所述参数信息获取对应类型的数据源;

[0079] 通过文件传送协议获取对应类型的数据源;

[0080] 将执行的SQL语句作为获取的对应类型的数据源。

[0081] 作为一种可选的实施方式,所述控制器具体被配置为通过如下任一或任多种方式根据所述参数信息获取对应类型的数据源:

[0082] 接收用户输入的数据库参数,根据所述数据库参数获取数据库类型的数据源;或,

[0083] 接收用户输入的接口参数,根据所述接口参数获取接口类型的数据源;或,

[0084] 获取用户上传的文本数据,将用户命名的所述文本数据确定为文本类型的数据源;或,

[0085] 接收用户输入的Redis参数,根据所述Redis参数获取Redis缓存类型的数据源;或,

[0086] 接收用户输入的SQL语句,将输入的SQL语句确定为SQL语句类型的数据源。

[0087] 作为一种可选的实施方式,所述控制器具体被配置为执行:

[0088] 通过SFTP的方式获取FTP服务器中的文件,将获取的文件确定为FTP类型的数据源。

[0089] 作为一种可选的实施方式,所述控制器具体被配置为执行:

[0090] 接收用户对已连接的数据源执行的SQL语句,将执行的SQL语句确定为SQL语句类型的数据源。

[0091] 作为一种可选的实施方式,所述控制器具体被配置为执行:

- [0092] 根据各类型的数据源的连接信息,分别建立与各类型的数据源的连接。
- [0093] 作为一种可选的实施方式,所述控制器具体被配置为执行:
- [0094] 将各类型的数据源的连接信息写入分布式查询引擎的配置文件中;
- [0095] 当启动分布式查询引擎时,根据配置文件中各类型的数据源的连接信息,分别建立与各类型的数据源的连接。
- [0096] 作为一种可选的实施方式,当所述数据源为数据库类型的数据源时,所述控制器具体被配置为执行:
- [0097] 根据数据库参数建立与所述数据库类型的数据源的连接,其中所述数据库参数表征连接数据库所需的参数。
- [0098] 作为一种可选的实施方式,当所述数据源为接口类型的数据源时,所述控制器具体被配置为执行:
- [0099] 根据接口参数运行接口得到JSON数据,对JSON数据进行解析,得到数据源参数;
- [0100] 根据解析出的数据源参数和所述接口参数,建立与接口类型的数据源的连接。
- [0101] 作为一种可选的实施方式,当所述数据源为文本类型的数据源时,所述控制器具体被配置为执行:
- [0102] 根据文件存储服务器存储的数据源,确定数据源参数;
- [0103] 根据文件存储服务器的服务器参数和所述数据源参数,建立与接口类型的数据源的连接。
- [0104] 作为一种可选的实施方式,所述数据源参数包括所述数据源标识、数据源的类型、库字段、表字段、列字段、列字段的字段类型中的至少一种。
- [0105] 作为一种可选的实施方式,当所述数据源为SQL语句类型的数据源时,所述控制器具体被配置为执行:
- [0106] 对SQL语句进行语法校验,确定语法校验通过后,对所述SQL语句进行解析,得到所述SQL语句中的表信息;
- [0107] 根据所述SQL语句和所述SQL语句中的表信息,建立与SQL语句类型的数据源的连接。
- [0108] 作为一种可选的实施方式,所述对所述SQL语句进行解析,得到所述SQL语句中的表信息之后,所述控制器具体还被配置为执行:
- [0109] 将所述SQL语句和所述SQL语句中的表信息存储到本地数据库;
- [0110] 利用存储的SQL语句和用户输入的SQL语句生成嵌套的SQL语句,将生成的嵌套的SQL语句确定为获取的SQL语句类型的数据源。
- [0111] 作为一种可选的实施方式,所述控制器具体被配置为执行:
- [0112] 根据各类型的数据源包含的每个数据源的连接池,构建共享数据源应用;
- [0113] 通过所述共享数据源应用建立各业务系统与各类型数据源的连接,其中所述共享数据源应用通过整合各类型数据源连接的能力,为各业务系统提供与各类型数据源连接的服务。
- [0114] 作为一种可选的实施方式,所述控制器具体被配置为执行:
- [0115] 根据元数据描述的各类型数据源中每个数据源的连接信息,建立共享数据源应用与各类型数据源的连接;

- [0116] 通过所述共享数据源应用,将与共享数据源应用建立连接的各种类型数据源,与各业务系统建立连接。
- [0117] 作为一种可选的实施方式,所述控制器具体被配置为执行:
- [0118] 通过所述共享数据源应用接收各业务系统的访问需求;
- [0119] 根据各业务系统的访问需求以及各数据源的连接池的连接数量,确定各业务系统对应的目标数据源的连接池;
- [0120] 通过所述目标数据源的连接池,建立各业务系统与对应的目标数据源的连接。
- [0121] 作为一种可选的实施方式,所述通过所述共享数据源应用建立各业务系统与各种类型数据源的连接之后,所述控制器具体还被配置为执行:
- [0122] 通过共享数据源应用,接收业务系统以元数据形式发送的操作指令;
- [0123] 对所述操作指令对应的数据源执行聚合、过滤、查询中的至少一种操作。
- [0124] 作为一种可选的实施方式,所述控制器具体被配置为执行:
- [0125] 响应于用户对显示的多个表的拖拽指令,确定所述拖拽指令对应的各个目标表的表信息;
- [0126] 接收用户输入的多个目标表间的关联关系,根据各个目标表的表信息和所述关联关系,生成目标数据集。
- [0127] 作为一种可选的实施方式,所述控制器具体被配置为执行:
- [0128] 根据所述关联关系确定多个目标表间相同的第一字段和多个目标表关联后保留的第二字段;
- [0129] 根据各个目标表的表信息、所述第一字段以及所述第二字段,生成SQL语句,执行所述SQL语句得到所述目标数据集。
- [0130] 作为一种可选的实施方式,所述控制器具体被配置为执行:
- [0131] 接收用户输入的过滤条件,其中所述过滤条件用于对多个目标表中的数据进行筛选;
- [0132] 根据所述过滤条件、多个目标表的表信息以及多个目标表间的关联关系,生成目标数据集。
- [0133] 作为一种可选的实施方式,所述控制器具体被配置为执行:
- [0134] 确定用户指定的图表类型以及目标数据集中的目标数据列;
- [0135] 将所述目标数据列作为所述图表类型对应的图表数据,利用图表组件绘制所述图表类型对应的图表;
- [0136] 将绘制的图表在可视化页面进行显示。
- [0137] 第三方面,本公开实施例提供一种可视化的数据分析设备,包括处理器和存储器,所述存储器用于存储所述处理器可执行的程序,所述处理器用于读取所述存储器中的程序并执行如下步骤:
- [0138] 获取多种类型的数据源,建立与各种类型数据源的连接,其中数据源的类型用于表征数据获取的来源;
- [0139] 通过可视化页面显示已连接的各种类型的数据源包含的各个表信息;
- [0140] 响应于用户对显示的多个表的关联操作,根据所述关联操作指示的多个表间的关联关系,生成目标数据集;

- [0141] 将所述目标数据集通过图表的方式在所述可视化页面进行显示。
- [0142] 作为一种可选的实施方式,所述处理器具体被配置为通过如下任一或任多种方式获取多种类型的数据源:
- [0143] 接收用户输入的参数信息,根据所述参数信息获取对应类型的数据源;
- [0144] 通过文件传送协议获取对应类型的数据源;
- [0145] 将执行的SQL语句作为获取的对应类型的数据源。
- [0146] 作为一种可选的实施方式,所述处理器具体被配置为通过如下任一或任多种方式根据所述参数信息获取对应类型的数据源:
- [0147] 接收用户输入的数据库参数,根据所述数据库参数获取数据库类型的数据源;或,
- [0148] 接收用户输入的接口参数,根据所述接口参数获取接口类型的数据源;或,
- [0149] 获取用户上传的文本数据,将用户命名的所述文本数据确定为文本类型的数据源;或,
- [0150] 接收用户输入的Redis参数,根据所述Redis参数获取Redis缓存类型的数据源;或,
- [0151] 接收用户输入的SQL语句,将输入的SQL语句确定为SQL语句类型的数据源。
- [0152] 作为一种可选的实施方式,所述处理器具体被配置为执行:
- [0153] 通过SFTP的方式获取FTP服务器中的文件,将获取的文件确定为FTP类型的数据源。
- [0154] 作为一种可选的实施方式,所述处理器具体被配置为执行:
- [0155] 接收用户对已连接的数据源执行的SQL语句,将执行的SQL语句确定为SQL语句类型的数据源。
- [0156] 作为一种可选的实施方式,所述处理器具体被配置为执行:
- [0157] 根据各类型的数据源的连接信息,分别建立与各类型的数据源的连接。
- [0158] 作为一种可选的实施方式,所述处理器具体被配置为执行:
- [0159] 将各类型的数据源的连接信息写入分布式查询引擎的配置文件中;
- [0160] 当启动分布式查询引擎时,根据配置文件中各类型的数据源的连接信息,分别建立与各类型的数据源的连接。
- [0161] 作为一种可选的实施方式,当所述数据源为数据库类型的数据源时,所述处理器具体被配置为执行:
- [0162] 根据数据库参数建立与所述数据库类型的数据源的连接,其中所述数据库参数表征连接数据库所需的参数。
- [0163] 作为一种可选的实施方式,当所述数据源为接口类型的数据源时,所述处理器具体被配置为执行:
- [0164] 根据接口参数运行接口得到JSON数据,对JSON数据进行解析,得到数据源参数;
- [0165] 根据解析出的数据源参数和所述接口参数,建立与接口类型的数据源的连接。
- [0166] 作为一种可选的实施方式,当所述数据源为文本类型的数据源时,所述处理器具体被配置为执行:
- [0167] 根据文件存储服务器存储的数据源,确定数据源参数;
- [0168] 根据文件存储服务器的服务器参数和所述数据源参数,建立与接口类型的数据源

的连接。

[0169] 作为一种可选的实施方式,所述数据源参数包括所述数据源标识、数据源的类型、库字段、表字段、列字段、列字段的字段类型中的至少一种。

[0170] 作为一种可选的实施方式,当所述数据源为SQL语句类型的数据源时,所述处理器具体被配置为执行:

[0171] 对SQL语句进行语法校验,确定语法校验通过后,对所述SQL语句进行解析,得到所述SQL语句中的表信息;

[0172] 根据所述SQL语句和所述SQL语句中的表信息,建立与SQL语句类型的数据源的连接。

[0173] 作为一种可选的实施方式,所述对所述SQL语句进行解析,得到所述SQL语句中的表信息之后,所述处理器具体还被配置为执行:

[0174] 将所述SQL语句和所述SQL语句中的表信息存储到本地数据库;

[0175] 利用存储的SQL语句和用户输入的SQL语句生成嵌套的SQL语句,将生成的嵌套的SQL语句确定为获取的SQL语句类型的数据源。

[0176] 作为一种可选的实施方式,所述处理器具体被配置为执行:

[0177] 根据各类型的数据源包含的每个数据源的连接池,构建共享数据源应用;

[0178] 通过所述共享数据源应用建立各业务系统与各类型数据源的连接,其中所述共享数据源应用通过整合各类型数据源连接的能力,为各业务系统提供与各类型数据源连接的服务。

[0179] 作为一种可选的实施方式,所述处理器具体被配置为执行:

[0180] 根据元数据描述的各类型数据源中每个数据源的连接信息,建立共享数据源应用与各类型数据源的连接;

[0181] 通过所述共享数据源应用,将与共享数据源应用建立连接的各类型数据源,与各业务系统建立连接。

[0182] 作为一种可选的实施方式,所述处理器具体被配置为执行:

[0183] 通过所述共享数据源应用接收各业务系统的访问需求;

[0184] 根据各业务系统的访问需求以及各数据源的连接池的连接数量,确定各业务系统对应的目标数据源的连接池;

[0185] 通过所述目标数据源的连接池,建立各业务系统与对应的目标数据源的连接。

[0186] 作为一种可选的实施方式,所述通过所述共享数据源应用建立各业务系统与各类型数据源的连接之后,所述处理器具体还被配置为执行:

[0187] 通过共享数据源应用,接收业务系统以元数据形式发送的操作指令;

[0188] 对所述操作指令对应的数据源执行聚合、过滤、查询中的至少一种操作。

[0189] 作为一种可选的实施方式,所述处理器具体被配置为执行:

[0190] 响应于用户对显示的多个表的拖拽指令,确定所述拖拽指令对应的各个目标表的表信息;

[0191] 接收用户输入的多个目标表间的关联关系,根据各个目标表的表信息和所述关联关系,生成目标数据集。

[0192] 作为一种可选的实施方式,所述处理器具体被配置为执行:

- [0193] 根据所述关联关系确定多个目标表间相同的第一字段和多个目标表关联后保留的第二字段；
- [0194] 根据各个目标表的表信息、所述第一字段以及所述第二字段，生成SQL语句，执行所述SQL语句得到所述目标数据集。
- [0195] 作为一种可选的实施方式，所述处理器具体还被配置为执行：
- [0196] 接收用户输入的过滤条件，其中所述过滤条件用于对多个目标表中的数据进行筛选；
- [0197] 根据所述过滤条件、多个目标表的表信息以及多个目标表间的关联关系，生成目标数据集。
- [0198] 作为一种可选的实施方式，所述处理器具体被配置为执行：
- [0199] 确定用户指定的图表类型以及目标数据集中的目标数据列；
- [0200] 将所述目标数据列作为所述图表类型对应的图表数据，利用图表组件绘制所述图表类型对应的图表；
- [0201] 将绘制的图表在可视化页面进行显示。
- [0202] 第四方面，本公开实施例还提供一种可视化的数据分析装置，该装置包括：
- [0203] 建立连接单元，用于获取多种类型的数据源，建立与各类型数据源的连接，其中数据源的类型用于表征数据获取的来源；
- [0204] 可视化显示单元，用于通过可视化页面显示已连接的各类型的数据源包含的各个表信息；
- [0205] 关联数据单元，用于响应于用户对显示的多个表的关联操作，根据所述关联操作指示的多个表间的关联关系，生成目标数据集；
- [0206] 图表显示单元，用于将所述目标数据集通过图表的方式在所述可视化页面进行显示。
- [0207] 作为一种可选的实施方式，所述建立连接单元具体用于通过如下任一或任多种方式获取多种类型的数据源：
- [0208] 接收用户输入的参数信息，根据所述参数信息获取对应类型的数据源；
- [0209] 通过文件传送协议获取对应类型的数据源；
- [0210] 将执行的SQL语句作为获取的对应类型的数据源。
- [0211] 作为一种可选的实施方式，所述建立连接单元具体用于通过如下任一或任多种方式根据所述参数信息获取对应类型的数据源：
- [0212] 接收用户输入的数据库参数，根据所述数据库参数获取数据库类型的数据源；或，
- [0213] 接收用户输入的接口参数，根据所述接口参数获取接口类型的数据源；或，
- [0214] 获取用户上传的文本数据，将用户命名的所述文本数据确定为文本类型的数据源；或，
- [0215] 接收用户输入的Redis参数，根据所述Redis参数获取Redis缓存类型的数据源；或，
- [0216] 接收用户输入的SQL语句，将输入的SQL语句确定为SQL语句类型的数据源。
- [0217] 作为一种可选的实施方式，所述建立连接单元具体用于：
- [0218] 通过SFTP的方式获取FTP服务器中的文件，将获取的文件确定为FTP类型的数据

源。

[0219] 作为一种可选的实施方式,所述建立连接单元具体用于:

[0220] 接收用户对已连接的数据源执行的SQL语句,将执行的SQL语句确定为SQL语句类型的数据源。

[0221] 作为一种可选的实施方式,所述建立连接单元具体用于:

[0222] 根据各类型的数据源的连接信息,分别建立与各类型的数据源的连接。

[0223] 作为一种可选的实施方式,所述建立连接单元具体用于:

[0224] 将各类型的数据源的连接信息写入分布式查询引擎的配置文件中;

[0225] 当启动分布式查询引擎时,根据配置文件中各类型的数据源的连接信息,分别建立与各类型的数据源的连接。

[0226] 作为一种可选的实施方式,当所述数据源为数据库类型的数据源时,所述建立连接单元具体用于:

[0227] 根据数据库参数建立与所述数据库类型的数据源的连接,其中所述数据库参数表征连接数据库所需的参数。

[0228] 作为一种可选的实施方式,当所述数据源为接口类型的数据源时,所述建立连接单元具体用于:

[0229] 根据接口参数运行接口得到JSON数据,对JSON数据进行解析,得到数据源参数;

[0230] 根据解析出的数据源参数和所述接口参数,建立与接口类型的数据源的连接。

[0231] 作为一种可选的实施方式,当所述数据源为文本类型的数据源时,所述建立连接单元具体用于:

[0232] 根据文件存储服务器存储的数据源,确定数据源参数;

[0233] 根据文件存储服务器的服务器参数和所述数据源参数,建立与接口类型的数据源的连接。

[0234] 作为一种可选的实施方式,所述数据源参数包括所述数据源标识、数据源的类型、库字段、表字段、列字段、列字段的字段类型中的至少一种。

[0235] 作为一种可选的实施方式,当所述数据源为SQL语句类型的数据源时,所述建立连接单元具体用于:

[0236] 对SQL语句进行语法校验,确定语法校验通过后,对所述SQL语句进行解析,得到所述SQL语句中的表信息;

[0237] 根据所述SQL语句和所述SQL语句中的表信息,建立与SQL语句类型的数据源的连接。

[0238] 作为一种可选的实施方式,所述对所述SQL语句进行解析,得到所述SQL语句中的表信息之后,所述建立连接单元具体还用于:

[0239] 将所述SQL语句和所述SQL语句中的表信息存储到本地数据库;

[0240] 利用存储的SQL语句和用户输入的SQL语句生成嵌套的SQL语句,将生成的嵌套的SQL语句确定为获取的SQL语句类型的数据源。

[0241] 作为一种可选的实施方式,所述建立连接单元具体用于:

[0242] 根据各类型的数据源包含的每个数据源的连接池,构建共享数据源应用;

[0243] 通过所述共享数据源应用建立各业务系统与各类型数据源的连接,其中所述共享

数据源应用通过整合各类型数据源连接的能力,为各业务系统提供与各类型数据源连接的服务。

[0244] 作为一种可选的实施方式,所述建立连接单元具体用于:

[0245] 根据元数据描述的各类型数据源中每个数据源的连接信息,建立共享数据源应用与各类型数据源连接;

[0246] 通过所述共享数据源应用,将与共享数据源应用建立连接的各类型数据源,与各业务系统建立连接。

[0247] 作为一种可选的实施方式,所述建立连接单元具体用于:

[0248] 通过所述共享数据源应用接收各业务系统的访问需求;

[0249] 根据各业务系统的访问需求以及各数据源连接池的连接数量,确定各业务系统对应的目标数据源连接池;

[0250] 通过所述目标数据源连接池,建立各业务系统与对应的目标数据源连接。

[0251] 作为一种可选的实施方式,所述通过所述共享数据源应用建立各业务系统与各类型数据源连接之后,还包括操作单元具体用于:

[0252] 通过共享数据源应用,接收业务系统以元数据形式发送的操作指令;

[0253] 对所述操作指令对应的数据源执行聚合、过滤、查询中的至少一种操作。

[0254] 作为一种可选的实施方式,所述关联数据单元具体用于:

[0255] 响应于用户对显示的多个表的拖拽指令,确定所述拖拽指令对应的各个目标表的表信息;

[0256] 接收用户输入的多个目标表间的关联关系,根据各个目标表的表信息和所述关联关系,生成目标数据集。

[0257] 作为一种可选的实施方式,所述关联数据单元具体用于:

[0258] 根据所述关联关系确定多个目标表间相同的第一字段和多个目标表关联后保留的第二字段;

[0259] 根据各个目标表的表信息、所述第一字段以及所述第二字段,生成SQL语句,执行所述SQL语句得到所述目标数据集。

[0260] 作为一种可选的实施方式,所述关联数据单元具体还用于:

[0261] 接收用户输入的过滤条件,其中所述过滤条件用于对多个目标表中的数据进行筛选;

[0262] 根据所述过滤条件、多个目标表的表信息以及多个目标表间的关联关系,生成目标数据集。

[0263] 作为一种可选的实施方式,所述图表显示单元具体用于:

[0264] 确定用户指定的图表类型以及目标数据集中的目标数据列;

[0265] 将所述目标数据列作为所述图表类型对应的图表数据,利用图表组件绘制所述图表类型对应的图表;

[0266] 将绘制的图表在可视化页面进行显示。

[0267] 第五方面,本公开实施例还提供计算机存储介质,其上存储有计算机程序,该程序被处理器执行时用于实现上述第一方面所述方法的步骤。

[0268] 本公开的这些方面或其他方面在以下的实施例的描述中会更加简明易懂。

附图说明

[0269] 为了更清楚地说明本公开实施例中的技术方案,下面将对实施例描述中所需要使用的附图作简要介绍,显而易见地,下面描述中的附图仅仅是本公开的一些实施例,对于本领域的普通技术人员来讲,在不付出创造性劳动性的前提下,还可以根据这些附图获得其他的附图。

- [0270] 图1为本公开实施例提供了一种可视化的数据分析方法的实施流程图;
- [0271] 图2A为本公开实施例提供了一种数据集生成的操作界面示意图;
- [0272] 图2B为本公开实施例提供了一种数据集生成的操作界面示意图;
- [0273] 图2C为本公开实施例提供了一种过滤数据集的操作界面图;
- [0274] 图3A为本公开实施例提供了一种显示图表的可视化页面操作示意图;
- [0275] 图3B为本公开实施例提供了一种显示图表的可视化页面操作示意图;
- [0276] 图4A为本公开实施例提供了一种获取数据库的操作界面图;
- [0277] 图4B为本公开实施例提供了一种获取数据库的操作界面图;
- [0278] 图5为本公开实施例提供了一种获取/创建Redis的连接操作界面图;
- [0279] 图6为本公开实施例提供了一种获取SQL数据源的操作界面图;
- [0280] 图7为本公开实施例提供了一种注册数据源的实施流程图;
- [0281] 图8A为本公开实施例提供了一种连接API数据源的操作界面示意图;
- [0282] 图8B为本公开实施例提供了一种连接API数据源的操作界面示意图;
- [0283] 图9为本公开实施例提供了一种建立API数据源的连接流程图;
- [0284] 图10为本公开实施例提供了一种连接SQL语句数据源的流程图;
- [0285] 图11为本公开实施例提供了一种配置SQL数据源的操作界面图;
- [0286] 图12为本公开实施例提供了一种SQL解析语法树的示意图;
- [0287] 图13为本公开实施例提供了一种传统的业务系统-数据源的连接关系示意图;
- [0288] 图14为本公开实施例提供了一种各业务系统和各数据源连接的架构示意图;
- [0289] 图15为本公开实施例提供了一种共享数据源的实施流程图;
- [0290] 图16为本公开实施例提供了一种可视化的数据分析系统示意图;
- [0291] 图17为本公开实施例提供了一种可视化的数据分析设备示意图;
- [0292] 图18为本公开实施例提供了一种可视化的数据分析装置示意图。

具体实施方式

[0293] 为了使本公开的目的、技术方案和优点更加清楚,下面将结合附图对本公开作进一步地详细描述,显然,所描述的实施例仅仅是本公开一部分实施例,而不是全部的实施例。基于本公开中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其它实施例,都属于本公开保护的范围。

[0294] 本公开实施例中术语“和/或”,描述关联对象的关联关系,表示可以存在三种关系,例如,A和/或B,可以表示:单独存在A,同时存在A和B,单独存在B这三种情况。字符“/”一般表示前后关联对象是一种“或”的关系。

[0295] 本公开实施例中的术语“数据源”,描述数据的来源,表示提供某种所需要数据的器件或原始媒体;

[0296] 本公开实施例中的术语“数据集”，又称为资料集、数据集合或资料集合，表示一种由数据所组成的集合。数据集是一个数据的集合，通常以表格形式出现。每一列代表一个特定变量。每一行都对应于某一用户的数据集。

[0297] 本公开实施例中的术语“数据库”，描述“按照数据结构来组织、存储和管理数据的仓库”。表示一个长期存储在计算机内的、有组织的、可共享的、统一管理的大量数据的集合。

[0298] 本公开实施例中的术语“Redis”，即远程字典服务，表示一个开源的使用ANSI C语言编写、支持网络、可基于内存亦可持久化的日志型、Key-Value数据库，并提供多种语言的API，常用于高并发下的缓存。

[0299] 本公开实施例中的术语“Kafka”，表示一种高吞吐量的分布式发布订阅消息系统，它可以处理消费者在网站中的所有动作流数据。这种动作（如网页浏览，搜索和其他用户的行动）是在现代网络上的许多社会功能的一个关键因素。这些数据通常是由于吞吐量的要求而通过处理日志和日志聚合来解决。对于像Hadoop一样的日志数据和离线分析系统，但又要求实时处理的限制，这是一个可行的解决方案。Kafka的目的是通过Hadoop的并行加载机制来统一线上和离线的消息处理，也是为了通过集群来提供实时的消息。

[0300] 本公开实施例中的术语“API”，是应用程序接口（Application Programming Interface, API），又称为应用编程接口，是软件系统不同组成部分衔接的约定。用于提供应用程序与开发人员以访问一组例程的能力，而又无需访问源码，或理解内部工作机制的细节。

[0301] 本公开实施例中的术语“SFTP”，在计算机领域，SSH文件传输协议（SSH File Transfer Protocol, 也称Secret File Transfer Protocol, 安全文件传送协议, Secure FTP或SFTP）是一种数据流连接，提供文件访问、传输和管理功能的网络传输协议。

[0302] 本公开实施例中的术语“Presto”，是一个facebook开源的分布式SQL查询引擎，适用于交互式分析查询，数据量支持GB到PB字节。presto的架构由关系型数据库的架构演化而来。

[0303] 本公开实施例中的术语“SQL”，是结构化查询语言（Structured Query Language）简称SQL，是一种特殊目的的编程语言，是一种数据库查询和程序设计语言，用于存取数据以及查询、更新和管理关系数据库系统。

[0304] 本公开实施例中的术语“CSV（Comma-Separated Values）”，又称逗号分隔值，是一种通用的、相对简单的文件格式。可在程序之间转移表格数据。

[0305] 本公开实施例中的术语“Minio”，是一个基于Apache License v2.0开源协议的对象存储服务。它兼容亚马逊S3云存储服务接口，非常适合于存储大容量非结构化的数据，例如图片、视频、日志文件、备份数据和容器/虚拟机镜像等，而一个对象文件可以是任意大小，从几kb到最大5T不等。

[0306] 本公开实施例描述的应用场景是为了更加清楚的说明本公开实施例的技术方案，并不构成对于本公开实施例提供的技术方案的限定，本领域普通技术人员可知，随着新应用场景的出现，本公开实施例提供的技术方案对于类似的技术问题，同样适用。其中，在本公开的描述中，除非另有说明，“多个”的含义是两个或两个以上。

[0307] 示例的，近些年来，各公司都在构建可视化数据分析系统，目前搭建的可视化平台

大多数都是针对某一具体数据源实现的。大数据的发展带来了数据的多元化,数据的来源不仅仅是从数据库中获取,还可以从外界开放的接口、一些产品运作时的临时缓存数据等,这些数据都可以通过一定的方式固化到数据库中,从而再通过数据库可视化系统进行可视化展示。但是,从开放接口获取数据或从临时缓存获取数据,并固化到数据库的方式,不仅会占用可视化系统自身的存储资源,而且不利于云平台的海量数据分析。

[0308] 目前,针对某些企业共用一套用户体系的情况,由于一套用户体系包括多种业务平台,每个用户都会在各个业务平台上留下大量用户数据,为了后续精准推送相关产品,需要汇总分析用户在不同业务平台的所有行为。每个业务平台都会涉及大量的表数据,例如Presto中的表数据,进行业务查询分析时,虽然用一条SQL语句可以将各业务系统中的数据组合起来,但每增加一个表的连接,连接的复杂度将会成指数级增长,这无疑会给查询引擎的性能带来挑战。并且,各业务平台的用户对其它平台的业务不了解,进行SQL关联前需要做大量的业务梳理工作。

[0309] 而本公开提供的数据分析方法,可以访问多种类型的数据源,并且可以通过简单的组合关联操作实现对各类数据源的组合分析,通过图表的方式在可视化页面进行显示。不仅操作简单,而且由于与各类型的数据源建立连接关系,不需要将数据源通过固化的方式进行存储,不仅能够实时进行数据的查询分析,还可以节省存储资源。本公开的数据分析方法的核心思想是,建立与各类型数据源的连接后,通过可视化页面显示各类数据源,通过用户在可视化界面对显示的多个表的关联操作,生成目标数据集,并将该目标数据集进行可视化显示。用户在操作的整个过程中,只需要简单的关联操作就能实现不同类型的数据源的组合分析,并进行可视化展示。

[0310] 如图1所示,本实施例提供的一种可视化的数据分析方法的具体实施流程如下所示:

[0311] 步骤100、获取多种类型的数据源,建立与各类型数据源的连接,其中数据源的类型用于表征数据获取的来源;

[0312] 实施中,本实施例可以建立与各类型数据源的连接,通过建立连接关系可以实时访问各类型的数据源,可选的,本实施例通过如下任一或任多种方式获取多种类型的数据源:

[0313] 方式(1)接收用户输入的参数信息,根据所述参数信息获取对应类型的数据源;

[0314] 在一些实施例中,本实施例中的参数信息包括但不限于数据库参数、接口参数、文本数据、Redis参数、SQL语句中的一种或多种;

[0315] 在一些实施例中,通过如下任一或任多种方式根据所述参数信息获取对应类型的数据源:

[0316] 接收用户输入的数据库参数,根据所述数据库参数获取数据库类型的数据源;或,

[0317] 接收用户输入的接口参数,根据所述接口参数获取接口类型的数据源;或,

[0318] 获取用户上传的文本数据,将用户命名的所述文本数据确定为文本类型的数据源;或,

[0319] 接收用户输入的Redis参数,根据所述Redis参数获取Redis缓存类型的数据源;或,

[0320] 接收用户输入的SQL语句,将输入的SQL语句确定为SQL语句类型的数据源。

[0321] 实施中,本实施例可以接收用户输入的多种类型的数据源的参数信息,根据多种参数信息,获取对应类型的数据源;例如,接收用户输入的数据库参数,根据所述数据库参数获取数据库类型的数据源;接收用户输入的接口参数,根据所述接口参数获取接口类型的数据源;以及接收用户输入的SQL语句,将输入的SQL语句确定为SQL语句类型的数据源。其中上述根据所述参数信息获取对应类型的数据源的方式中,可以选择其中一种或多种组合,本实施例对此不作过多限定。

[0322] 方式(2)通过文件传送协议获取对应类型的数据源;

[0323] 在一些实施例中,通过SFTP的方式获取FTP服务器中的文件,将获取的文件确定为FTP类型的数据源。

[0324] 方式(3)将执行的SQL语句作为获取的对应类型的数据源。

[0325] 在一些实施例中,接收用户对已连接的数据源执行的SQL语句,将执行的SQL语句确定为SQL语句类型的数据源。

[0326] 实施中,本实施例可以将上述方式(1)、方式(2)和方式(3)进行组合,通过组合后的方式获取多种类型的数据源,本实施例对具体的组合方式不作过多限定。

[0327] 在一些实施例中,本实施例中的数据源包括但不限于如下任多种:

[0328] 第1种、数据库类型的数据源,包括但不限于Mysql(关系型数据库管理系统)、PostgreSQL(是一个自由的对象-关系数据库服务器(数据库管理系统))、Oracle(甲骨文,是一个大型数据库软件)、达梦(数据库)、Hive(是基于Hadoop构建的一套数据仓库分析系统,它提供了丰富的SQL查询方式来分析存储在Hadoop分布式文件系统的数据)、Hbase(是一个分布式的、面向列的开源数据库)、InfluxDB(是一个开源的时序数据库,使用GO语言开发,特别适合用于处理和分析资源监控数据这种时序相关数据)中的至少一种;

[0329] 第2种、接口类型的数据源,包括但不限于API接口;可选的,提供的API协议包括但不限于:HTTP协议、RPC(Remote Procedure Call,远程过程调用)协议、socket(套接字)协议、SDK(Software Development Kit,软件开发工具包)协议中的至少一种。

[0330] 第3种、文本类型的数据源,包括但不限于Excel文本、CSV文本、TXT文本中的至少一种;

[0331] 第4种、FTP类型的数据源,包括但不限于SFTP类型、FTP类型中的至少一种;

[0332] 第5种、Redis缓存类型的数据源,包括但不限于Redis缓存或其他缓存中的至少一种;

[0333] 第6种、SQL语句类型的数据源,包括但不限于用户输入的SQL语句、执行的SQL语句、存储的SQL语句、生成的SQL语句中的至少一种。

[0334] 第7种、其他类型的数据源,包括但不限于本地文件、ES(文件浏览器)、kafka(是一种高吞吐量的分布式发布订阅消息系统,它可以处理消费者在网站中的所有动作流数据)、clickhost中的至少一种。

[0335] 可选的,本实施例利用Presto组件获取并连接各类型的数据源。

[0336] 步骤101、通过可视化页面显示已连接的各类型的数据源包含的各个表信息;

[0337] 在一些实施例中,本实施例通过URL嵌入到web、终端等的形式配置可视化页面,不需要进行web端与后端定义接口进行联调等,使得可视化展示不强依赖于前后端开发。

[0338] 在一些实施例中,本实施例中的表信息包括但不限于表所属的数据源标识、表字

段名称、列字段名称、列字段的字段类型中的至少一种。

[0339] 实施中,每个类型的数据源包括一个或多个表信息,以数据库为例,包括至少一个库,每个库包括至少一个表,可以将该数据库的各个库的各个表中的列信息确定为表信息。

[0340] 本实施例可以显示各类型的数据源包含的各表中的各列信息,例如在可视化页面的右侧显示每个数据源中的各列字段名。

[0341] 步骤102、响应于用户对显示的多个表的关联操作,根据所述关联操作指示的多个表间的关联关系,生成目标数据集;

[0342] 实施中,由于在可视化页面已经将各类型的数据源中的表信息都显示出来,用户只需要通过简单的关联操作就可以建立两个或两个以上的表间的关联关系,最终通过执行SQL语句的方式,根据多个表间的关联关系生成目标数据集。

[0343] 在一些实施例中,本实施例中的关联操作包括但不限于:拖拽操作、点击操作、输入关联信息操作中的至少一种,本实施例对此不作过多限定。实施中,用户可以通过简单的拖拽操作,将显示的需要关联的多个表信息拖拽到指定区域,其中拖拽操作执行时会调用后端接口获取表信息对应的表的全部信息,包括所属数据源、各列字段等信息,然后将指定区域中的多个表进行关联,生成目标数据集。

[0344] 在一些实施例中,本实施例通过如下方式生成目标数据集:

[0345] 响应于用户对显示的多个表的拖拽指令,确定所述拖拽指令对应的各个目标表的表信息;接收用户输入的多个目标表间的关联关系,根据各个目标表的表信息和所述关联关系,生成目标数据集。

[0346] 可选的,本实施例可以将各类数据源中的数据信息通过简单的拖拽方式实现数据的聚合。实施中,如图2A-2B所示,本实施例提供一种数据集生成的操作界面示意图,其中,如图2A所示,用户可选择任意已经建立连接的数据源(对应图中的区域1),选择数据源之后,会显示该数据源下的所有表信息(对应图中的区域2),用户选择多个目标表,将多个目标表的表信息拖拽到指定区域(对应图中的区域3),拖拽表信息时,后端调用后端接口获取该目标表的全部信息,包括数据源、所有列字段等,然后用户可以指定多个目标表之间的关系,即多个目标表中的某些列字段一致,从而将多个目标表关联到一起,其中,图中的区域4为属性区,可以针对产生的目标数据集中的各属性进行重命名,复制属性、删除属性等操作,其中属性是指表字段、列字段等表属性信息。图中的区域5为预览区,给用户直观的展示数据聚合后的目标数据集是否符合预期。如图2B所示,用户可以输入多个目标表间的关联关系,即定义多个目标表中的某些列字段相同,以此确定多个目标表间的关联关系从而生成目标数据集。

[0347] 在一些实施例中,本实施例通过如下方式根据各个目标表的表信息和所述关联关系,生成目标数据集:

[0348] 根据所述关联关系确定多个目标表间相同的第一个字段和多个目标表关联后保留的第二个字段;根据各个目标表的表信息、所述第一个字段以及所述第二个字段,生成SQL语句,执行所述SQL语句得到所述目标数据集。

[0349] 在一些实施例中,本实施例还可以接收用户输入的过滤条件,其中所述过滤条件用于对多个目标表中的数据进行筛选;根据所述过滤条件、多个目标表的表信息以及多个目标表间的关联关系,生成目标数据集。

[0350] 实施中,数据集可由多个数据源中的“表”进行简单的拖拽组合生成,对应的连接可为SQL中左外连接和内连接,两个表的关联需要一个桥梁,因此两表关联时需要指定相等的属性(如列字段相同)。除了进行关联外,还可在关联的基础上增加过滤条件,如图2C所示,本实施例提供的一种过滤数据集的操作界面。例如,有一张表,包含了用户购买商品相关的信息,现需要创建一个衣服品类的用户购买信息,则需要增加过滤条件,将商品的品类匹配为衣服。

[0351] 下面通过具体例子对本实施例中的数据关联和过滤的过程进行说明:

[0352] 例如表A为商品表,表B为用户表,表C为用户购买商品记录表,各表间的关联关系为表A连接表B连接表C,关联关系具体包括表A的商品ID等于表C的商品ID,表B的用户ID等于表C的用户ID。过滤条件为表B的商品类型为衣服。实施中,前端可以将表A,表B和表C的各表数据源ID(用户进行拖拽时都会调用后端接口获取到,包括后续所需数据源的各种信息),各表关联后所保留的字段及各表关联时相等的字段发送给后端,后端按照如下格式生成SQL语句,然后通过调用Presto获取SQL结果回显到界面:

[0353] SELECT A表保留属性,B表保留属性,C表保留属性

[0354] FROM A(left) join B(left) join C on A.id=C.produc_id and B.id=C.user_id WHERE A.product_type= ‘衣服’

[0355] 可选的,本实施例中的属性是指数据源ID及其类型、表字段及其类型、表中的各列字段及其类型等相关信息。

[0356] 在一些实施例中,可以将生成的目标数据集作为一个新的数据源添加到本执行主体中,以供后续使用。可选的,可以将目标数据集存储到业务数据库以供后续使用。

[0357] 步骤103、将所述目标数据集通过图表的方式在所述可视化页面进行显示。

[0358] 在一些实施例中,本实施例通过如下方式绘制图表并显示:

[0359] 确定用户指定的图表类型以及目标数据集中的目标数据列;

[0360] 将所述目标数据列作为所述图表类型对应的图表数据,利用图表组件绘制所述图表类型对应的图表;

[0361] 将绘制的图表在可视化页面进行显示。

[0362] 实施中,本实施例首先指定需要绘制的图表类型,然后通过拖拽的方式将需要绘制的目标数据集中的目标数据列拖拽到指定区域,利用图表组件绘制图表并进行可视化显示。

[0363] 在一些实施例中,本实施例中的图表组件包括但不限于前端开源组件Echart,用户通过点击选中一种图表类型,生成一个图表,然后针对选中的图表配置图表数据。如图3A-3B所示,本实施例提供一种显示图表的可视化页面操作示意图,其中,用户选中折线图后,可以对折线图进行设置,例如更改样式、插入多媒体数据、输入文字等编辑操作,设置完成后,如图3B所示,在页面右侧栏显示的各数据源的表信息中选择所需显示的目标数据集(对应图中标注的区域1),选择完目标数据集后会列出该目标数据集中的所有数据列(对应图中标注的区域2),用户从所有数据列中选择目标数据列,将所述目标数据列作为所述图表类型对应的图表数据,拖拽到指定区域(对应图中标注的区域3),利用图表组件绘制并显示基于目标数据列生成的折线图(对应图中标注的区域4)。

[0364] 在一些实施例中,确定用户指定的图表类型以及目标数据集中的目标数据列之

后,还包括:

[0365] 接收用户输入的过滤条件(对应图3B中标注的区域5),其中所述过滤条件用于对目标数据列中的数据进行筛选;将筛选后的目标数据列作为所述图表类型对应的图表数据,利用图表组件绘制所述图表类型对应的图表;将绘制的图表在可视化页面进行显示。

[0366] 可选的,用户还可以对显示的图表的颜色、文字格式、背景等进行编辑操作,本实施例对此不作过多限定。

[0367] 需要说明的是,本实施例中与各类型的数据源建立连接主要包括两个方面,一方面侧重于连接关系的建立,另一方面侧重于连接关系的共享。其中,连接关系的建立主要包括数据源的获取和注册(即连接)的过程,连接关系的共享主要包括从业务系统和数据库连接的整体架构上,提供一种共享数据源的连接关系。

[0368] 第一方面,连接关系的建立。

[0369] 在一些实施例中,本实施例通过如下任多种方式获取多种类型的数据源:

[0370] 方式1)接收用户输入的数据库参数,根据所述数据库参数获取数据库类型的数据源;

[0371] 在一些实施例中,本实施例中的数据库参数包括但不限于IP地址、端口号、数据库名、数据库类型、登录用户名、登录密码、数据源名称等中的至少一种或多种。

[0372] 可选的,本实施例利用Presto组件获取并连接各类型的数据源。其中,Presto内部集成了一些数据库的连接器的,如Mysql、PostgreSql、Oracle等数据库,针对不同数据库可输入不同的数据库参数,具体可参考Presto官方文档。对于未支持的数据库类型,可以针对Prsto源码进行插件开发,例如可以对达梦数据库进行连接功能的开发。用户选择直连数据库(内部集成的连接器对应的数据库)的方式时,需要具体指定数据库的类型,针对不同数据库的类型填入的数据库参数也存在差异,以Mysql和PostgreSql为例,如图4A-图4B所示,本实施例提供的一种获取数据库的操作界面图。其中“*”对应的内容表示用户需要输入的数据库参数。用户输入数据库参数后,根据用户所输入的数据库参数,后端服务可以利用Presto连接对应的数据库,来校验输入的数据库参数是否正确。如果错误则反馈给用户,正确则提示用户保存,将用户输入的数据库参数信息保存在本地数据库中。

[0373] 方式2)接收用户输入的接口参数,根据所述接口参数获取接口类型的数据源;

[0374] 在一些实施例中,本实施例中的接口参数包括但不限于接口名称、接口调用方式、接口路径中的至少一种信息。其中接口路径包括接口IP地址和端口。

[0375] 方式3)获取用户上传的文本数据,将用户命名的所述文本数据确定为文本类型的数据源;

[0376] 在一些实施例中,本实施例中的文本数据包括但不限于Excel文本、CSV文本、TXT文本中的至少一种。

[0377] 在实际开发过程中,难免会用一些开源数据集,当开源数据集的格式为Excel/CSV格式时,本实施例可以支持用户将历史保存的数据以Excel/CSV/TXT文本的方式上传,用户只需命名数据源名称即可。当利用Presto组件获取并连接各类型的数据源时,由于Presto可识别CSV格式的数据,可以将用户上传的文本数据都转换为CSV格式,以文本形式存储到本地存储器,供后续使用,由于存储的是文本形式,因此并不会占用较多的存储空间。

[0378] 方式4)通过SFTP的方式获取FTP服务器中的文件,将获取的文件确定为FTP类型的

数据源；

[0379] 实施中,鉴于早期的企业,有很多数据是存储于ftp服务器上,为了更好提供服务,本实施例还支持用户将文件通过sftp的方式从FTP服务器中获取文件,并注册到本执行主体中,支持的文件格式为Excel、CSV、TXT格式。其中,本实施例的执行主体可以是一个平台、系统、设备中的一种,本实施例对此不作过多限定。

[0380] 方式5)接收用户输入的Redis参数,根据所述Redis参数获取Redis缓存类型的数据源;

[0381] 本实施例还支持Redis缓存作为数据源,在特定环境下,例如双11电商大促时,服务器在短时间内会收到大量订单信息,若直接把订单信息存入到数据库中,高频率的写操作极大可能将数据库搞垮,造成服务异常。这种情况下通常都会先把订单信息存储在缓存中,一段时间内再同步到数据库。如果想要及时的分析当前销售情况,获取缓存中数据就很有必要,本实施例提供一种实施分析当前购买信息的方法,通过获取Redis缓存中的数据源,并实时进行分析以用于对用户推荐更适合的商品。

[0382] 需要说明的是,本实施例中获取Redis缓存类型的数据源后,便认为建立了与Redis缓存类型的数据源的连接关系,其中,如图5所示,本实施例提供一种获取/创建Redis的连接操作界面,用户需要提供数据源类型:Redis缓存类型;数据源名称:Redis缓存名称;数据源地址:Redis缓存地址;数据源端口号:Redis缓存端口号;登录用户名;登录密码等。

[0383] 方式6)接收用户输入的SQL语句,将输入的SQL语句确定为SQL语句类型的数据源;或,接收用户对已连接的数据源执行的SQL语句,将执行的SQL语句确定为SQL语句类型的数据源。

[0384] 如图6所示,本实施例提供一种获取SQL数据源的操作界面,其中用户需要输入自定义的SQL语句的名称即可。

[0385] 实施中,本实施例可以针对已经建立连接(已经注册过)的数据源,通过运行SQL语句,将各数据源联系起来,并将SQL语句作为一个中间过程重新作为一个数据源中的一种表信息注册回Presto中,从而可以实现该数据源的复用。创建SQL数据源时只需要输入数据源类型为SQL类型,输入数据源名称即可。

[0386] 例如,获取在第一平台和第二平台下同时购买风衣的用户基本信息,简单来说,至少需要3张表,一张为用户信息表记为表A,一张为第一平台的用户购买记录记为表B,一张为第二平台的用户购买记录记为表C,假设风衣的商品ID共用一个,可以将获取在第一平台和第二平台下同时购买风衣的用户基本信息分为三个步骤执行:步骤一,可以先在表C中取出购买过风衣的用户ID;步骤二,再在A表中查询出购买过风衣的用户,同时还在步骤一的结果中的用户ID,步骤三,将步骤二的结果关联用户基本表,得到在第一平台和第二平台上同时购买风衣的用户基本信息。对于步骤二来说,可以复用步骤一执行的SQL语句,只需要再添加一些区别于步骤一的筛选条件,对于步骤三来说,同样可以复用步骤二的SQL语句,添加相关的筛选条件。由于本实施例将SQL语句作为一种数据源,在执行复杂的数据组合查询时,可以通过生成嵌套SQL语句的方式,将生成的嵌套SQL语句作为一个数据源,而无需将每次执行的SQL语句的结果作为一个数据源继续增加表的连接,造成多表关联的复杂度成指数级增长,本实施例基于本方法能够适用于任何复杂的SQL语句,将复杂的SQL语句简单化,通过生成嵌套SQL语句,并直接执行最终嵌套的SQL语句的方式降低了查询复杂数据组

合时占用的资源,使得SQL执行的结果集无需存储在物理空间,而是将SQL语句本身作为一种数据源进行复用,有效提高了查询效率。

[0387] 在一些实施例中,本实施例通过如下方式建立与各类型数据源的连接:

[0388] 根据各类型的数据源的连接信息,分别建立与各类型的数据源的连接。

[0389] 在一些实施例中,本实施例中的连接信息包括但不限于:数据库参数、接口参数、数据源参数、服务器参数、SQL语句、SQL语句中的表信息中的至少一种,具体根据数据源的类型定义,本实施例对此不作过多限定。

[0390] 在一些实施例中,本实施例根据各类型的数据源的连接信息,通过如下方式分别建立与各类型的数据源的连接:

[0391] 将各类型的数据源的连接信息写入分布式查询引擎的配置文件中;

[0392] 当启动分布式查询引擎时,根据配置文件中各类型的数据源的连接信息,分别建立与各类型的数据源的连接。

[0393] 本实施例以Presto为例,利用Presto分布式查询引擎的特点,可以将多个数据源之间建立联系。Presto引擎中有目录(catalog),图解(schema),表(table)三个概念。其中catalog可以理解为数据源,schema理解为模式,对应数据库中的具体某一个数据库,table对应于数据库中的表信息。Presto中内置多种数据源的连接器,如Mysql、PostgreSql、Hive、Kafka、Redis等。

[0394] 对于Presto中内置连接器的数据源类型,只需将数据源连接信息(如数据库的数据库参数例如URL,用户名,密码等)写入Presto的配置文件中即可,如图7所示,本实施例还提供一种注册数据源的实施流程,具体注册流程(即建立连接流程)如下所示:

[0395] 步骤700、Presto服务启动;

[0396] 步骤701、初始化查询已建立连接的数据源信息;

[0397] 步骤702、将查询出来的数据源信息写入Presto的配置文件中,生成注册Presto的配置信息;

[0398] 步骤703、通过Http接口向Presto发送配置信息,Presto根据接收的配置信息更新本地数据库。

[0399] 实施中,Presto服务启动时,会将在本实施例中获取的数据源连接信息通过Http接口修改到Presto的Catalog下,从而将数据源信息注册到Presto中。

[0400] 在使用过程中,如需要编辑数据源,则可以通过http接口删除数据源,再重新注册即可。Presto中的数据源名称唯一,为了便于管理维护,本实施例还创建了各个数据源的数据源ID,将创建的数据源ID作为Presto中连接的数据源名称。

[0401] 在一些实施例中,本实施例根据不同类型的数据源,提供对应的连接信息,通过如下任一种情况建立和数据源间的连接关系:

[0402] 情况1、数据源为数据库类型的数据源。

[0403] 可选的,根据数据库参数建立与所述数据库类型的数据源的连接,其中所述数据库参数表征连接数据库所需的参数。

[0404] 在一些实施例中,连接信息包括数据库参数,本实施例中的数据库参数包括但不限于:IP地址、端口号、数据库库名、数据库类型、登录用户名、登录密码、数据源名称等中的至少一种或多种。

[0405] 情况2、数据源为接口类型的数据源。

[0406] 可选的,根据接口参数运行接口得到JSON数据,对JSON数据进行解析,得到数据源参数;根据解析出的数据源参数和所述接口参数,建立与接口类型的数据源的连接。

[0407] 在一些实施例中,连接信息包括数据源参数和接口参数。可选的,接口参数包括但不限于用户定义的接口名称、接口调用方式、IP地址、端口、接口路径等接口信息。

[0408] 实施中,以API接口类型的数据源为例,如图8A-8B所示,本实施例提供一种连接API数据源的操作界面示意图,图8A中,用户在创建API数据源时,在操作界面中输入接口参数,包括接口名称、接口调用方式、IP、端口、接口路径(如URL(Universal Resource Locator,统一资源定位符))等,从而获取到API数据源,获取到API数据源之后,如图8B所示,运行API接口,得到JSON(JavaScriptObject Notation,JS对象简谱,一种轻量级的数据交换格式)数据,对JSON数据进行解析,得到数据源参数;

[0409] 其中,解析得到的数据源参数包括但不限于:数据源标识、数据源的类型、库字段、表字段、列字段、列字段的字段类型中的至少一种;根据解析出的数据源参数和所述接口参数,建立与接口类型的数据源的连接。

[0410] 如图9所示,以建立连接的数据源为接口类型的数据源为例,本实施例提供一种建立API数据源的连接流程,用于说明当所述数据源为接口类型的数据源时,如何获取该数据源并根据该数据源的连接信息,建立与该数据源的连接,该流程的实施步骤如下所示:

[0411] 步骤900、接收用户输入的API数据源,指定API数据源的IP和端口;

[0412] 步骤901、接收用户指定的API数据源的URL、接口名称、调用方式;

[0413] 步骤902、接收用户输入的API调用时所需参数、消息头信息等;

[0414] 实施中,本实施例接收用户输入的接口参数,根据所述接口参数获取接口类型的数据源,其中接口参数包括API接口参数,可选的,本实施例中的API接口参数包括但不限于API数据源的IP地址、端口、API数据源的URL、接口名称、调用方式、API调用时所需参数、消息头信息中的一种或多种。

[0415] 步骤903、根据调用方式和调用时所需参数、消息头信息运行API,得到JSON数据;

[0416] 步骤904、对JSON数据进行解析,得到数据源参数;

[0417] 其中,数据源参数包括所述数据源标识、数据源的类型、库字段、表字段、列字段、列字段的字段类型中的至少一种。

[0418] 步骤905、根据解析出的数据源参数和所述接口参数,建立与接口类型的数据源的连接。

[0419] 实施中,本实施例根据接口参数运行接口得到JSON数据,对JSON数据进行解析,得到数据源参数;根据解析出的数据源参数和所述接口参数,建立与接口类型的数据源的连接。其中,接口参数包括API接口参数。

[0420] 实施中,利用JavaScript把接口返回的JSON数据读取成一个对象,然后根据用户输入的数据名称解析出对应的数据源参数,并将请求解析出数据这一过程存储在本地的数据库中。其中,更新数据源采用的是将数据源在Presto中删除,然后重新注册数据源。注册数据源时,以API数据源为例,需要给Presto提供预设格式的信息,该信息按预设格式,将数据源参数和所述接口参数提供给Presto,从而建立Presto与API数据源的连接。

[0421] 在一些实施例中,本实施例中的预设格式如下:

```

    {
      "schema":[{
        "name":"table1",
        "columns":[
          {
            "name":"key1",
            "type":"bigint"
          },
          {
            [0422] "name":"key2",
            "type":"varchar"
          }
        ],
        "sources":[
          http://localhost:9080/data.csv;
        ]
      }
    ]
  [0423] }

```

[0424] 其中,上述格式中的“sources”用于表示数据的来源,当数据源为数据库时,“sources”为数据库来源,如数据库名称、IP地址、端口号等信息,当数据源为接口数据源时,“sources”为接口来源,如接口名称、IP地址、端口号等信息,其他类型的数据源同理,“sources”对应数据的来源,用于填写各类型数据源的来源信息。

[0425] 实施中,按上述预设格式将数据源的连接信息写入分布式查询引擎的配置文件中,从而当启动分布式查询引擎时,根据配置文件中各类型的数据源的连接信息,分别建立与各类型的数据源的连接。

[0426] 情况3、数据源为文本类型的数据源。

[0427] 可选的,根据文件存储服务器存储的数据源,确定数据源参数;根据文件存储服务器的服务器参数和所述数据源参数,建立与接口类型的数据源的连接。

[0428] 可选的,本实施例中的服务器参数包括但不限于服务器IP地址、端口号等,本实施例中的数据源参数包括所述数据源标识、数据源的类型、库字段、表字段、列字段、列字段的字段类型中的至少一种。

[0429] 实施中,如果用户以Excel/CSV/TXT格式的数据进行数据源的创建,本实施例没有将以上文件中的数据写入到本地数据库中,而是将文件上传到Minio服务器上,并提供一个查询文件内容的接口放在通过Http增加数据源的方式的source字段中,详见上述预设格式,可以将该服务器参数添加到上述预设格式的source字段中,从而将该数据源注册到Presto中。

[0430] 可选的,对于数据源为FTP类型的数据源,可以将文件通过SFTP的方式从网络注册到Presto中。

[0431] 情况4、数据源为SQL语句类型的数据源。

[0432] 可选的,对SQL语句进行语法校验,确定语法校验通过后,对所述SQL语句进行解析,得到所述SQL语句中的表信息;根据所述SQL语句和所述SQL语句中的表信息,建立与SQL语句类型的数据源的连接。

[0433] 实施中,如图10所示,以建立连接的数据源为SQL语句类型的数据源为例,本实施例提供一种连接SQL语句数据源的流程,用于说明当所述数据源为SQL语句类型的数据源时,如何获取该数据源并根据该数据源的连接信息,建立与该数据源的连接,该流程的实施过程如下所示:

[0434] 步骤1000、接收用户输入的SQL语句;

[0435] 实施中,本实施例接收用户输入的SQL语句,将输入的SQL语句确定为SQL语句类型的数据源。

[0436] 实施中,常规SQL的语法为SELECT查询字段FROM表名WHERE条件GROUP BY等内容。在本实施例中用户只需按照指定格式例如[“ID”.“Schema”.“表名”]来替代常规SQL中的表名(“ID”.“Schema”及表信息),就可实现多数据源之间的数据查询。其中,“ID”指用户指定的数据源ID,“Schema”为模式,其中数据源类型不同对应的Schema也不同,数据库类型的数据源有其自有的schema,其他方式如接口数据源可以指定名称,本实施中指定接口的模式为schema,“表名”在数据库中指表名、其他方式如接口数据源为用户定义的接口名;如图11所示,本实施例还提供一种配置SQL数据源的操作界面,根据该界面中左侧区域1中数据源的各表信息,用户可以根据显示的各表信息按指定格式在区域2输入SQL语句,使得操作界面更加便捷。

[0437] 步骤1001、对SQL语句进行语法校验,确定语法校验通过;

[0438] 实施中,用户点击执行SQL,调用SQL校验模块,返回SQL执行结果,用户看到预览的结果无误后执行后续步骤,否则修改SQL语句;其中,校验模块调用Presto执行SQL语句,执行成功后会返回SQL结果集,封装成结果返回给用户,若失败,则将错误信息返回给用户,以提示用户修改SQL语句。经过SQL校验模块后,可以保证SQL的准确性。

[0439] 步骤1002、对所述SQL语句进行解析,得到所述SQL语句中的表信息;

[0440] 实施中,根据所述SQL语句和所述SQL语句中的表信息,建立与SQL语句类型的数据源的连接。

[0441] 实施中,用户保存SQL,后端服务会调用SQL解析模块,解析出SQL语句中的表信息,包括但不限于表所属的数据源标识、表字段名称、列字段名称、列字段的字段类型中的至少一种。

[0442] 其中,通过SQL解析模块,解析出注册“表”的属性名称、属性类型、属性备注等信

息。实施中,能够解析出表所属的数据源标识、表字段名称、列字段名称、列字段的字段类型等信息。

[0443] 实施中,SQL的结构为SELECT属性名FROM表名WHERE条件GROUP BY分组属性HAVING分组条件,其中FROM和WHERE中依然可以嵌套SQL语句。假设最外一层的SELECT属性名FROM表名WHERE条件GROUP BY分组属性HAVING分组条件为第一层,SQL解析模块只需要解析出第一层的SELECT中的属性名对应的实际物理“表”中的名称、数据类型、备注信息即可,第一层中的FROM中就描述了这些属性所归属的表信息,WHERE、GROUP、HAVING等条件都不需要关注。由于FROM中可以嵌套SQL语句,因此需要依次递归解析FROM中的SELECT,FROM信息,由此形成一个语法树,其中每一层节点记录了每一层的属性及其所在的表信息,叶子节点作为实际连接的表信息,根节点为查询属性分别归属于实际的哪个表。接下来只需要从叶子节点出发,向根节点进行遍历,就可以最终确定SQL最后要查询的属性分别对应的是物理存储的哪张“表”。

[0444] 可选的,本实施例中的属性可以理解为表字段名及其类型,列字段名及其类型、库字段名及其类型、数据源名称及其类型等。

[0445] 如图12所示,本实施例提供一种SQL解析语法树的示意图,其中,设有3张表,分别为表(table)1,表2,表3,对应分别为学生(student)表,教师(teacher)表,班级(class)表。按照上述描述方法,将SQL分析出语法树为3层,根节点:查询表1中的名称字段,表示4中的老师字段和班级字段。则在根节点有两个儿子节点,一个为表1,一个为表4,表4在SQL中为一个临时表,并且表4是由表2和表3生成的临时表,描述的是教师和班级之间的关系,并且查询的字段是由表2中的名称(name)字段重命名为的老师(teacher)字段和表3中的ID字段及名称重命名的班级(class)字段。因此表4会有两个儿子节点,分别为表2,表3,其中表2查询名称字段,表3查询名称字段。最终确定出该SQL最后查询的字段为表1中的名称字段及表2中的名称字段,表3中的名称字段。从最下层(第三层)的叶子节点出发,进行树的后序遍历,每到达根节点时,找出根节点中列(column)与叶子节点的对应关系,并且把根节点的列和叶子节点的表关系对应上,直到遍历结束,便可最终获得所有属性对应的表信息。图中对应的解析结果为:学生对应“1”.public.student的名称字段;老师对应“2”.public.teacher的名称字段;班级对应“3”.schema.class的名称字段。

[0446] 步骤1003、调用SQL注册模块,将SQL的信息注册到Presto中;

[0447] 实施中,根据所述SQL语句和所述SQL语句中的表信息,建立与SQL语句类型的数据源的连接。

[0448] 其中,由于SQL结果的数据量不确定,将SQL结果保存到内存中显然不太可能。本实施例中,将SQL结果以接口的形式注册到Presto中,我们只需要在后端提供一个接口,用于返回执行SQL结果,并将该接口放在上述提供给Presto的预设格式中的source字段中,SQL语句中表信息中的字段信息添加到接口注册的列字段中,调用Presto重新加载该SQL语句数据源即可。也就是说,本实施例中并不存储SQL结果,而是通过提供的接口返回SQL结果,从而有效节省服务器的物理内存资源。

[0449] 步骤1004、将SQL语句和SQL语句中的表信息存储在本地数据库中,以供后续复用该SQL语句。

[0450] 实施中,还可以利用存储的SQL语句和用户再次输入的SQL语句生成嵌套的SQL语

句,将生成的嵌套的SQL语句确定为获取的SQL语句类型的数据源,从而实现对存储的SQL语句的复用。

[0451] 其中,无需将SQL语句的执行结果保存,有效节约服务器的物理内存。

[0452] 在一些实施例中,对所述SQL语句进行解析,得到所述SQL语句中的表信息之后,还可以将所述SQL语句和所述SQL语句中的表信息存储到本地数据库;利用存储的SQL语句和用户输入的SQL语句生成嵌套的SQL语句,将生成的嵌套的SQL语句确定为获取的SQL语句类型的数据源。

[0453] 其中,在执行复杂的数据组合查询时,通过生成嵌套SQL语句的方式,将生成的嵌套SQL语句作为一个数据源,而无需将每次执行的SQL语句的结果作为一个数据源继续增加表的连接,造成多表关联的复杂度成指数级增长,将复杂的SQL语句简单化,通过生成嵌套SQL语句,并直接执行最终嵌套的SQL语句的方式降低了查询复杂数据组合时占用的资源,使得SQL执行的结果集无需存储在物理空间,而是将SQL语句本身作为一种数据源进行复用,有效提高了查询效率。

[0454] 本实施例提供的一种可视化的数据分析方法,能够支持多种数据源,打破了传统只能从数据库中展示数据的单一方式;不仅可支持多种数据源,还可将多种数据源的数据聚合(即关联)在一起;实现一种SQL数据源方式,而且执行的SQL结果集无需存储在物理空间上,依然可以作为一种数据源进行复用,并且将SQL结果注册到Presto的一种解决方式,为日后拓展其他业务提供思路;将复杂SQL简单化,可兼容支持各类复杂SQL;提供了用户拖式的页面配置,简化了前后端的开发的耦合性。用户组合操作后的数据集可用户数据分析,生成知识图谱,为企业各业务的开展提供可靠支持。

[0455] 第二方面,连接关系的共享。

[0456] 需要说明的是,如图13所示,本实施例提供一种传统的业务系统-数据源的连接关系示意图,目前每个业务系统需要自己创建维护数据源,导致占用系统资源(包括应用系统本身的物理资源(如内存)以及访问数据库时占用的公共资源),各业务或应用系统无法使用数据库的最大资源。

[0457] 为了解决上述问题,本实施例提供一种共享数据源应用的方法,通过将多个业务系统通过共享数据源资源池的方式,与各数据源建立连接,使得上层业务或应用系统不再关心及实现数据控制层,应用系统不需要再访问数据库,进行数据查询等工作,释放该层在业务系统占用的资源。另外,还可以通过元数据描述的方式将数据源注册到共享数据源应用中,再根据业务或应用需求,通过元数据描述语言进行数据查询即可。

[0458] 本实施例中的共享数据源应用,可以维护相同数据源的资源唯一性,最大程度利用数据库自身的连接池,由于涉及多个业务系统,那么根据各个业务系统的连接需求,最大程度进行数据库的高并发连接。同时提供了丰富的聚合拆分及联邦查询能力(可以进行跨数据源的链表关联等查询操作),降低上层业务或应用系统对数据的处理复杂度,同时该共享数据源应用提供丰富的扩展工具,如可视化的数据集编辑器,数据性能分析等,提升使用者效率。

[0459] 在一些实施例中,通过如下方式建立与各类型数据源的连接:

[0460] 根据各类型的数据源包含的每个数据源的连接池,构建共享数据源应用;

[0461] 通过所述共享数据源应用建立各业务系统与各类型数据源的连接,其中所述共享

数据源应用通过整合各类型数据源连接的能力,为各业务系统提供与各类型数据源连接的服务。

[0462] 可选的,本实施例中的共享数据源应用是一个以服务为主的应用,可以是Sass (Syntactically Awesome Stylesheets)应用,其中,Sass应用是一个最初由Hampton Catlin设计并由Natalie Weizenbaum开发的层叠样式表语言。在开发最初版本之后,Weizenbaum和Chris Eppstein继续通过SassScript来继续扩充Sass的功能。SassScript是一个在Sass文件中使用的小型脚本语言。

[0463] 在一些实施例中,通过所述共享数据源应用建立各业务系统与各类型数据源的连接,具体执行步骤如下所示:

[0464] 根据元数据描述的各类型数据源中每个数据源的连接信息,建立共享数据源应用与各类型数据源的连接;

[0465] 通过所述共享数据源应用,将与共享数据源应用建立连接的各类型数据源,与各业务系统建立连接。

[0466] 实施中,例如通过元数据描述进行数据源注册(即建立连接),以mysql为例,有如下描述:

[0467] connector.name=mysql//数据源类型

[0468] connection-url=jdbc:mysql://192.168.52.1:3306//数据源地址

[0469] connection-user=root//用户名

[0470] connection-password=123456//密码

[0471] 可选的,当数据源注册时判断是否该数据源已注册,如果已注册,则绑定该租户(或用户)的数据源,如果未注册,则动态创建数据源,并绑定该租户(或用户)数据源关系。

[0472] 在一些实施例中,通过所述共享数据源应用建立各业务系统与各类型数据源的连接,如图14所示,本实施例提供一种各业务系统和各数据源连接的架构示意图,基于该架构示意图,执行如下流程:

[0473] 通过所述共享数据源应用接收各业务系统的访问需求;根据各业务系统的访问需求以及各数据源的连接池的连接数量,确定各业务系统对应的目标数据源的连接池;通过所述目标数据源的连接池,建立各业务系统与对应的目标数据源的连接。其中,连接池表示创建和管理一个连接的缓冲池的技术,这些连接可以被任何需要它们的线程使用。

[0474] 可选的,如图14所示,各业务系统还可以通过多租户技术共享给多个租户使用。其中,多租户技术(multi-tenancy technology)或称多重租赁技术,是一种软件架构技术,它是在探讨与实现如何于多用户的环境下共用相同的系统或程序组件,并且仍可确保各用户间数据的隔离性。

[0475] 在一些实施例中,基于上述架构,当多个租户或用户同时访问同一数据库时,通过http建立连接,首先确定租户或用户名称,并确定是否具备访问数据库权限,如果具备访问权限则可以利用JDBC访问搜索引擎或本实施例中的Presto,经过对数据库中数据的处理,将处理结果返回业务系统。

[0476] 在一些实施例中,通过共享数据源应用,接收业务系统以元数据形式发送的操作指令;对所述操作指令对应的数据源执行聚合、过滤、查询中的至少一种操作。其中,元数据主要是描述数据属性的信息,用来支持如指示存储位置、历史数据、资源查找、文件纪录等

功能。可选的,所有基于该共享数据源应用的操作都会被记录到日志中。本实施例中的各业务或应用系统都可以对数据库中的原始数据进行加工梳理,比如聚合、过滤,或者将多个数据源的数据先查询出来,再在代码层面进行数据处理,共享数据源应用提供了丰富的聚合、过滤、联邦及可视化能力,能够大幅度减少开发人员代码编写及错误出现率。

[0477] 实施中,应用系统可以通过一个API接口访问数据源的表,直接返回查询结果,例如通过元数据描述的形式查询,查询信息如下所示:

```
{
  "id": "1971",
  "row": [{
    "caption": "code",
    "colType": "character",
    "filter": {
      "componentType": "conditionInput",
      "config": {
        "joinType": "or",
        "conditions": [{
          "conditionValue": "=",
          "value": "能效管理平台"
        }]
      }
    },
    "itemType": "dimension",
    "name": "code",
    "owner": "e2ff664bcb3d",
```

```

        "pathId": "f_r9FILrmv.204.public.cto_view_time_section_6.code",
        "remark": ""
    }],
    "column": [{
        "caption": "id",
        "colType": "bigint",
        "itemType": "measure",
        "name": "id",
        "owner": "e2ff664bcb3d",
        "pathId": "f_0wcT1QY8.204.public.cto_view_time_section_6.id",
        "remark": ""
    }],
    [0479] "filter": [{
        "caption": "create_time ( quarter ) ",
        "colType": "quarter",
        "itemType": "datetime",
        "name": "create_time",
        "owner": "e2ff664bcb3d",
        "pathId":
        "f_mxk19Ky6.204.public.cto_view_time_section_6.create_time_quarter",
        "remark": ""
    }],
    "order": [],
    "limit": 1000
}

```

[0480] 其中，一级描述key如下所示，包括：

[0481] Row:描述了科目，在聚合是可以分型分组的资源，即sql中的group by；

[0482] column:描述了需要聚合的资源，即sql中的max, sum等；

[0483] filter:描述了需要过滤的资源，即sql中的where；

[0484] order:描述了需要排序的资源，即sql中的order；

[0485] limit:描述了需要查询的条数，即sql中的limit；

[0486] 其中，二级描述key如下所示，包括：

- [0487] Caption:描述了一个资源字段的备注等;
- [0488] ColType:描述了一个资源字段的数据库类型;
- [0489] ItemType:描述了一个资源字段是字符串、数字或时间;
- [0490] Name:描述了一个资源字段的原始命名;
- [0491] Owner:描述了一个资源字段的唯一映射;
- [0492] pathId:描述了这个资源的来源(数据源、schema、数据库表、字段);
- [0493] remark:描述了自定义备注说明;
- [0494] 其中,filter:描述过滤如下所示,包括:
- [0495] componentType:描述了过滤的类型;
- [0496] config:描述了过滤的配置;
- [0497] joinType:描述了多个过滤条件间的关系;
- [0498] conditions:描述了过滤的匹配规则;
- [0499] conditionValue:描述了过滤的公式;
- [0500] value:描述了过滤的值。

[0501] 在一些实施例中,本实施例还可以建立租户-数据源之间的绑定关系,便于后期系统维护。可选的,可以构建租户ID、用户ID、数据源ID三者之间的对应关系,还可以构建数据源ID、数据源类型、数据源IP、数据源端口、数据库名称、用户名、密码、schema中多者间的对应关系。本实施例对此不作过多限定。

[0502] 如图15所示,本实施例还提供一种共享数据源的实施流程,该流程的具体实施步骤如下所示:

[0503] 步骤1500、根据各类型的数据源包含的每个数据源的连接池,构建共享数据源应用;

[0504] 其中,共享数据源应用通过整合各类型数据源连接的能力,为各业务系统提供与各类型数据源连接的服务。

[0505] 步骤1501、根据元数据描述的各类型数据源中每个数据源的连接信息,建立共享数据源应用与各类型数据源的连接;

[0506] 步骤1502、通过所述共享数据源应用,将与共享数据源应用建立连接的各类型数据源,与各业务系统建立连接;

[0507] 步骤1503、通过所述共享数据源应用接收各业务系统的访问需求;

[0508] 步骤1504、根据各业务系统的访问需求以及共享数据源应用中各数据源的连接池的连接数量,确定各业务系统对应的目标数据源的连接池;

[0509] 实施中,各独立的业务或应用系统都会对相同的数据库保持一定的资源占用,如数据库连接池连接的数据库数量是有限的,本实施例通过共享数据源应用实现数据库资源的最大利用,减少上层业务或应用系统运行环境资源,降低上层业务或应用系统开发复杂度。

[0510] 步骤1505、通过所述目标数据源的连接池,建立各业务系统与对应的目标数据源的连接。

[0511] 由于业务或应用系统经常同时连接访问相同数据源的情况,而这些业务或应用系统通常是独立的,需要各自开发实现对数据库连接与操作,且需要耗费一定的系统资源。本

实施例通过共享数据源应用进行中心化管理、监控,并提供服务,通过整合的所有数据库连接的能力,并可根据业务系统实际情况限流熔断,最大程度发挥数据库自身的全量资源能力,共享数据源应用提供了强大的数据内存计算能力,将原来在业务或应用系统中的对大量数据单点计算的形式,变为了在高速内存中分布式处理的方式。另外,数据库通常都是敏感的,对安全要求高,同一个数据库服务器需要对各业务或应用系统都开放网络连接权限,维护成本大,而本实施例通过共享数据源应用管理数据库资源的方式,可以保证数据库服务的安全性。共享数据源应用还提供了基于元数据描述的语言,开发者或者不会sql语言的业务人员,可以通过简单的语言描述即可实现业务数据操作。

[0512] 本实施例建立和各类型数据源的连接,从各应用系统或业务系统和各类型数据源的连接架构而言,通过共享数据源应用中心化的布局方式,将各应用系统和各类型的数据源通过共享数据源资源池的方式进行连接,当确定某个应用系统通过共享数据源资源池中的某个数据源的资源池,和该数据源建立连接时,可以根据该数据源的连接信息,建立和该数据源的连接,一方面,能够最大程度发挥数据库自身的全量资源能力,另一方面,能够实时进行各类型数据的查询分析,通过可视化页面显示各类数据源,通过用户在可视化界面对显示的多个表的关联操作,生成目标数据集,并将该目标数据集进行可视化显示。

[0513] 示例的,基于相同的发明构思,本公开实施例还提供了一种可视化的数据分析系统,由于该系统即是本公开实施例中的方法中的系统,并且该系统解决问题的原理与该方法相似,因此该系统的实施可以参见方法的实施,重复之处不再赘述。

[0514] 如图16所示,该系统包括显示器1600和控制器1601:

[0515] 所述显示器1600被配置为通过交互界面实现与用户的人机交互操作,并进行可视化页面的显示;

[0516] 所述控制器1601被配置为基于人机交互操作执行如下步骤:

[0517] 获取多种类型的数据源,建立与各类型数据源的连接,其中数据源的类型用于表征数据获取的来源;

[0518] 通过可视化页面显示已连接的各类型的数据源包含的各个表信息;

[0519] 响应于用户对显示的多个表的关联操作,根据所述关联操作指示的多个表间的关联关系,生成目标数据集;

[0520] 将所述目标数据集通过图表的方式在所述可视化页面进行显示。

[0521] 作为一种可选的实施方式,所述控制器1601具体被配置为通过如下任一或任多种方式获取多种类型的数据源:

[0522] 接收用户输入的参数信息,根据所述参数信息获取对应类型的数据源;

[0523] 通过文件传送协议获取对应类型的数据源;

[0524] 将执行的SQL语句作为获取的对应类型的数据源。

[0525] 作为一种可选的实施方式,所述控制器1601具体被配置为通过如下任一或任多种方式根据所述参数信息获取对应类型的数据源:

[0526] 接收用户输入的数据库参数,根据所述数据库参数获取数据库类型的数据源;或,

[0527] 接收用户输入的接口参数,根据所述接口参数获取接口类型的数据源;或,

[0528] 获取用户上传的文本数据,将用户命名的所述文本数据确定为文本类型的数据源;或,

- [0529] 接收用户输入的Redis参数,根据所述Redis参数获取Redis缓存类型的数据源;或,
- [0530] 接收用户输入的SQL语句,将输入的SQL语句确定为SQL语句类型的数据源。
- [0531] 作为一种可选的实施方式,所述控制器1601具体被配置为执行:
- [0532] 通过SFTP的方式获取FTP服务器中的文件,将获取的文件确定为FTP类型的数据源。
- [0533] 作为一种可选的实施方式,所述控制器1601具体被配置为执行:
- [0534] 接收用户对已连接的数据源执行的SQL语句,将执行的SQL语句确定为SQL语句类型的数据源。
- [0535] 作为一种可选的实施方式,所述控制器1601具体被配置为执行:
- [0536] 根据各类型的数据源的连接信息,分别建立与各类型的数据源的连接。
- [0537] 作为一种可选的实施方式,所述控制器1601具体被配置为执行:
- [0538] 将各类型的数据源的连接信息写入分布式查询引擎的配置文件中;
- [0539] 当启动分布式查询引擎时,根据配置文件中各类型的数据源的连接信息,分别建立与各类型的数据源的连接。
- [0540] 作为一种可选的实施方式,当所述数据源为数据库类型的数据源时,所述控制器1601具体被配置为执行:
- [0541] 根据数据库参数建立与所述数据库类型的数据源的连接,其中所述数据库参数表征连接数据库所需的参数。
- [0542] 作为一种可选的实施方式,当所述数据源为接口类型的数据源时,所述控制器1601具体被配置为执行:
- [0543] 根据接口参数运行接口得到JSON数据,对JSON数据进行解析,得到数据源参数;
- [0544] 根据解析出的数据源参数和所述接口参数,建立与接口类型的数据源的连接。
- [0545] 作为一种可选的实施方式,当所述数据源为文本类型的数据源时,所述控制器1601具体被配置为执行:
- [0546] 根据文件存储服务器存储的数据源,确定数据源参数;
- [0547] 根据文件存储服务器的服务器参数和所述数据源参数,建立与接口类型的数据源的连接。
- [0548] 作为一种可选的实施方式,所述数据源参数包括所述数据源标识、数据源的类型、库字段、表字段、列字段、列字段的字段类型中的至少一种。
- [0549] 作为一种可选的实施方式,当所述数据源为SQL语句类型的数据源时,所述控制器1601具体被配置为执行:
- [0550] 对SQL语句进行语法校验,确定语法校验通过后,对所述SQL语句进行解析,得到所述SQL语句中的表信息;
- [0551] 根据所述SQL语句和所述SQL语句中的表信息,建立与SQL语句类型的数据源的连接。
- [0552] 作为一种可选的实施方式,所述对所述SQL语句进行解析,得到所述SQL语句中的表信息之后,所述控制器1601具体还被配置为执行:
- [0553] 将所述SQL语句和所述SQL语句中的表信息存储到本地数据库;

- [0554] 利用存储的SQL语句和用户输入的SQL语句生成嵌套的SQL语句,将生成的嵌套的SQL语句确定为获取的SQL语句类型的数据源。
- [0555] 作为一种可选的实施方式,所述控制器1601具体被配置为执行:
- [0556] 根据各类型的数据源包含的每个数据源的连接池,构建共享数据源应用;
- [0557] 通过所述共享数据源应用建立各业务系统与各类型数据源的连接,其中所述共享数据源应用通过整合各类型数据源连接的能力,为各业务系统提供与各类型数据源连接的服务。
- [0558] 作为一种可选的实施方式,所述控制器1601具体被配置为执行:
- [0559] 根据元数据描述的各类型数据源中每个数据源的连接信息,建立共享数据源应用与各类型数据源的连接;
- [0560] 通过所述共享数据源应用,将与共享数据源应用建立连接的各类型数据源,与各业务系统建立连接。
- [0561] 作为一种可选的实施方式,所述控制器1601具体被配置为执行:
- [0562] 通过所述共享数据源应用接收各业务系统的访问需求;
- [0563] 根据各业务系统的访问需求以及各数据源的连接池的连接数量,确定各业务系统对应的目标数据源的连接池;
- [0564] 通过所述目标数据源的连接池,建立各业务系统与对应的目标数据源的连接。
- [0565] 作为一种可选的实施方式,所述通过所述共享数据源应用建立各业务系统与各类型数据源的连接之后,所述控制器1601具体还被配置为执行:
- [0566] 通过共享数据源应用,接收业务系统以元数据形式发送的操作指令;
- [0567] 对所述操作指令对应的数据源执行聚合、过滤、查询中的至少一种操作。
- [0568] 作为一种可选的实施方式,所述控制器1601具体被配置为执行:
- [0569] 响应于用户对显示的多个表的拖拽指令,确定所述拖拽指令对应的各个目标表的表信息;
- [0570] 接收用户输入的多个目标表间的关联关系,根据各个目标表的表信息和所述关联关系,生成目标数据集。
- [0571] 作为一种可选的实施方式,所述控制器1601具体被配置为执行:
- [0572] 根据所述关联关系确定多个目标表间相同的第一字段和多个目标表关联后保留的第二字段;
- [0573] 根据各个目标表的表信息、所述第一字段以及所述第二字段,生成SQL语句,执行所述SQL语句得到所述目标数据集。
- [0574] 作为一种可选的实施方式,所述控制器1601具体被配置为执行:
- [0575] 接收用户输入的过滤条件,其中所述过滤条件用于对多个目标表中的数据进行筛选;
- [0576] 根据所述过滤条件、多个目标表的表信息以及多个目标表间的关联关系,生成目标数据集。
- [0577] 作为一种可选的实施方式,所述控制器1601具体被配置为执行:
- [0578] 确定用户指定的图表类型以及目标数据集中的目标数据列;
- [0579] 将所述目标数据列作为所述图表类型对应的图表数据,利用图表组件绘制所述图

表类型对应的图表；

[0580] 将绘制的图表在可视化页面进行显示。

[0581] 示例的，基于相同的发明构思，本公开实施例还提供了一种可视化的数据分析设备，由于该设备即是本公开实施例中的方法中的设备，并且该设备解决问题的原理与该方法相似，因此该设备的实施可以参见方法的实施，重复之处不再赘述。

[0582] 如图17所示，该设备包括处理器1700和存储器1701，所述存储器1701用于存储所述处理器1700可执行的程序，所述处理器1700用于读取所述存储器1701中的程序并执行如下步骤：

[0583] 获取多种类型的数据源，建立与各类型数据源的连接，其中数据源的类型用于表征数据获取的来源；

[0584] 通过可视化页面显示已连接的各类型的数据源包含的各个表信息；

[0585] 响应于用户对显示的多个表的关联操作，根据所述关联操作指示的多个表间的关联关系，生成目标数据集；

[0586] 将所述目标数据集通过图表的方式在所述可视化页面进行显示。

[0587] 作为一种可选的实施方式，所述处理器1700具体被配置为通过如下任一或任多种方式获取多种类型的数据源：

[0588] 接收用户输入的参数信息，根据所述参数信息获取对应类型的数据源；

[0589] 通过文件传送协议获取对应类型的数据源；

[0590] 将执行的SQL语句作为获取的对应类型的数据源。

[0591] 作为一种可选的实施方式，所述处理器1700具体被配置为通过如下任一或任多种方式根据所述参数信息获取对应类型的数据源：

[0592] 接收用户输入的数据库参数，根据所述数据库参数获取数据库类型的数据源；或，

[0593] 接收用户输入的接口参数，根据所述接口参数获取接口类型的数据源；或，

[0594] 获取用户上传的文本数据，将用户命名的所述文本数据确定为文本类型的数据源；或，

[0595] 接收用户输入的Redis参数，根据所述Redis参数获取Redis缓存类型的数据源；或，

[0596] 接收用户输入的SQL语句，将输入的SQL语句确定为SQL语句类型的数据源。

[0597] 作为一种可选的实施方式，所述处理器1700具体被配置为执行：

[0598] 通过SFTP的方式获取FTP服务器中的文件，将获取的文件确定为FTP类型的数据源。

[0599] 作为一种可选的实施方式，所述处理器1700具体被配置为执行：

[0600] 接收用户对已连接的数据源执行的SQL语句，将执行的SQL语句确定为SQL语句类型的数据源。

[0601] 作为一种可选的实施方式，所述处理器1700具体被配置为执行：

[0602] 根据各类型的数据源的连接信息，分别建立与各类型的数据源的连接。

[0603] 作为一种可选的实施方式，所述处理器1700具体被配置为执行：

[0604] 将各类型的数据源的连接信息写入分布式查询引擎的配置文件中；

[0605] 当启动分布式查询引擎时，根据配置文件中各类型的数据源的连接信息，分别建

立与各类型的数据源的连接。

[0606] 作为一种可选的实施方式,当所述数据源为数据库类型的数据源时,所述处理器1700具体被配置为执行:

[0607] 根据数据库参数建立与所述数据库类型的数据源的连接,其中所述数据库参数表征连接数据库所需的参数。

[0608] 作为一种可选的实施方式,当所述数据源为接口类型的数据源时,所述处理器1700具体被配置为执行:

[0609] 根据接口参数运行接口得到JSON数据,对JSON数据进行解析,得到数据源参数;

[0610] 根据解析出的数据源参数和所述接口参数,建立与接口类型的数据源的连接。

[0611] 作为一种可选的实施方式,当所述数据源为文本类型的数据源时,所述处理器1700具体被配置为执行:

[0612] 根据文件存储服务器存储的数据源,确定数据源参数;

[0613] 根据文件存储服务器的服务器参数和所述数据源参数,建立与接口类型的数据源的连接。

[0614] 作为一种可选的实施方式,所述数据源参数包括所述数据源标识、数据源的类型、库字段、表字段、列字段、列字段的字段类型中的至少一种。

[0615] 作为一种可选的实施方式,当所述数据源为SQL语句类型的数据源时,所述处理器1700具体被配置为执行:

[0616] 对SQL语句进行语法校验,确定语法校验通过后,对所述SQL语句进行解析,得到所述SQL语句中的表信息;

[0617] 根据所述SQL语句和所述SQL语句中的表信息,建立与SQL语句类型的数据源的连接。

[0618] 作为一种可选的实施方式,所述对所述SQL语句进行解析,得到所述SQL语句中的表信息之后,所述处理器1700具体还被配置为执行:

[0619] 将所述SQL语句和所述SQL语句中的表信息存储到本地数据库;

[0620] 利用存储的SQL语句和用户输入的SQL语句生成嵌套的SQL语句,将生成的嵌套的SQL语句确定为获取的SQL语句类型的数据源。

[0621] 作为一种可选的实施方式,所述处理器1700具体被配置为执行:

[0622] 根据各类型的数据源包含的每个数据源的连接池,构建共享数据源应用;

[0623] 通过所述共享数据源应用建立各业务系统与各类型数据源的连接,其中所述共享数据源应用通过整合各类型数据源连接的能力,为各业务系统提供与各类型数据源连接的服务。

[0624] 作为一种可选的实施方式,所述处理器1700具体被配置为执行:

[0625] 根据元数据描述的各类型数据源中每个数据源的连接信息,建立共享数据源应用与各类型数据源的连接;

[0626] 通过所述共享数据源应用,将与共享数据源应用建立连接的各类型数据源,与各业务系统建立连接。

[0627] 作为一种可选的实施方式,所述处理器1700具体被配置为执行:

[0628] 通过所述共享数据源应用接收各业务系统的访问需求;

- [0629] 根据各业务系统的访问需求以及各数据源的连接池的连接数量,确定各业务系统对应的目标数据源的连接池;
- [0630] 通过所述目标数据源的连接池,建立各业务系统与对应的目标数据源的连接。
- [0631] 作为一种可选的实施方式,所述通过所述共享数据源应用建立各业务系统与各类型数据源的连接之后,所述处理器1700具体还被配置为执行:
- [0632] 通过共享数据源应用,接收业务系统以元数据形式发送的操作指令;
- [0633] 对所述操作指令对应的数据源执行聚合、过滤、查询中的至少一种操作。
- [0634] 作为一种可选的实施方式,所述处理器1700具体被配置为执行:
- [0635] 响应于用户对显示的多个表的拖拽指令,确定所述拖拽指令对应的各个目标表的表信息;
- [0636] 接收用户输入的多个目标表间的关联关系,根据各个目标表的表信息和所述关联关系,生成目标数据集。
- [0637] 作为一种可选的实施方式,所述处理器1700具体被配置为执行:
- [0638] 根据所述关联关系确定多个目标表间相同的第一字段和多个目标表关联后保留的第二字段;
- [0639] 根据各个目标表的表信息、所述第一字段以及所述第二字段,生成SQL语句,执行所述SQL语句得到所述目标数据集。
- [0640] 作为一种可选的实施方式,所述处理器1700具体还被配置为执行:
- [0641] 接收用户输入的过滤条件,其中所述过滤条件用于对多个目标表中的数据进行筛选;
- [0642] 根据所述过滤条件、多个目标表的表信息以及多个目标表间的关联关系,生成目标数据集。
- [0643] 作为一种可选的实施方式,所述处理器1700具体被配置为执行:
- [0644] 确定用户指定的图表类型以及目标数据集中的目标数据列;
- [0645] 将所述目标数据列作为所述图表类型对应的图表数据,利用图表组件绘制所述图表类型对应的图表;
- [0646] 将绘制的图表在可视化页面进行显示。
- [0647] 示例的,基于相同的发明构思,本公开实施例还提供了一种可视化的数据分析装置,由于该装置即是本公开实施例中的方法中的装置,并且该装置解决问题的原理与该方法相似,因此该装置的实施可以参见方法的实施,重复之处不再赘述。
- [0648] 如图18所示,该装置包括:
- [0649] 建立连接单元1800,用于获取多种类型的数据源,建立与各类型数据源的连接,其中数据源的类型用于表征数据获取的来源;
- [0650] 可视化显示单元1801,用于通过可视化页面显示已连接的各类型的数据源包含的各个表信息;
- [0651] 关联数据单元1802,用于响应于用户对显示的多个表的关联操作,根据所述关联操作指示的多个表间的关联关系,生成目标数据集;
- [0652] 图表显示单元1803,用于将所述目标数据集通过图表的方式在所述可视化页面进行显示。

- [0653] 作为一种可选的实施方式,所述建立连接单元1800具体用于通过如下任一或任多种方式获取多种类型的数据源:
- [0654] 接收用户输入的参数信息,根据所述参数信息获取对应类型的数据源;
- [0655] 通过文件传送协议获取对应类型的数据源;
- [0656] 将执行的SQL语句作为获取的对应类型的数据源。
- [0657] 作为一种可选的实施方式,所述建立连接单元1800具体用于通过如下任一或任多种方式根据所述参数信息获取对应类型的数据源:
- [0658] 接收用户输入的数据库参数,根据所述数据库参数获取数据库类型的数据源;或,
- [0659] 接收用户输入的接口参数,根据所述接口参数获取接口类型的数据源;或,
- [0660] 获取用户上传的文本数据,将用户命名的所述文本数据确定为文本类型的数据源;或,
- [0661] 接收用户输入的Redis参数,根据所述Redis参数获取Redis缓存类型的数据源;或,
- [0662] 接收用户输入的SQL语句,将输入的SQL语句确定为SQL语句类型的数据源。
- [0663] 作为一种可选的实施方式,所述建立连接单元1800具体用于:
- [0664] 通过SFTP的方式获取FTP服务器中的文件,将获取的文件确定为FTP类型的数据源。
- [0665] 作为一种可选的实施方式,所述建立连接单元1800具体用于:
- [0666] 接收用户对已连接的数据源执行的SQL语句,将执行的SQL语句确定为SQL语句类型的数据源。
- [0667] 作为一种可选的实施方式,所述建立连接单元1800具体用于:
- [0668] 根据各类型的数据源的连接信息,分别建立与各类型的数据源的连接。
- [0669] 作为一种可选的实施方式,所述建立连接单元1800具体用于:
- [0670] 将各类型的数据源的连接信息写入分布式查询引擎的配置文件中;
- [0671] 当启动分布式查询引擎时,根据配置文件中各类型的数据源的连接信息,分别建立与各类型的数据源的连接。
- [0672] 作为一种可选的实施方式,当所述数据源为数据库类型的数据源时,所述建立连接单元1800具体用于:
- [0673] 根据数据库参数建立与所述数据库类型的数据源的连接,其中所述数据库参数表征连接数据库所需的参数。
- [0674] 作为一种可选的实施方式,当所述数据源为接口类型的数据源时,所述建立连接单元1800具体用于:
- [0675] 根据接口参数运行接口得到JSON数据,对JSON数据进行解析,得到数据源参数;
- [0676] 根据解析出的数据源参数和所述接口参数,建立与接口类型的数据源的连接。
- [0677] 作为一种可选的实施方式,当所述数据源为文本类型的数据源时,所述建立连接单元1800具体用于:
- [0678] 根据文件存储服务器存储的数据源,确定数据源参数;
- [0679] 根据文件存储服务器的服务器参数和所述数据源参数,建立与接口类型的数据源的连接。

- [0680] 作为一种可选的实施方式,所述数据源参数包括所述数据源标识、数据源的类型、库字段、表字段、列字段、列字段的字段类型中的至少一种。
- [0681] 作为一种可选的实施方式,当所述数据源为SQL语句类型的数据源时,所述建立连接单元1800具体用于:
- [0682] 对SQL语句进行语法校验,确定语法校验通过后,对所述SQL语句进行解析,得到所述SQL语句中的表信息;
- [0683] 根据所述SQL语句和所述SQL语句中的表信息,建立与SQL语句类型的数据源的连接。
- [0684] 作为一种可选的实施方式,所述对所述SQL语句进行解析,得到所述SQL语句中的表信息之后,所述建立连接单元1800具体还用于:
- [0685] 将所述SQL语句和所述SQL语句中的表信息存储到本地数据库;
- [0686] 利用存储的SQL语句和用户输入的SQL语句生成嵌套的SQL语句,将生成的嵌套的SQL语句确定为获取的SQL语句类型的数据源。
- [0687] 作为一种可选的实施方式,所述建立连接单元1800具体用于:
- [0688] 根据各类型的数据源包含的每个数据源的连接池,构建共享数据源应用;
- [0689] 通过所述共享数据源应用建立各业务系统与各类型数据源的连接,其中所述共享数据源应用通过整合各类型数据源连接的能力,为各业务系统提供与各类型数据源连接的服务。
- [0690] 作为一种可选的实施方式,所述建立连接单元1800具体用于:
- [0691] 根据元数据描述的各类型数据源中每个数据源的连接信息,建立共享数据源应用与各类型数据源的连接;
- [0692] 通过所述共享数据源应用,将与共享数据源应用建立连接的各类型数据源,与各业务系统建立连接。
- [0693] 作为一种可选的实施方式,所述建立连接单元1800具体用于:
- [0694] 通过所述共享数据源应用接收各业务系统的访问需求;
- [0695] 根据各业务系统的访问需求以及各数据源的连接池的连接数量,确定各业务系统对应的目标数据源的连接池;
- [0696] 通过所述目标数据源的连接池,建立各业务系统与对应的目标数据源的连接。
- [0697] 作为一种可选的实施方式,所述通过所述共享数据源应用建立各业务系统与各类型数据源的连接之后,还包括操作单元具体用于:
- [0698] 通过共享数据源应用,接收业务系统以元数据形式发送的操作指令;
- [0699] 对所述操作指令对应的数据源执行聚合、过滤、查询中的至少一种操作。
- [0700] 作为一种可选的实施方式,所述关联数据单元1802具体用于:
- [0701] 响应于用户对显示的多个表的拖拽指令,确定所述拖拽指令对应的各个目标表的表信息;
- [0702] 接收用户输入的多个目标表间的关联关系,根据各个目标表的表信息和所述关联关系,生成目标数据集。
- [0703] 作为一种可选的实施方式,所述关联数据单元1802具体用于:
- [0704] 根据所述关联关系确定多个目标表间相同的第一字段和多个目标表关联后保留

的第二字段；

[0705] 根据各个目标表的表信息、所述第一字段以及所述第二字段，生成SQL语句，执行所述SQL语句得到所述目标数据集。

[0706] 作为一种可选的实施方式，所述关联数据单元1802具体还用于：

[0707] 接收用户输入的过滤条件，其中所述过滤条件用于对多个目标表中的数据进行筛选；

[0708] 根据所述过滤条件、多个目标表的表信息以及多个目标表间的关联关系，生成目标数据集。

[0709] 作为一种可选的实施方式，所述图表显示单元1803具体用于：

[0710] 确定用户指定的图表类型以及目标数据集中的目标数据列；

[0711] 将所述目标数据列作为所述图表类型对应的图表数据，利用图表组件绘制所述图表类型对应的图表；

[0712] 将绘制的图表在可视化页面进行显示。

[0713] 基于相同的发明构思，本公开实施例还提供了一种计算机存储介质，其上存储有计算机程序，该程序被处理器执行时用于实现如下步骤：

[0714] 获取多种类型的数据源，建立与各类型数据源的连接，其中数据源的类型用于表征数据获取的来源；

[0715] 通过可视化页面显示已连接的各类型的数据源包含的各个表信息；

[0716] 响应于用户对显示的多个表的关联操作，根据所述关联操作指示的多个表间的关联关系，生成目标数据集；

[0717] 将所述目标数据集通过图表的方式在所述可视化页面进行显示。

[0718] 本领域内的技术人员应明白，本公开的实施例可提供为方法、系统、或计算机程序产品。因此，本公开可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且，本公开可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器和光学存储器等)上实施的计算机程序产品的形式。

[0719] 本公开是参照根据本公开实施例的方法、设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器，使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的设备。

[0720] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中，使得存储在该计算机可读存储器中的指令产生包括指令设备的制品，该指令设备实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0721] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上，使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理，从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一

个方框或多个方框中指定的功能的步骤。

[0722] 显然,本领域的技术人员可以对本公开进行各种改动和变型而不脱离本公开的精神和范围。这样,倘若本公开的这些修改和变型属于本公开权利要求及其等同技术的范围之内,则本公开也意图包含这些改动和变型在内。

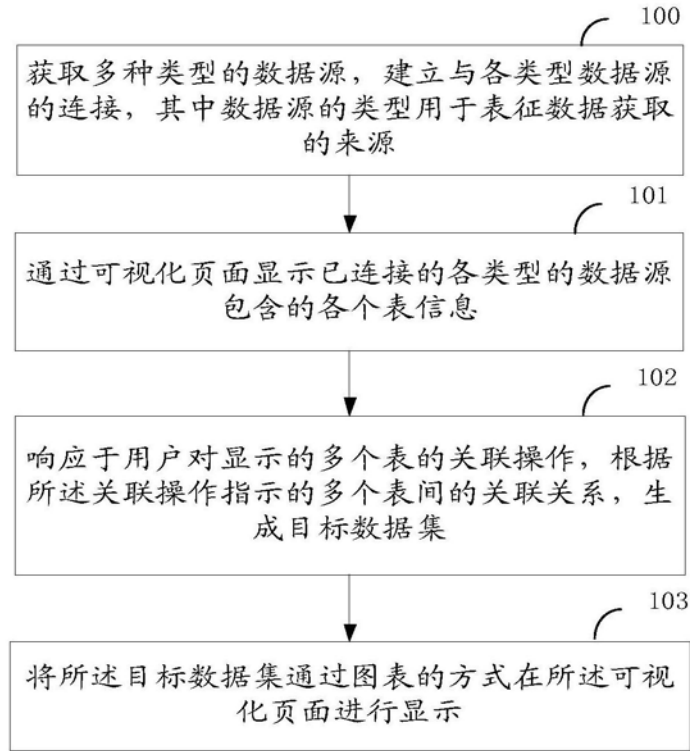


图1

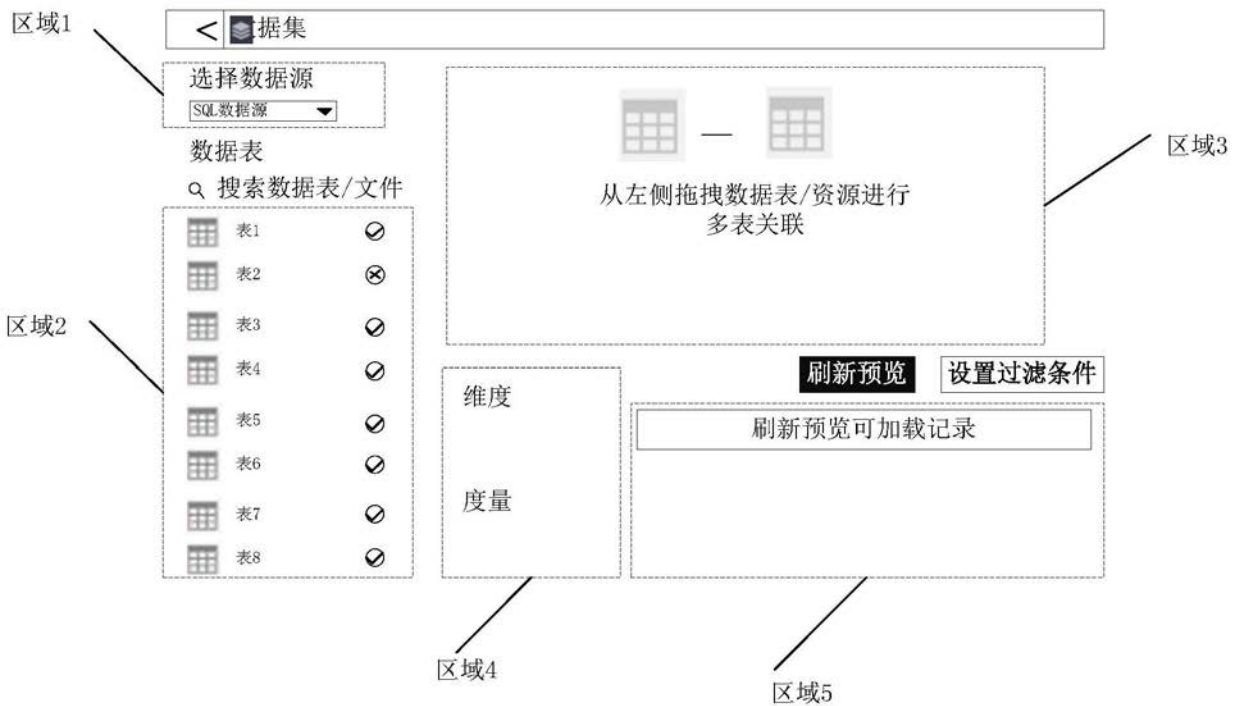


图2A

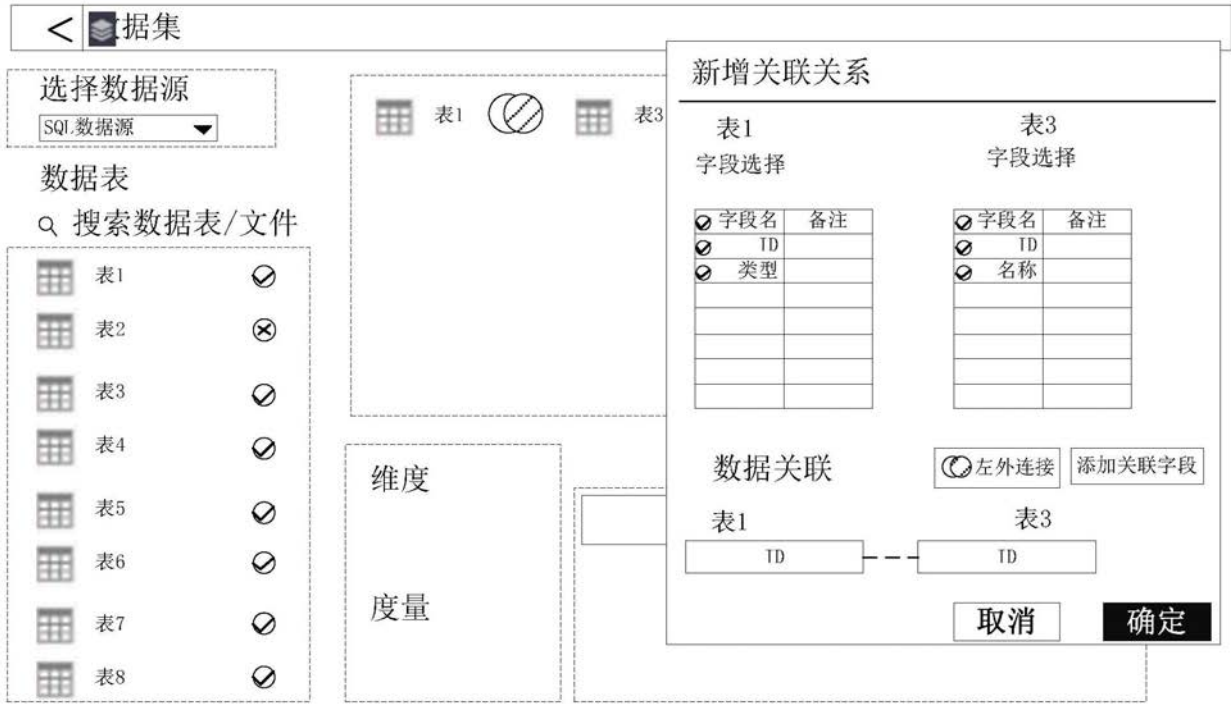


图2B



图2C

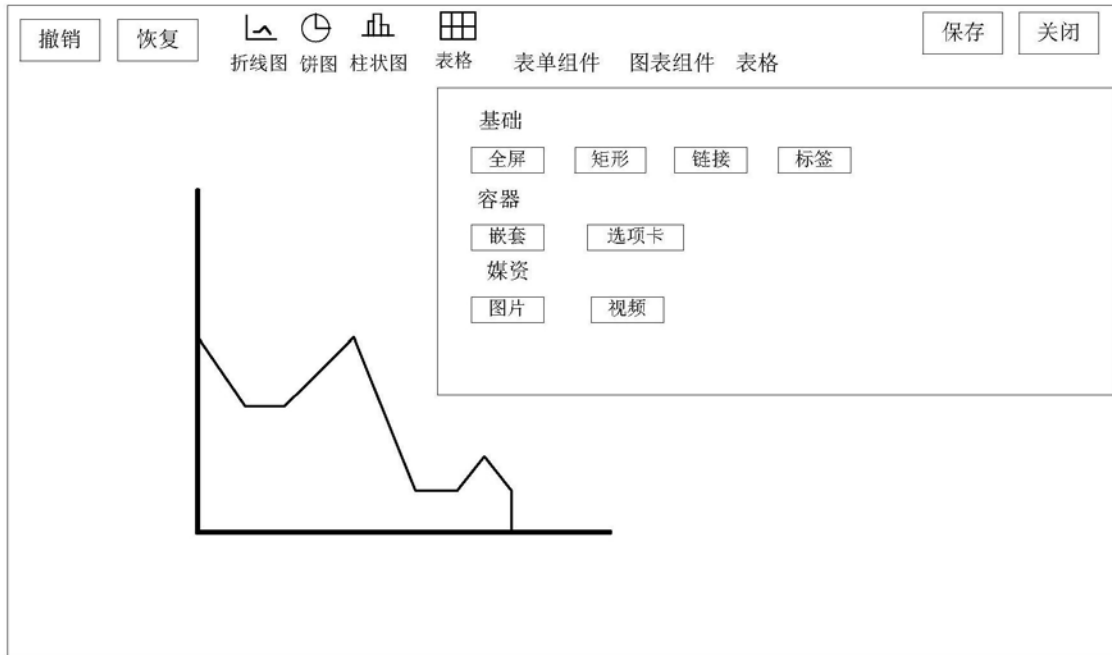


图3A

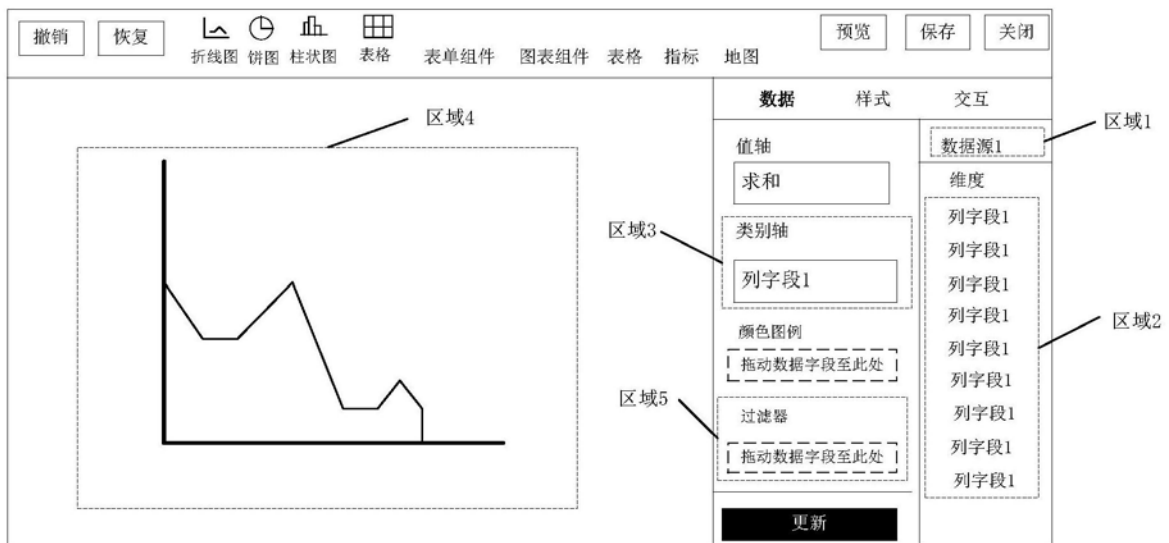


图3B

创建数据源 ✕

* 类型

* 名称

* 数据库地址

* 端口

* 数据库

用户名

密码

公钥 私有 公有

图4A

创建数据源 ✕

* 类型

* 名称

* 数据库地址

* 端口

* 数据库

schema

用户名

密码

公钥 私有 公有

图4B

创建数据源 ✕

* 类型

* 名称

* 数据库地址

* 端口

用户名

密码

公钥 私有 公有

图5

创建SQL ✕

* SQL名称

图6

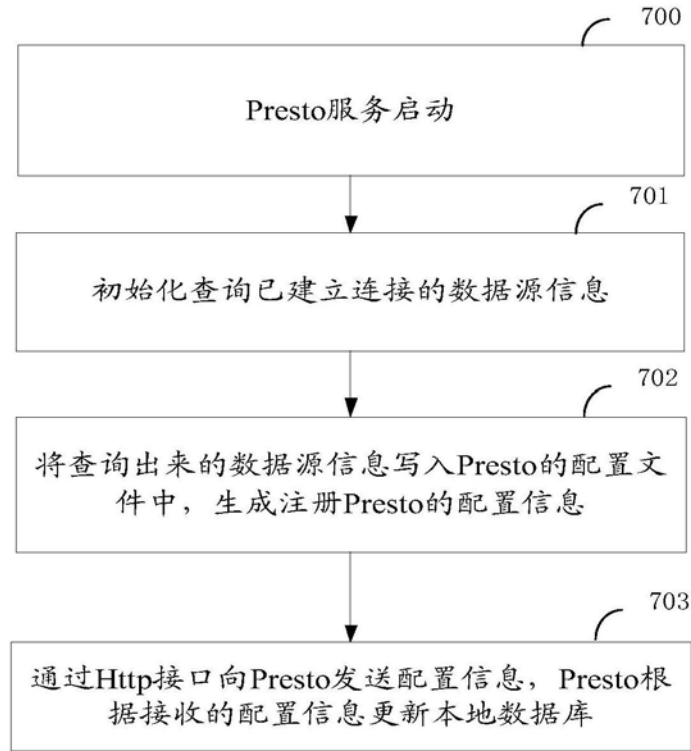


图7

创建数据源 ✕

| | |
|--------|----------------------|
| * 接口名称 | <input type="text"/> |
| * 调用方式 | <input type="text"/> |
| 服务基础路径 | <input type="text"/> |
| * 接口路径 | <input type="text"/> |

图8A

配置数据源✕

| | | | | | |
|--------|--|--------|----------------------------------|-----------|--|
| * 接口名称 | <input type="text" value="role"/> | * 调用方式 | <input type="text" value="GET"/> | | |
| 服务基础路径 | <input type="text"/> | | | 添加 | |
| * 接口路径 | <input type="text" value="/sysRole/page"/> | | | | |
| 完整路径 | <input type="text" value="10.10.239.120.8072/sysRole/page"/> | | | 运行 | |
| 提取数据 | <input type="text" value="普通 JSON提取"/> | | | | |

| | |
|---|-----------------------|
| 1 | { |
| 2 | "code" :1000, |
| 3 | "message" :无效用户, |
| 4 | "timestamp" :16522476 |
| 5 | } |

图8B

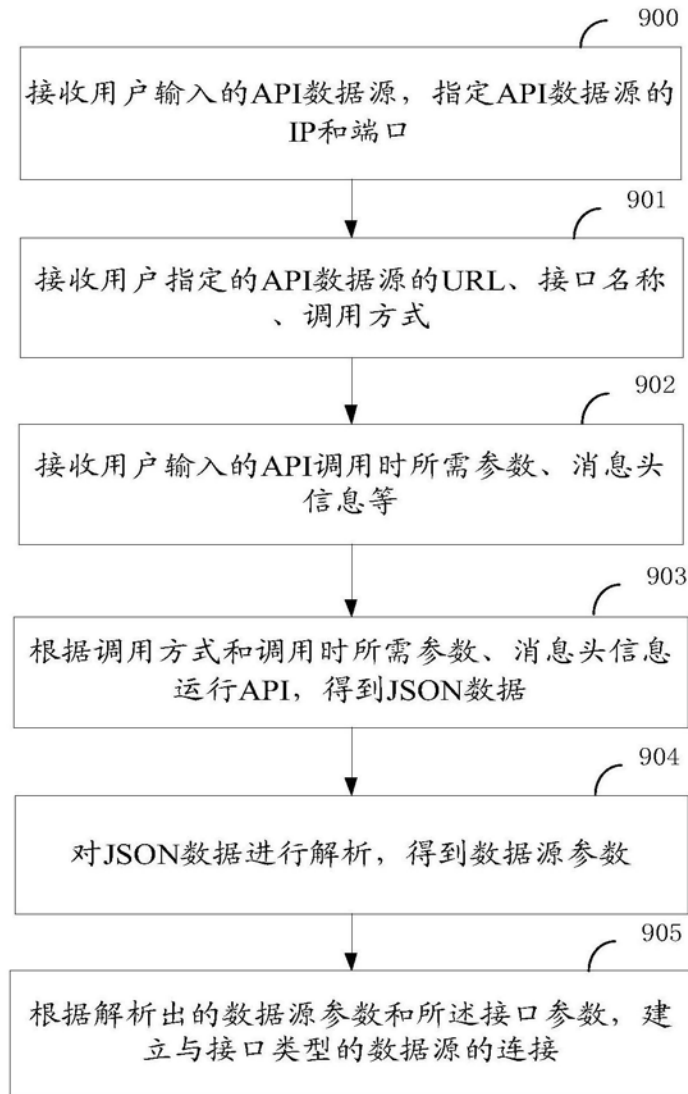


图9

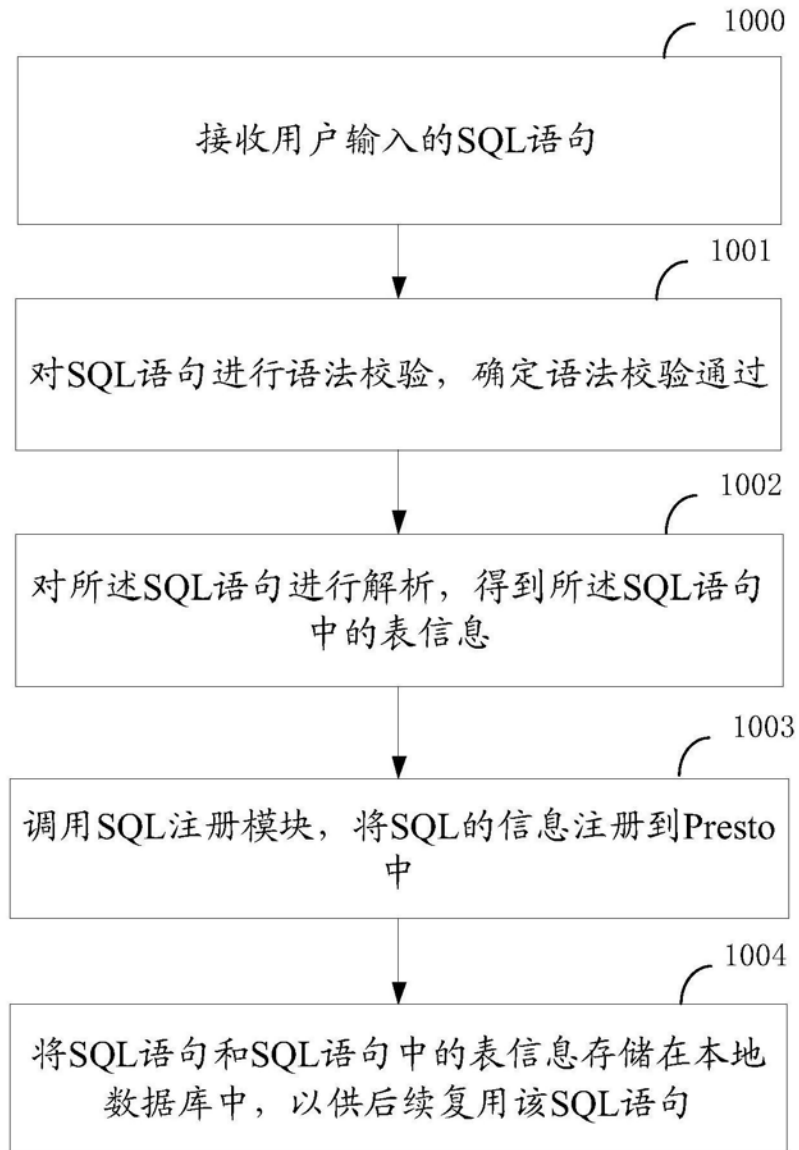


图10



图11

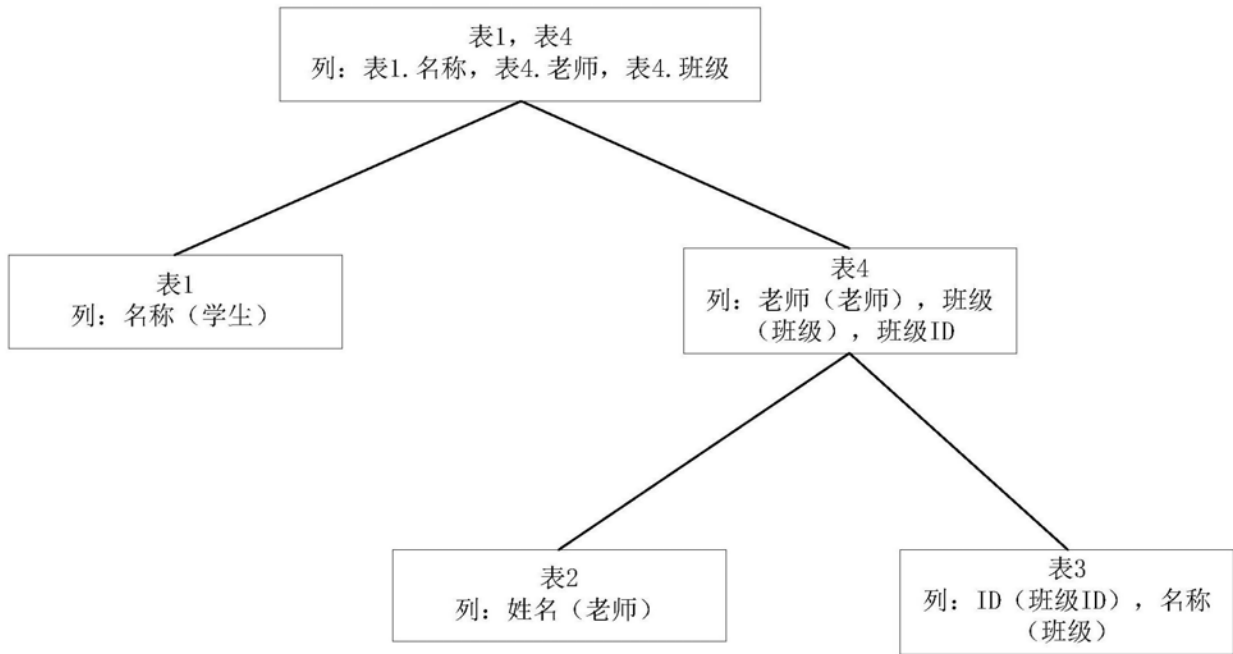


图12

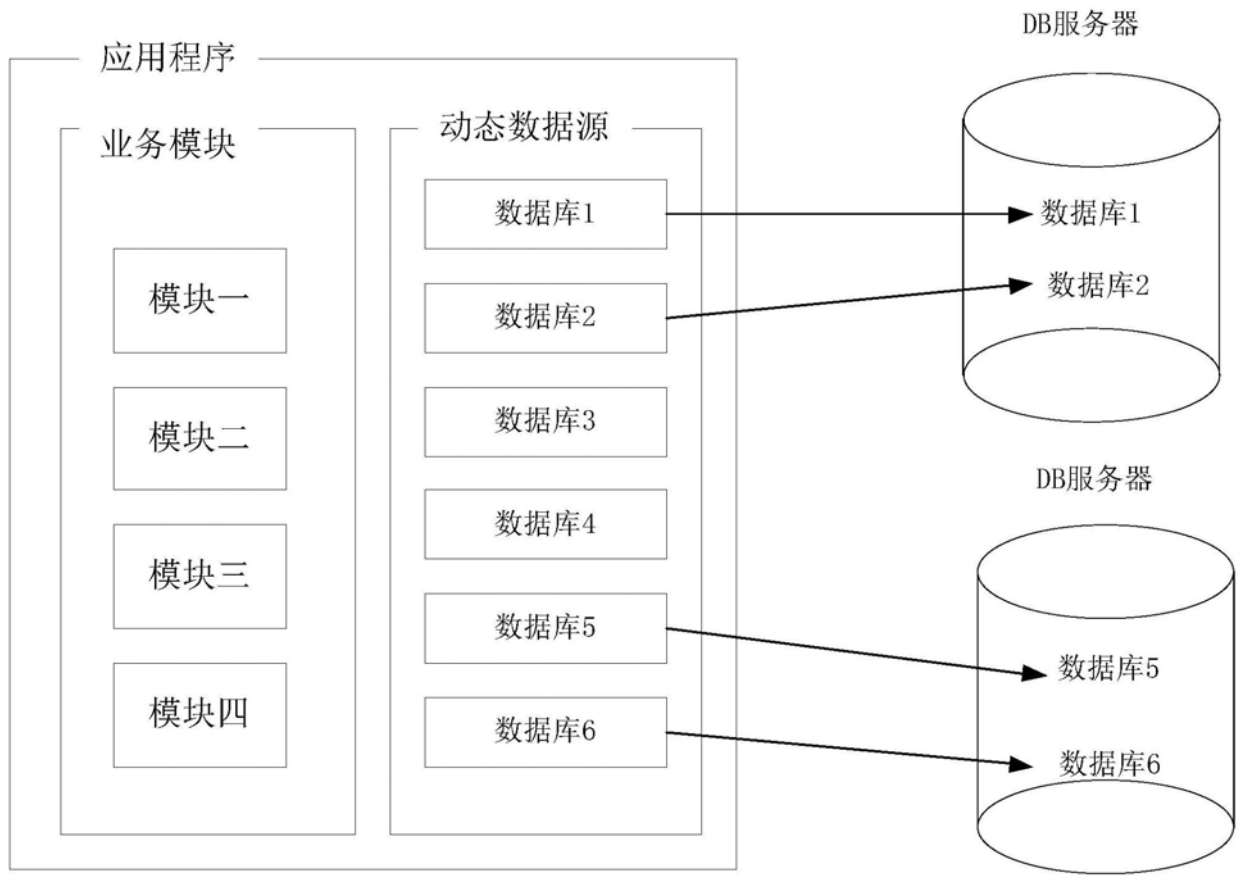


图13

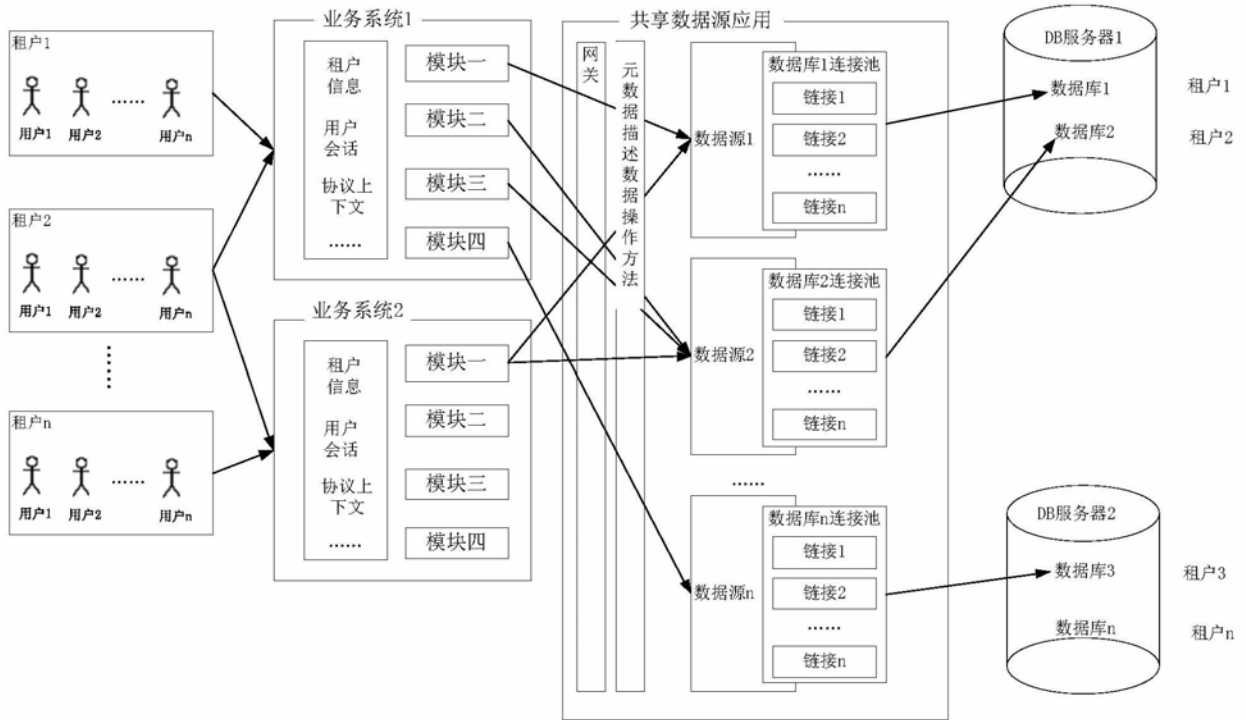


图14

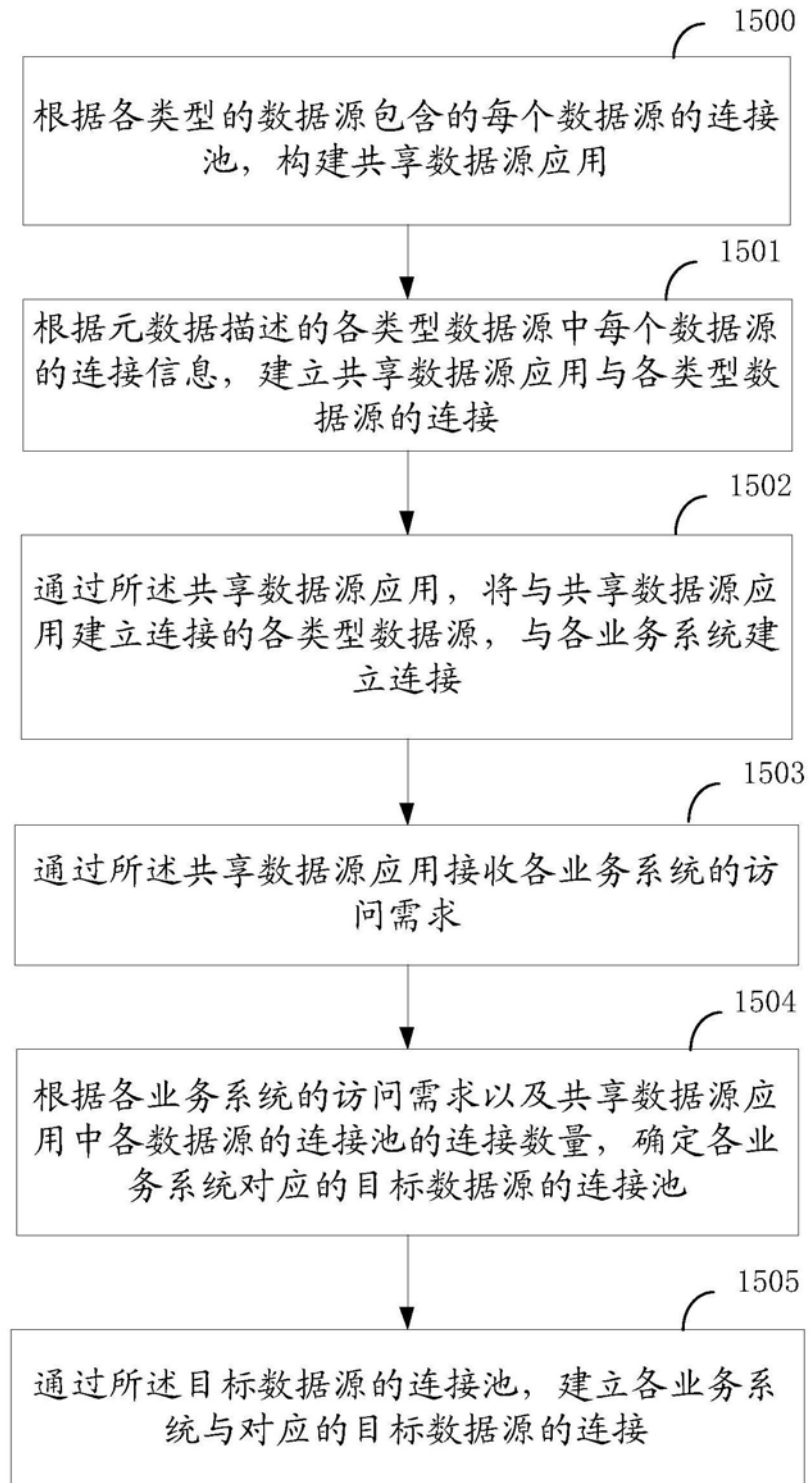


图15



图16



图17

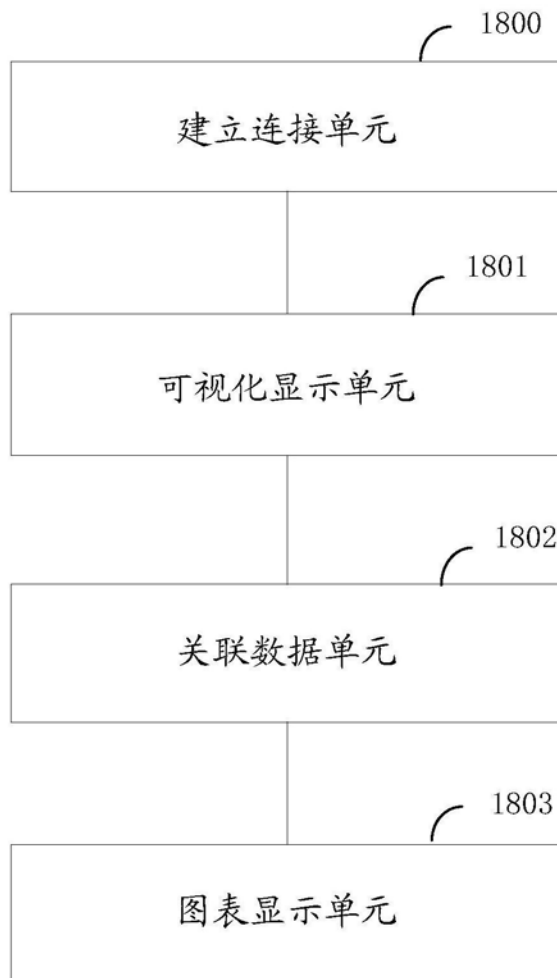


图18