

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2024年5月10日 (10.05.2024)



(10) 国际公布号
WO 2024/092926 A1

- (51) 国际专利分类号:
G06F 40/18 (2020.01)
- (21) 国际申请号: PCT/CN2022/135215
- (22) 国际申请日: 2022年11月30日 (30.11.2022)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202211351865.3 2022年10月31日 (31.10.2022) CN
- (71) 申请人: 蚂蚁区块链科技(上海)有限公司
(**ANT BLOCKCHAIN TECHNOLOGY (SHANGHAI) CO., LTD.**) [CN/CN]; 中国上海市黄浦区外马路618号8层803室, Shanghai 200010 (CN).
- (72) 发明人: 翟艳堂 (**ZHAI, Yantang**); 中国浙江省杭州市西湖区西溪路556号8层B段801-11, Zhejiang

310000 (CN)。杨仁慧 (**YANG, Renhui**); 中国浙江省杭州市西湖区西溪路556号8层B段801-11, Zhejiang 310000 (CN)。孙善禄 (**SUN, Shanlu**); 中国浙江省杭州市西湖区西溪路556号8层B段801-11, Zhejiang 310000 (CN)。

(74) 代理人: 北京博思佳知识产权代理有限公司 (**BEIJING BESTIPR INTELLECTUAL PROPERTY LAW CORPORATION**); 中国北京市海淀区上地三街9号嘉华大厦B座409, Beijing 100085 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN,

(54) **Title:** METHOD AND DEVICE FOR GENERATING DATA TABLE

(54) 发明名称: 生成数据表的方法及装置

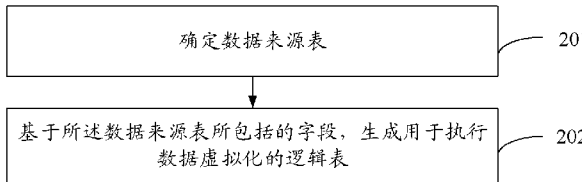


图 2

- 201 Determine a data source table
202 On the basis of fields comprised in the data source table, generate a logical table used for executing data virtualization

(57) **Abstract:** One or more embodiments of the present description provide a method and device for generating a data table. The method comprises: determining a data source table, wherein the data source table is at least one of a data warehouse table of a current organizational unit, a data federation table across organizational units, and an raw data table across organizational units; and on the basis of fields comprised in the data source table, generating a logical table used for executing data virtualization, wherein the logical table is used for providing a data query result for data application. In the present disclosure, the logical table used for executing data virtualization can be generated, and compatibility and unification of data objects on a current organizational unit data warehouse system and a cross-organizational-unit data fusion system are achieved by means of data virtualization objects of the logical table.

(57) 摘要: 本说明书一个或多个实施例提供一种生成数据表的方法及装置, 其中, 该方法包括: 确定数据来源表, 其中, 所述数据来源表是当前机构域的数据仓库表、跨机构域的数据联邦表以及跨机构域的原始数据表中的至少一个; 基于所述数据来源表所包括的字段, 生成用于执行数据虚拟化的逻辑表; 其中, 所述逻辑表用于为数据应用提供数据查询结果。在本公开中, 可以生成用于执行数据虚拟化的逻辑表, 通过逻辑表的数据虚拟化对象, 达到本域数据仓库和跨机构域数据融合两种系统之上数据对象的兼容统一。

MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

- (84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

- 包括国际检索报告(条约第21条(3))。

生成数据表的方法及装置

技术领域

[01] 本说明书一个或多个实施例涉及终端技术领域，尤其涉及一种生成数据表的方法及装置。

背景技术

[02] 随着数字化程度越来越高，越来越多的机构建设或者使用了数据仓库系统，为上层数据应用提供支撑服务。随着数据驱动力越来越强，而且数据应用对效果的要求越来越高，数据应用不仅使用本机构的数据，还寻求使用其他机构的数据。随着数据合规要求越来越严格，若想使用其他机构的数据，直接连接或者直接采集的方式越来越少，更多的选择建设或者使用合规的跨机构域数据融合系统，比如基于隐私计算技术构建的系统。这样数据应用的数据来源既有来自本域数据仓库系统的数据，又有来自跨机构域数据融合系统的数据。

[03] 但是本机构域数据仓库系统和跨机构域数据融合系统相互独立，两种系统中的数据相互隔离，对于上层数据应用没有形成统一的数据视角，也没有形成统一的服务对接。

发明内容

[04] 有鉴于此，本说明书一个或多个实施例提供一种生成数据表的方法及装置，通过逻辑表的数据虚拟化对象，达到本机构域数据仓库和跨机构域数据融合两种系统之上，数据对象的兼容统一。

[05] 根据本说明书一个或多个实施例的第一方面，提出了一种生成数据表的方法，包括：确定数据来源表，其中，所述数据来源表是当前机构域的数据仓库表、跨机构域的数据联邦表以及跨机构域的原始数据表中的至少一个；基于所述数据来源表所包括的字段，生成用于执行数据虚拟化的逻辑表；其中，所述逻辑表用于为数据应用提供数据查询结果。

[06] 根据本说明书一个或多个实施例的第二方面，提出了一种生成数据表的装置，包括：处理模块，用于确定数据来源表，其中，所述数据来源表是当前机构域的数据仓库表、跨机构域的数据联邦表以及跨机构域的原始数据表中的至少一个；生成模块，用于

基于所述数据来源表所包括的字段，生成用于执行数据虚拟化的逻辑表；其中，所述逻辑表用于为数据应用提供数据查询结果。

[07] 根据本说明书一个或多个实施例的第三方面，提出了一种数据虚拟化系统，包括：用于执行数据虚拟化的逻辑表，所述逻辑表是基于数据来源表所包括的字段生成的，且所述逻辑表用于为数据应用提供数据查询结果，所述数据来源表是当前机构域的数据仓库表、跨机构域的数据联邦表以及跨机构域的原始数据表中的至少一个；查询引擎，用于为数据应用提供查询所述逻辑表的所述数据查询结果。

[08] 根据本说明书一个或多个实施例的第四方面，提出了一种电子设备，包括：处理器；用于存储处理器可执行指令的存储器；其中，所述处理器通过运行所述可执行指令以实现如第一方面中任一项所述的生成数据表的方法。

[09] 根据本说明书一个或多个实施例的第五方面，提出了一种计算机可读存储介质，其上存储有计算机指令，该指令被处理器执行时实现如第一方面中任一项所述生成数据表的方法的步骤。

[10] 本说明书的实施例提供的技术方案可以包括以下有益效果：在本公开中，可以生成用于执行数据虚拟化的逻辑表，该逻辑表可以用于为数据应用提供数据查询结果，通过逻辑表的数据虚拟化对象，达到本域数据仓库和跨机构域数据融合两种系统之上数据对象的兼容统一。此外，可以实现跨机构域数据融合和本域数据仓库两种系统之上的统一的查询。且实现了逻辑表的自动物化，可用性高。

附图说明

[11] 图 1 是一示例性实施例提供的两种数据系统相互粒度的场景示意图。

[12] 图 2 是一示例性实施例提供的一种生成数据表的方法的流程图。

[13] 图 3 是一示例性实施例提供的一种基于逻辑表生成新的逻辑表的场景示意图。

[14] 图 4 是一示例性实施例提供的另一种生成数据表的方法的流程图。

[15] 图 5 是一示例性实施例提供的一种查询引擎的结构示意图。

[16] 图 6A 至图 6B 是一示例性实施例提供的查询计划示意图。

[17] 图 7 是一示例性实施例提供的另一种生成数据表的方法的流程图。

[18] 图 8 是一示例性实施例提供的一种存储物化表的场景示意图。

[19] 图 9 是一示例性实施例提供的一种数据虚拟化系统的结构示意图。

[20] 图 10 是一示例性实施例提供的一种电子设备的结构示意图。

[21] 图 11 是一示例性实施例提供的一种生成数据表的装置的框图。

具体实施方式

[22] 这里将详细地对示例性实施例进行说明，其示例表示在附图中。下面的描述涉及附图时，除非另有表示，不同附图中的相同数字表示相同或相似的要素。以下示例性实施例中所描述的实施方式并不代表与本发明说明书一个或多个实施例相一致的所有实施方式。相反，它们仅是与如所附权利要求书中所详述的、本发明说明书一个或多个实施例的一些方面相一致的装置和方法的例子。

[23] 需要说明的是：在其他实施例中并不一定按照本发明说明书示出和描述的顺序来执行相应方法的步骤。在一些其他实施例中，其方法所包括的步骤可以比本发明说明书所描述的更多或更少。此外，本发明说明书中所描述的单个步骤，在其他实施例中可能被分解为多个步骤进行描述；而本发明说明书中所描述的多个步骤，在其他实施例中也可能被合并为单个步骤进行描述。

[24] 在介绍本发明公开提供的方案之前，先介绍一下本发明公开涉及到的术语。

[25] 数据仓库：数据仓库是信息的中央集中存储库，用于存储和处理数据，以便对数据进行分析。数据一般定期或者实时从在线事务系统、关系型数据库、消息系统和其他系统采集流入到数据仓库。本提案中提到的数据仓库特指大数据仓库，是基于大数据存储和计算系统构建的数据仓库。

[26] 本域数据仓库：归属于本机构域的数据仓库，存储的数据归属于本机构法律主体或者合规授权给本机构法律主体。

[27] 跨机构域数据融合：不同机构之间数据的流转、共享、分析、计算等，和合规要求下建设在不同的机构之间，为了打破机构之间数据孤岛，共同发挥数据的更大价值。

[28] 隐私计算：从计算讲，它是一组技术统称，以解决数据计算过程中的数据安全和隐私保护的问题。以多方安全计算（Secure Multi-Party Computation, MPC）、联邦学习（Federated Learning, FL）、可信执行环境（Trusted Execution Environment, TEE）等技术为代表。

[29] 联邦表：一种跨机构域数据融合系统之上的数据虚拟化对象，对上层屏蔽多数据集的分散形态。

[30] 动态数据管理框架 Apache Calcite: Apache Calcite 用于构建数据库或者数据管理系统的开源框架。包括一个结构化查询语言（Structured Query Language, SQL）解析器，一个用于在关系代数中构建表达式的应用程序接口（Application Program Interface, API）和一个查询计划引擎。

[31] 开源语法分析器 ANTLR: ANTLR 的全名是 ANother Tool for Language Recognition, 是基于 LL 算法实现的语法解析器生成器，广泛用于构建语言、工具和框架。

[32] 元数据：用于描述数据属性（property）的信息，某个数据表的元数据可以理解为该数据表的字段名，例如，某个数据表的元数据包括：用户标识 id、性别 gender、年龄 age 等。

[33] 物理数据：用于描述数据的具体信息，某个数据表的物理数据可以理解为该数据表的字段值，例如，某个数据表的元数据包括：用户标识 id、性别 gender、年龄 age 等，物理数据包括：id#1、女、28 岁等。

[34] 数据虚拟化：用来描述数据管理方法，这些方法允许应用程序，例如数据应用检索并管理数据，且不需要数据相关的技术细节，例如数据格式化的方式或物理位置所在。其中，物理位置在本公开中可以理解为机构域所对应的地理位置。

[35] 随着数字化程度越来越高，越来越多的机构建设或者使用了数据仓库系统，为上层数据应用提供支撑服务。随着数据驱动力越来越强，而且数据应用对效果的要求越来越高，数据应用不仅使用本机构的数据，还寻求使用其他机构的数据。随着数据合规要求越来越严格，若想使用其他机构的数据，直接连接或者直接采集的方式越来越少，更多的选择建设或者使用合规的跨机构域数据融合系统，比如基于隐私计算技术构建的系统。

[36] 这样数据应用的数据来源既有来自本域数据仓库系统的数据，又有来自跨机构域数据融合系统的数据。但是本域数据仓库系统和跨机构域数据融合系统相互独立，两种系统中的数据相互隔离，对于上层数据应用没有形成统一的数据视角，也没有形成统一的服务对接，例如图 1 所示。

[37] 示例性地，某风控类的数据应用，既使用本域内的数据仓库系统的数据，又通过数据融合系统使用其他机构的数据做联合风控，使用的数据集既有数据仓库的数仓表，

又有数据融合系统的联邦表或者原始数据表，而且该数据应用需要分别对接这两种系统。

[38] 但是本机构域数据仓库系统和跨机构域数据融合系统相互独立，两种系统中的数据相互隔离，对于上层数据应用没有形成统一的数据视角，也没有形成统一的服务对接。

[39] 为了解决这一技术问题，本公开提供了以下生成数据的方法及装置。

[40] 图 2 是一示例性实施例提供的一种生成数据表的方法流程图。请参考图 2，该方法可以由服务器执行，该服务器可以是当前机构域内用于提供数据服务，例如数据查询、数据存储、数据更新等的服务器，包括步骤 201 至步骤 202。

[41] 在步骤 201 中，确定数据来源表。

[42] 在本公开实施例中，所述数据来源表可以是当前机构域的数据仓库表、跨机构域的数据联邦表以及跨机构域的原始数据表中的至少一个。

[43] 其中，跨机构域可以理解为跨域名，一般情况下，不同的域名对应不同的地理区域。

[44] 在一个示例中，假设当前机构域为机构域 B，数据来源表可以包括机构域 C 的数据仓库表。

[45] 在另一个示例中，假设当前机构域为机构域 B，数据来源表可以包括跨机构域 A 和机构域 C 的联邦表。

[46] 在另一个示例中，假设当前机构域为机构域 B，数据来源表可以包括机构域 B 的数据仓库表和跨机构域 A 和机构域 C 的原始数据表。

[47] 以上仅为示例性说明，数据来源表至少是当前机构域的数据仓库表、跨机构域的数据联邦表以及跨机构域的原始数据表中的至少一个的任意一种情况，均应属于本公开的保护范围。

[48] 在步骤 202 中，基于所述数据来源表所包括的字段，生成用于执行数据虚拟化的逻辑表。

[49] 在本公开实施例中，逻辑表可以用于为数据应用提供数据查询结果。

[50] 逻辑表是从业务视角定义的一份数据视图，数据来源于本机构域内的数据仓库表和跨机构域数据融合的联邦表或者跨机构域的原始数据表，本公开对此不作限定。

[51] 在一个示例中，逻辑表可以只存储元数据，不存储物理数据。逻辑表的目的是为

了统一规范的数据建模，基于逻辑表可以把数据仓库数据和数据融合数据放在一起建设统一的数据模型。基于逻辑表还可以创建其他的逻辑表，例如图 3 所示，在基于本机构域的数据仓库表（即数仓表）和跨机构域的联邦表生成一张逻辑表后，可以基于该逻辑表生成其他的逻辑表，图 3 中又生成了两张逻辑表。基于逻辑表生成新的逻辑表的方式与生成逻辑表的方式类似，生成逻辑表的方式如下：在一个示例中，可以采用预定义规则，确定逻辑表名。

[52] 进一步地，可以在所述数据来源表所包括的字段中，确定生成所述逻辑表所需要的指定字段，其中，指定字段是数据来源表所包括的一个或多个字段。

[53] 进一步地，可以确定所述逻辑表的输出字段与所述指定字段之间的计算逻辑关系。

[54] 在本公开实施例中，逻辑表的输出字段可以与指定字段相同，或者逻辑表的输出字段需要基于指定字段之间的计算逻辑关系确定。

[55] 例如，指定字段包括买方标识 buyer_id、卖方标识 seller_id。逻辑表的输出字段可以包括买方标识 buyer_id、卖方标识 seller_id。

[56] 再例如，指定字段包括买方标识 buyer_id、卖方标识 seller_id、订单交易量 quantity，逻辑表的输出字段包括日最大订单交易量 quantity_max_1d。

[57] 其中，逻辑表的输出字段 quantity_max_1d 与所述指定字段之间的计算逻辑关系为：按照 buyer_id、seller_id 分组（即统计同一天内）对 quantity 求最大值。

[58] 再进一步地，可以基于所述逻辑表名、所述数据来源表、所述指定字段以及所述计算逻辑关系中的至少一项，生成所述逻辑表。

[59] 示例性地，可以基于预定义规则，确定逻辑表名为 trade_indicator。

[60] 进一步地，数据来源表包括：当前机构域的数据仓库表 MAXCOMPUTE.default.order，以及机构域 A 和机构 C 的联邦表 FeDX.default.event，数据来源表包括多个时，多个数据来源表之间的关联条件是 join，join 的字段和条件是 order.id = event.id。

[61] 选择的数据来源表的指定字段包括：买方标识 event.buyer_id、卖方标识 event.seller_id、订单交易量 order.quantity、交易渠道 order.channel、价格 event.price。

[62] 逻辑表的输出字段包括：买方标识 buyer_id、卖方标识 seller_id、日最大订单交易量 quantity_max_1d、云上的周订单交易收入总和 amount_cloud_7d。

[63] 计算逻辑关系包括：输出字段 quantity_max_1d 与指定字段 buyer_id、seller_id、quantity 之间的计算逻辑关系包括：按照 buyer_id、seller_id 分组对 quantity 求最大值；输出字段 amount_cloud_7d 字段与指定字段 buyer_id、seller_id、quantity、price 之间的计算逻辑关系包括：统计 7 天数据量，按照 buyer_id、seller_id 分组（统计七日内）对 quantity × price 求和，限制条件是交易渠道 channel = 'cloud'（云上）。

[64] 以上仅为示例性说明，实际应用中，将当前机构域的数据仓库表、跨机构域的数据联邦表以及跨机构域的原始数据表中的至少一个作为数据来源表，从而生成用于执行数据虚拟化的逻辑表的方案均应属于本公开的保护范围。

[65] 上述实施例中，可以生成用于执行数据虚拟化的逻辑表，该逻辑表可以用于为数据应用提供数据查询结果，通过逻辑表的数据虚拟化对象，达到本域数据仓库和跨机构域数据融合两种系统之上数据对象的兼容统一，可用性高。

[66] 在一些可选实施例中，图 4 是基于图 2 所示实施例提供的另一种生成数据表的方法流程图。请参考图 4，该方法可以由服务器执行，该服务器可以是当前机构域内用于提供数据服务，例如数据查询、数据存储、数据更新等的服务器，所述方法还包括步骤 203。

[67] 在步骤 203 中，为数据应用提供查询所述逻辑表的查询引擎。

[68] 在本公开实施例中，可以为数据应用提供查询上述逻辑表的查询引擎，示例性地，该查询引擎的结构可以例如图 5 所示。

[69] 在一个示例中，查询引擎支持的查询语言为 SQL 语言，可以为标准 SQL 语言的子集，并有适量语法扩展，本公开对此不作限定。该查询引擎针对数据应用的对接接口同样支持 SQL 语句。

[70] 其中，图 5 中查询引擎的解析可以由 SQL parser 执行，SQL parser 可以采用 Apache Calcit 或 ANTLR 来实现，本公开对此不作限定。

[71] 在另一个示例中，查询引擎用于存储所述逻辑表的元数据。即查询引擎不直接存储各个机构域物理表的元数据。

[72] 在本公开实施例中，逻辑表的元数据校验和/或鉴权由查询引擎来实现，物理表的元数据校验和/或鉴权通过查询引擎调用物理表各自所在机构域的计算引擎来实现。

[73] 在另一个示例中，可以由所述查询引擎生成跨机构域的查询计划，其中，该查询

计划可以基于数据应用提供的查询语句来生成。

[74] 进一步地，由所述查询引擎将所述查询计划拆分为至少一个查询子计划，其中，所述查询子计划与所述数据来源表的计算引擎一一对应。也就是说，由查询引擎将查询计划按照每个数据来源表所对应的计算引擎分别进行拆分。

[75] 每个计算引擎执行查询子计划得到查询子结果后，可以由查询引擎合并所述查询子结果，并将得到的所述数据查询结果提供给所述数据应用。

[76] 示例性地，查询引擎接收到来自数据应用的查询语句如下：

```
SELECT COUNT(f1), SUM(f2) FROM
(
    (SELECT id as f1, frequency as f2 FROM FeDX.default.consumer_features
    WHERE id <> '1234') T1
    UNION ALL
    (SELECT c1 as f1, c2 as f2 FROM MAXCOMPUTE.default.dim_user WHERE id
    <> '1234') T2
)
```

[77] 进一步地，查询引擎将其翻译成查询计划，例如图 6A 所示。

[78] 其中逻辑表对应的数据来源表包括当前机构域的数据仓库表和数据融合系统的联邦表，分别对应的计算引擎为数据仓库计算引擎、数据融合系统计算引擎。查询引擎将上述的查询计划拆分成三个部分，包括由查询引擎所执行的第一部分，由数据仓库计算引擎所执行的第二部分和由数据融合系统计算引擎所执行的第三部分，例如图 6B 所示。

[79] 第二部分和第三部分的查询子计划分别由数据仓库计算引擎和数据融合系统计算引擎执行，得到对应的查询子结果，查询引擎合并所述查询子结果，最终得到数据查询结果。

[80] 上述实施例中，通过查询引擎可以实现跨机构域数据融合和本机构域数据仓库两种系统之上的统一的查询。

[81] 在一些可选实施例中，图 7 是基于图 4 所示实施例提供的另一种生成数据表的方法流程图。请参考图 7，该方法可以由服务器执行，该服务器可以是当前机构域内用于提供数据服务，例如数据查询、数据存储、数据更新等的服务器，所述方法还包括步骤

204。

[82] 在步骤 204 中，当对所述逻辑表执行物化时，生成用于承载所述逻辑表的物理数据的物化表。

[83] 在一个示例中，可以由物化引擎生成所述物化表。

[84] 在一个示例中，逻辑表的物化与数据库中的物化视图主要区别点有以下几项。

[85] 首先，数据库中的物化视图是一定有数据存储的；而逻辑表不一定有数据存储，在打开物化的时候，才会有数据存储，如果没有打开物化，是没有数据存储的。

[86] 其次，物化视图的数据存储是由其本身存储；而逻辑表在物化的时候，逻辑表的物化数据存储由单独的物理表来承载，逻辑表和其对应的物理表共同组成了物化了的逻辑表。

[87] 再次，一个物化视图就是一个数据对象；而打开物化的一个逻辑表可以对应多个物理表。

[88] 再次，物化视图是直接面对使用者的；而打开物化的逻辑表，所产生的物理表是不直接面对使用者的，使用者还是直接使用逻辑表而不是所产生的物理表。

[89] 另外，对于物化到数据融合系统的数据，在一些合规要求下，有些数据是分散物化存储到各个机构域。

[90] 承载逻辑表物化数据存储的表称为物化表，物化表存储在底层的各个系统，逻辑表和物化表的对应关系如图 8 所示，其中，物化表 1、物化表 2、物化表 3 分别存储在 A 机构域内、当前机构域（即 B 机构域）的数据仓库、C 机构域内。

[91] 首先，可以在所述数据来源表所包括的字段中，确定生成所述逻辑表所需要的所述指定字段。确定指定字段的方式已经在上述实施例进行了介绍，此处不再赘述。

[92] 其次，需要确定所述物化表的输出字段与所述指定字段之间的物化逻辑关系。

[93] 进一步地，基于所述指定字段和所述物化逻辑关系，确定与所述物化表的输出字段所对应的物理数据。

[94] 进一步地，可以基于物化表名、所述物化表的输出字段和所述物理数据中的至少一项，生成所述物化表。

[95] 生成物化表的过程与生成逻辑表的过程类似，只是在生成物化表时需要确定物化

表输出字段所对应的物理数据。

[96] 示例性地，上文的逻辑表 trade_indicator，当对所述逻辑表执行物化时，即打开物化时，会自动生成两张物化表：trade_1d、trade_7d。

[97] 其中，物化表 trade_1d 对应的输出字段包括：buyer_id、seller_id、quantity_max_1d、dt；对应的物化逻辑如下：

```
INSERT OVERWRITE TABLE trade_1d PARTITION (dt)
SELECT
    trade.buyer_id as buyer_id,
    trade.seller_id as seller_id,
    max(trade.quantity) as quantity_max_1d,
    '${bizdate}' as dt
FROM
    (SELECT event.buyer_id as buyer_id, event.seller_id as seller_id, order.quantity as
quantity, order.dt as dt
    from MAXCOMPUTE.default.order join FeDX.default.event on order.id = event.id and
order.dt = event.dt
    ) as trade
WHERE trade.dt = '${bizdate}'
GROUP BY trade.buyer_id, trade.seller_id;
```

[98] 其中，物化表 trade_7d 对应输出的字段如下：buyer_id、seller_id、amount_cloud_7d、dt；对应的物化逻辑如下：

```
INSERT OVERWRITE TABLE trade_7d PARTITION (dt)
SELECT
    trade.buyer_id as buyer_id,
    trade.seller_id as seller_id,
    sum (trade.quantity * trade.price ) as amount_cloud_7d,
    '${bizdate}' as dt
FROM
```

```
(SELECT
    event.buyer_id as buyer_id, event.seller_id as seller_id,
    order.quantity as quantity, order.channel, event.price as price
    order.dt as dt
    from MAXCOMPUTE.default.order join FeDX.default.event on order.id = event.id and
order.dt = event.dt
) as trade
WHERE trade.dt <= '${bizdate}' and trade.dt >= '${bizdate-7d}'
AND trade.channel = 'cloud'
GROUP BY trade.buyer_id, trade.seller_id;
```

[99] 以上仅为示例性说明，本领域技术人员能够理解的是，基于逻辑表生成物化表的方案均应属于本公开的保护范围。

[100] 上述实施例中，可以实现逻辑表的自动物化，提高了数据查询效率，可用性高。

[101] 在一些可选实施例中，本公开提供了一种数据虚拟化系统，该系统可以部署在本机构域数据仓库系统和跨机构域数据融合系统之上，其结构示意图参考图9所示，包括：逻辑表、查询引擎，还可以包括物化引擎。

[102] 其中，逻辑表的生成方式与上述图2所示实施例类似，在此不再赘述。

[103] 其中，查询引擎的结构与上述图5类似，查询引擎所执行的操作与上述图4所示类似，在此同样不再赘述。

[104] 其中，物化引擎生成物化表的方式已经在上述实施例进行了介绍，此处不再赘述。

[105] 上述实施例中，可以通过逻辑表的数据虚拟化对象，达到本域数据仓库和跨机构域数据融合两种系统之上数据对象的兼容统一。此外，可以实现跨机构域数据融合和本域数据仓库两种系统之上的统一的查询，且实现了逻辑表的自动物化，提高了数据查询效率，可用性高。

[106] 图10是一示例性实施例提供的一种电子设备的示意结构图，该电子设备可以为数据服务器，本公开对此不作限定。请参考图10，在硬件层面，该设备包括处理器1002、内部总线1004、网络接口1006、内存1008以及非易失性存储器1010，当然还可能包括其他业务所需要的硬件。本说明书一个或多个实施例可以基于软件方式来实现，比如由

处理器 1002 从非易失性存储器 1010 中读取对应的计算机程序到内存 1008 中然后运行。当然，除了软件实现方式之外，本说明书一个或多个实施例并不排除其他实现方式，比如逻辑器件抑或软硬件结合的方式等等，也就是说以下处理流程的执行主体并不限定于各个逻辑单元，也可以是硬件或逻辑器件。

[107] 请参考图 11，生成数据表的装置可以应用于如图 10 所示的设备中，以实现本说明书的技术方案。其中，该生成数据表的装置可以包括：处理模块 1101，用于确定数据来源表，其中，所述数据来源表是当前机构域的数据仓库表、跨机构域的数据联邦表以及跨机构域的原始数据表中的至少一个；生成模块 1102，用于基于所述数据来源表所包括的字段，生成用于执行数据虚拟化的逻辑表；其中，所述逻辑表用于为数据应用提供数据查询结果。

[108] 上述实施例阐明的系统、装置、模块或单元，具体可以由计算机芯片或实体实现，或者由具有某种功能的产品来实现。一种典型的实现设备为计算机，计算机的具体形式可以是个人计算机、膝上型计算机、蜂窝电话、相机电话、智能电话、个人数字助理、媒体播放器、导航设备、电子邮件收发设备、游戏控制台、平板计算机、可穿戴设备或者这些设备中的任意几种设备的组合。

[109] 在一个典型的配置中，计算机包括一个或多个处理器（CPU）、输入/输出接口、网络接口和内存。

[110] 内存可能包括计算机可读介质中的非永久性存储器，随机存取存储器（RAM）和/或非易失性内存等形式，如只读存储器（ROM）或闪存（flash RAM）。内存是计算机可读介质的示例。

[111] 计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括，但不限于相变内存（PRAM）、静态随机存取存储器（SRAM）、动态随机存取存储器（DRAM）、其他类型的随机存取存储器（RAM）、只读存储器（ROM）、电可擦除可编程只读存储器（EEPROM）、快闪记忆体或其他内存技术、只读光盘只读存储器（CD-ROM）、数字多功能光盘（DVD）或其他光学存储、磁盒式磁带、磁盘存储、量子存储器、基于石墨烯的存储介质或其他磁性存储设备或任何其他非传输介质，可用于存储可以被计算设备访问的信息。按照本文中的界定，计算机可读介质不包括暂存电脑可读媒体（transitory media），如调制的数据信号和载波。

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

[113] 上述对本说明书特定实施例进行了描述。其它实施例在所附权利要求书的范围内。在一些情况下，在权利要求书中记载的动作或步骤可以按照不同于实施例中的顺序来执行并且仍然可以实现期望的结果。另外，在附图中描绘的过程不一定要求示出的特定顺序或者连续顺序才能实现期望的结果。在某些实施方式中，多任务处理和并行处理也是可以的或者可能是有利的。

[114] 在本说明书一个或多个实施例使用的术语是仅仅出于描述特定实施例的目的，而非旨在限制本说明书一个或多个实施例。在本说明书一个或多个实施例和所附权利要求书中所使用的单数形式的“一种”、“所述”和“该”也旨在包括多数形式，除非上下文清楚地表示其他含义。还应当理解，本文中使用的术语“和/或”是指并包含一个或多个相关联的列出项目的任何或所有可能组合。

[115] 应当理解，尽管在本说明书一个或多个实施例可能采用术语第一、第二、第三等来描述各种信息，但这些信息不应限于这些术语。这些术语仅用来将同一类型的信息彼此区分开。例如，在不脱离本说明书一个或多个实施例范围的情况下，第一信息也可以被称为第二信息，类似地，第二信息也可以被称为第一信息。取决于语境，如在此所使用的词语“如果”可以被解释成为“在……时”或“当……时”或“响应于确定”。

[116] 以上所述仅为本说明书一个或多个实施例的较佳实施例而已，并不用以限制本说明书一个或多个实施例，凡在本说明书一个或多个实施例的精神和原则之内，所做的任何修改、等同替换、改进等，均应包含在本说明书一个或多个实施例保护的范围之内。

权利要求书

1、一种生成数据表的方法，其特征在于，包括：

确定数据来源表，其中，所述数据来源表是当前机构域的数据仓库表、跨机构域的数据联邦表以及跨机构域的原始数据表中的至少一个；

基于所述数据来源表所包括的字段，生成用于执行数据虚拟化的逻辑表；其中，所述逻辑表用于为数据应用提供数据查询结果。

2、根据权利要求1所述的方法，其特征在于，所述基于所述数据来源表所包括的字段，生成用于执行数据虚拟化的逻辑表，包括：

确定逻辑表名；

在所述数据来源表所包括的字段中，确定生成所述逻辑表所需要的指定字段；

确定所述逻辑表的输出字段与所述指定字段之间的计算逻辑关系；

基于所述逻辑表名、所述数据来源表、所述指定字段以及所述计算逻辑关系中的至少一项，生成所述逻辑表。

3、根据权利要求1所述的方法，其特征在于，还包括：

为数据应用提供查询所述逻辑表的查询引擎。

4、根据权利要求3所述的方法，其特征在于，所述查询引擎用于存储所述逻辑表的元数据。

5、根据权利要求4所述的方法，其特征在于，还包括：

由所述查询引擎对所述逻辑表的元数据进行校验和/或鉴权。

6、根据权利要求4所述的方法，其特征在于，还包括：

由所述查询引擎生成跨机构域的查询计划；

由所述查询引擎将所述查询计划拆分为至少一个查询子计划，其中，所述查询子计划与所述数据来源表的计算引擎一一对应；

由所述查询引擎获取每个所述计算引擎执行对应的所述查询子计划的查询子结果；

由所述查询引擎合并所述查询子结果，并将合并得到的所述数据查询结果提供给所述数据应用。

7、根据权利要求1-6任一项所述的方法，其特征在于，还包括：

当对所述逻辑表执行物化时，生成用于承载所述逻辑表的物理数据的物化表。

8、根据权利要求7所述的方法，其特征在于，所述生成用于承载所述逻辑表的物理数据的物化表，包括：

在所述数据来源表所包括的字段中，确定生成所述逻辑表所需要的指定字段；

确定所述物化表的输出字段与所述指定字段之间的物化逻辑关系；

基于所述指定字段和所述物化逻辑关系，确定与所述物化表的输出字段所对应的物理数据；

基于物化表名、所述物化表的输出字段和所述物理数据中的至少一项，生成所述物化表。

9、根据权利要求 7 所述的方法，其特征在于，所述生成用于承载所述逻辑表的物理数据的物化表，包括：

由物化引擎生成所述物化表。

10、一种生成数据表的装置，其特征在于，包括：

处理模块，用于确定数据来源表，其中，所述数据来源表是当前机构域的数据仓库表、跨机构域的数据联邦表以及跨机构域的原始数据表中的至少一个；

生成模块，用于基于所述数据来源表所包括的字段，生成用于执行数据虚拟化的逻辑表；其中，所述逻辑表用于为数据应用提供数据查询结果。

11、一种数据虚拟化系统，其特征在于，包括：

用于执行数据虚拟化的逻辑表，所述逻辑表是基于数据来源表所包括的字段生成的，且所述逻辑表用于为数据应用提供数据查询结果，所述数据来源表是当前机构域的数据仓库表、跨机构域的数据联邦表以及跨机构域的原始数据表中的至少一个；

查询引擎，用于为数据应用提供查询所述逻辑表的所述数据查询结果。

12、根据权利要求 11 所述的系统，其特征在于，还包括：

物化引擎，用于当对所述逻辑表执行物化时，生成承载所述逻辑表的物理数据的物化表。

13、一种电子设备，其特征在于，包括：

处理器；

用于存储处理器可执行指令的存储器；

其中，所述处理器通过运行所述可执行指令以实现如权利要求 1-9 中任一项所述的生成数据表的方法。

14、一种计算机可读存储介质，其上存储有计算机指令，其特征在于，该指令被处理器执行时实现如权利要求 1-9 中任一项所述生成数据表的方法的步骤。

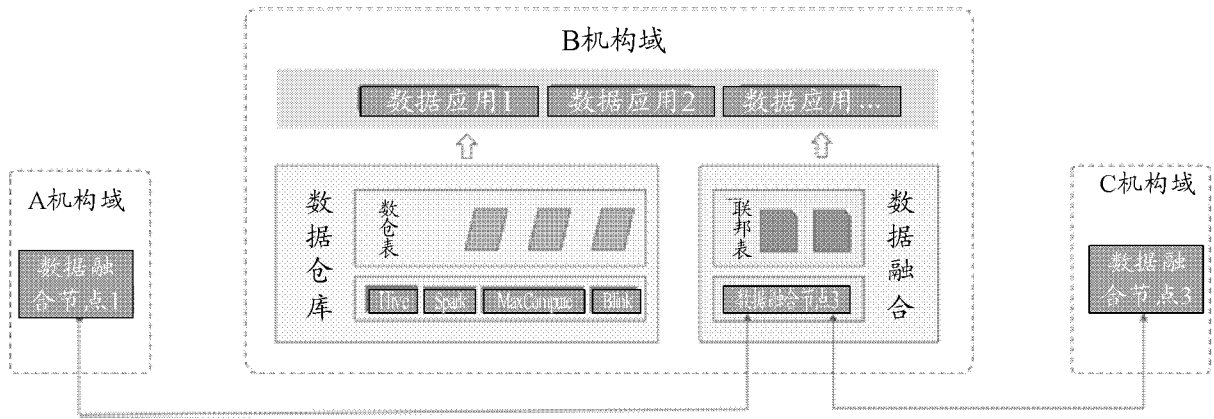


图 1

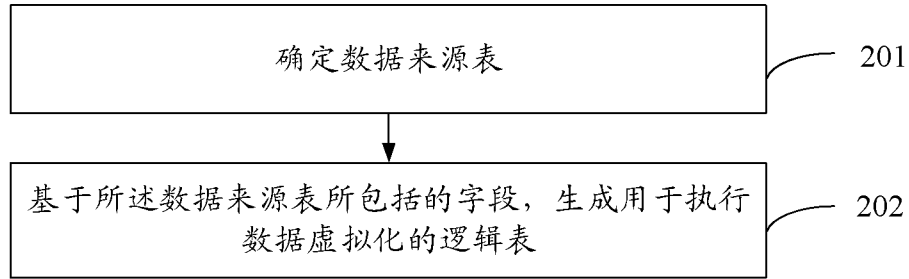


图 2

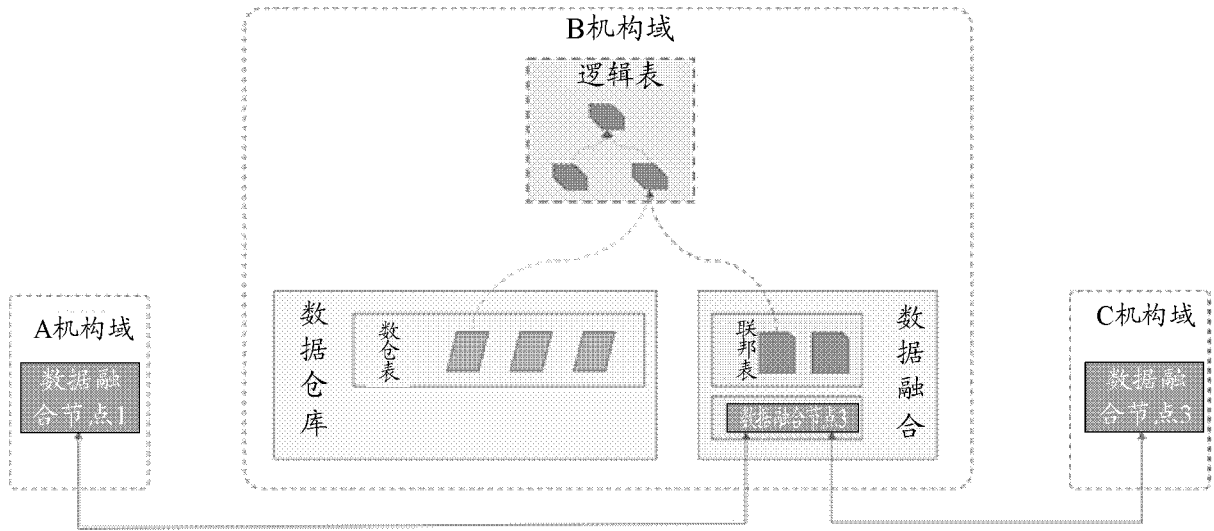


图 3

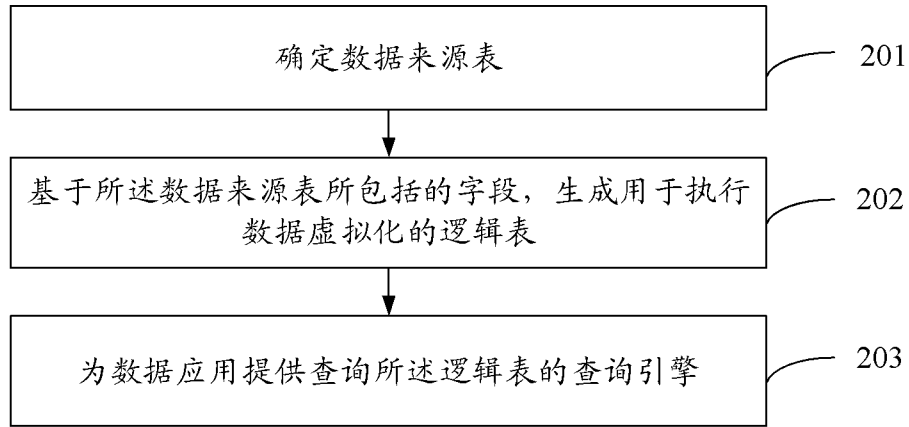


图 4

SQL

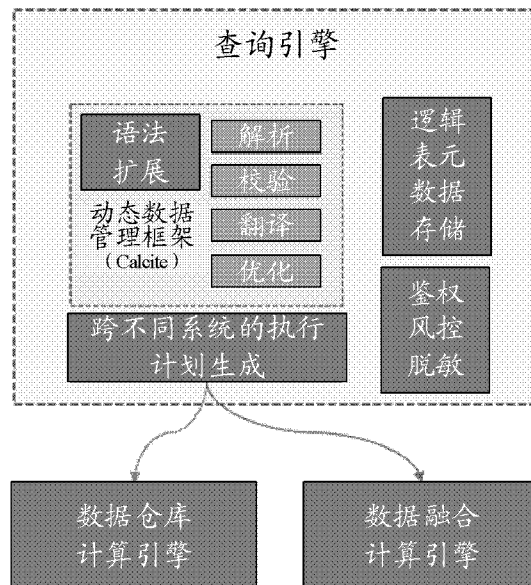


图 5

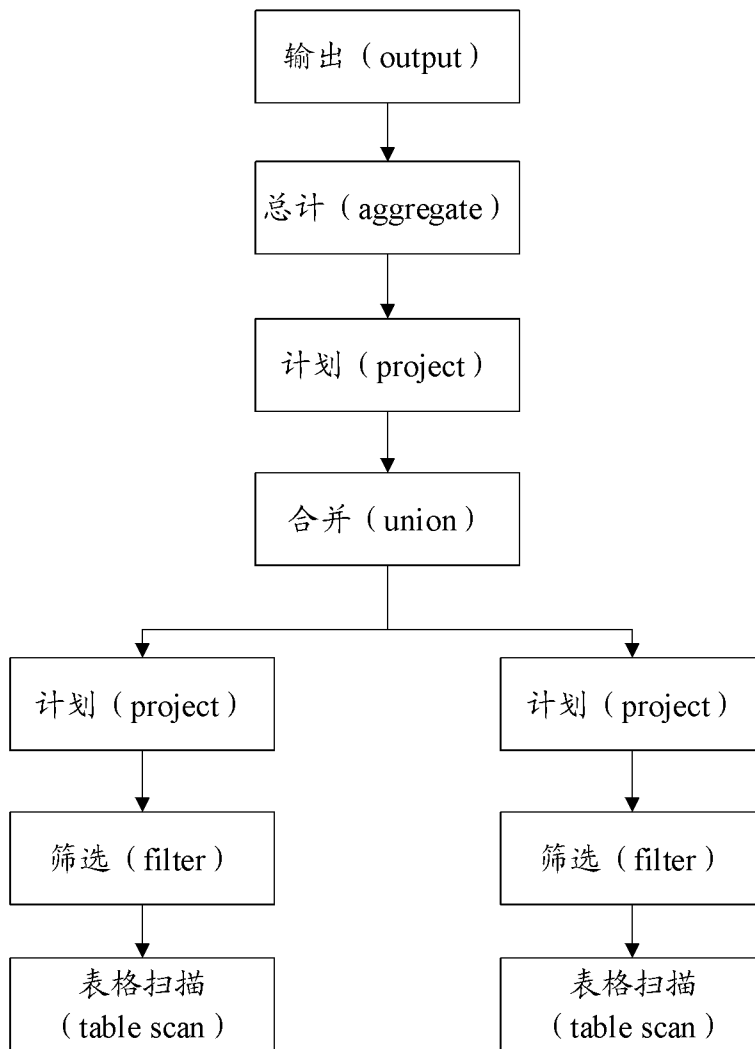


图 6A

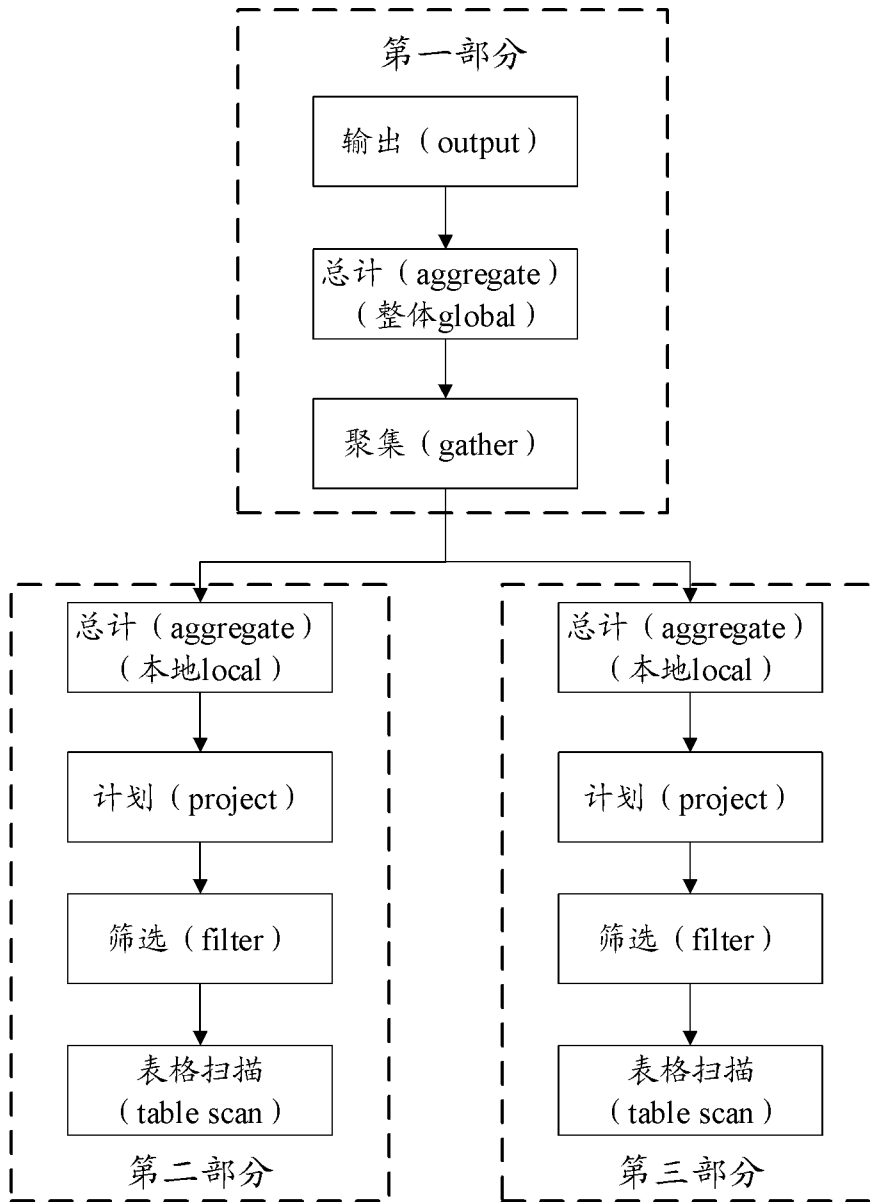


图 6B

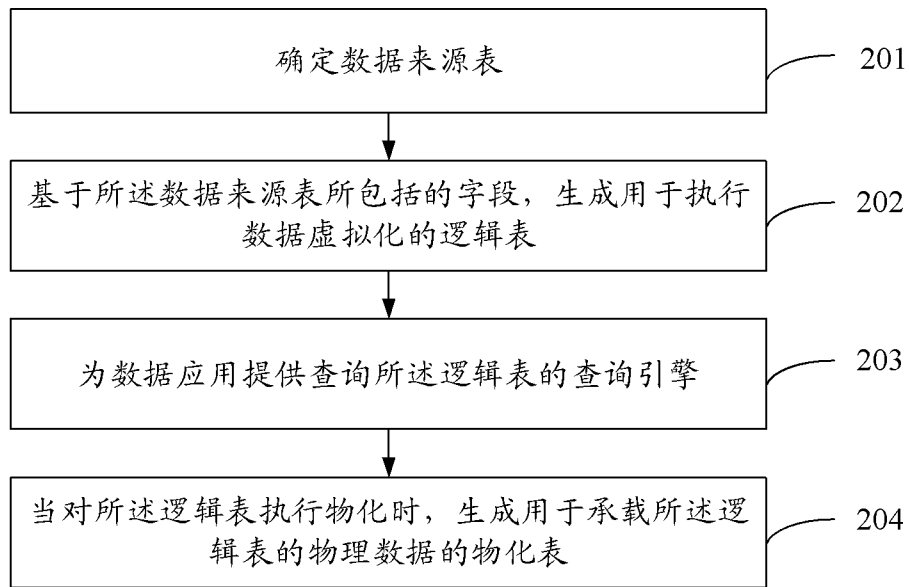


图 7

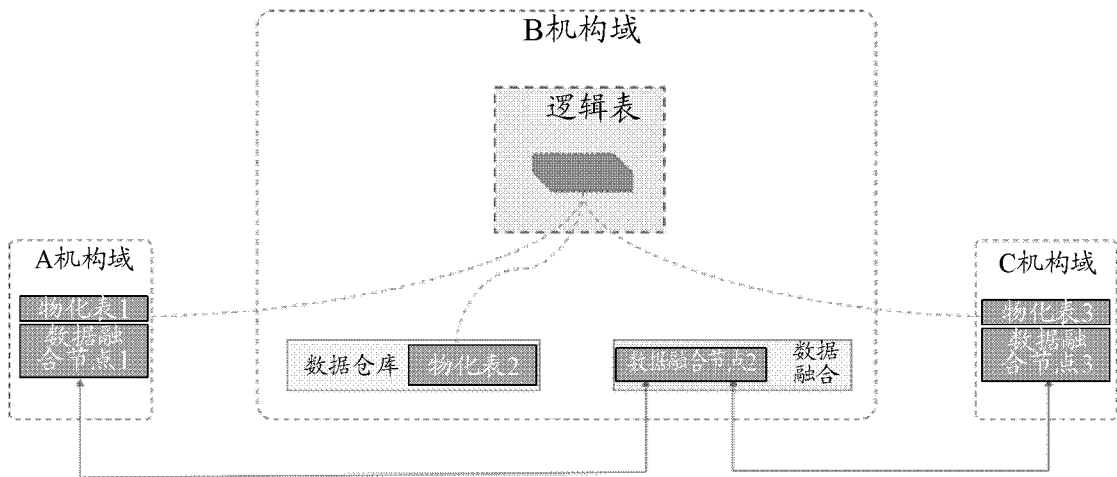


图 8

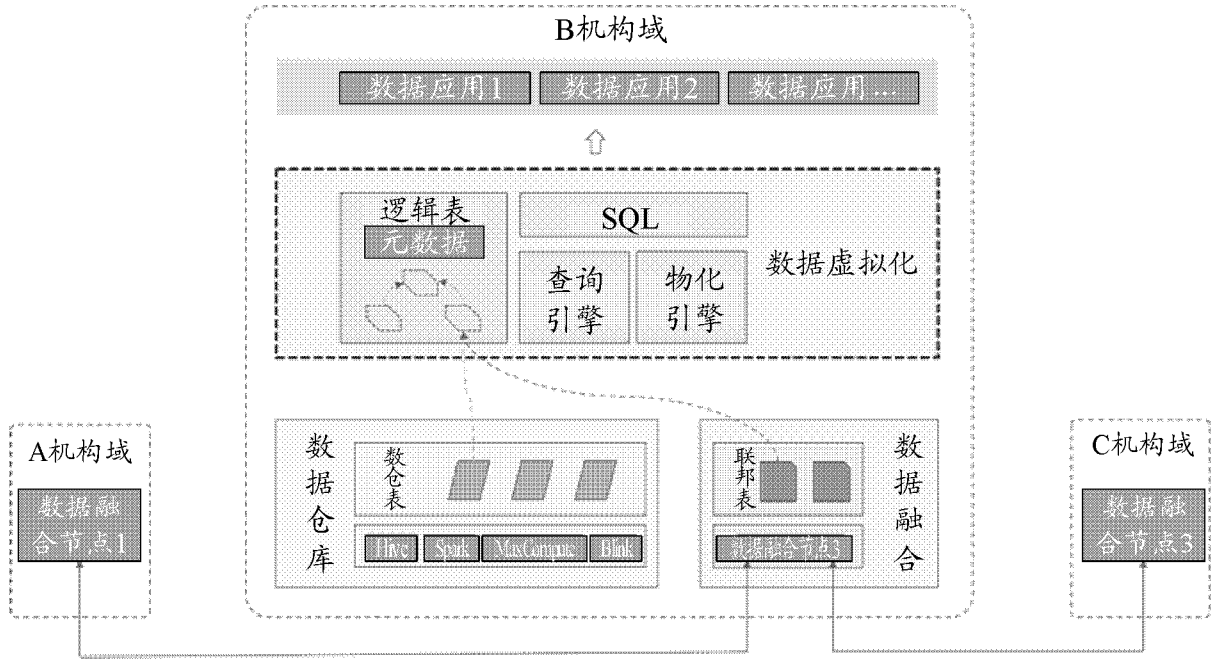


图 9

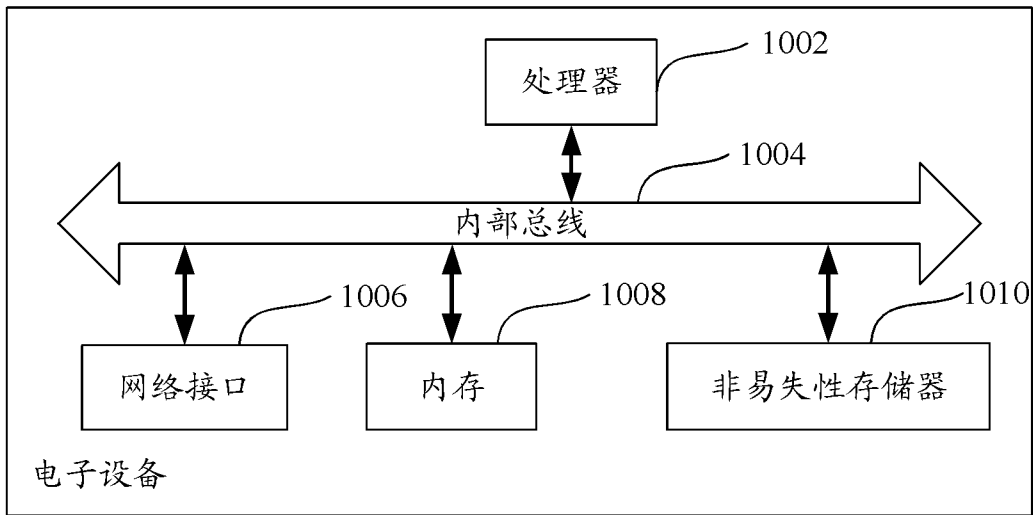


图 10

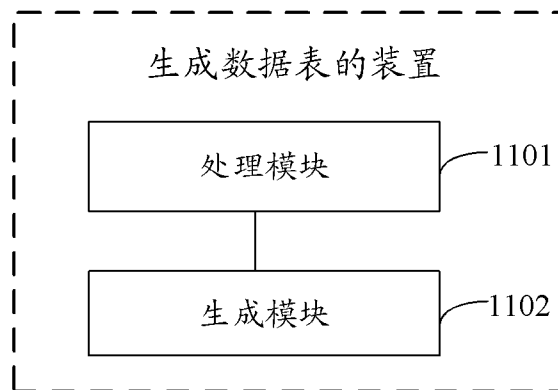


图 11

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2022/135215

A. CLASSIFICATION OF SUBJECT MATTER G06F40/18(2020.01)i According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC:G06F Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) VEN, CNABS, CNTXT, WOTXT, EPTXT, USTXT, CNKI, IEEE: 数据表, 逻辑表, 虚拟, 原始, 联邦, 融合, 关联, 字段, datasheet, logic, table, list, sheet, dummy, fictitious, original, federation, amalgamation, associate, conjunction, relating, field		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 114647716 A (TIANJIN NANKAI UNIVERSITY GENERAL DATA TECHNOLOGIES CO., LTD.) 21 June 2022 (2022-06-21) claims 1-10, and description, paragraphs [0025]-[0067]	1-14
Y	CN 112905595 A (TENCENT TECHNOLOGY (SHENZHEN) CO., LTD.) 04 June 2021 (2021-06-04) claims 1-10, and description, paragraphs [0027]-[0126]	1-14
A	CN 112966004 A (BEIJING BAIDU NETCOM SCIENCE & TECHNOLOGY CO., LTD.) 15 June 2021 (2021-06-15) entire document	1-14
A	CN 1790322 A (INTERNATIONAL BUSINESS MACHINES CORP.) 21 June 2006 (2006-06-21) entire document	1-14
A	US 2012124081 A1 (VERIZON PATENT AND LICENSING INC.) 17 May 2012 (2012-05-17) entire document	1-14
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "D" document cited by the applicant in the international application "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search 09 June 2023		Date of mailing of the international search report 19 June 2023
Name and mailing address of the ISA/CN China National Intellectual Property Administration (ISA/CN) China No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088		Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No. PCT/CN2022/135215

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	114647716	A	21 June 2022	None	
CN	112905595	A	04 June 2021	None	
CN	112966004	A	15 June 2021	None	
CN	1790322	A	21 June 2006	US 2006136469 A1	22 June 2006
				US 8112459 B2	07 February 2012
US	2012124081	A1	17 May 2012	None	

<p>A. 主题的分类 G06F40/18(2020.01) i</p> <p>按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类</p>																				
<p>B. 检索领域</p> <p>检索的最低限度文献(标明分类系统和分类号) IPC:G06F</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用)) VEN, CNABS, CNTXT, WOTXT, EPTXT, USTXT, CNKI, IEEE: 数据表, 逻辑表, 虚拟, 原始, 联邦, 融合, 关联, 字段, datasheet, logic, table, list, sheet, dummy, fictitious, original, federation, amalgamation, associate, conjunction, relating, field</p>																				
<p>C. 相关文件</p> <table border="1"> <thead> <tr> <th>类型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>Y</td> <td>CN 114647716 A (天津南大通用数据技术股份有限公司) 2022年6月21日 (2022 - 06 - 21) 权利要求1-10, 说明书第[0025]-[0067]段</td> <td>1-14</td> </tr> <tr> <td>Y</td> <td>CN 112905595 A (腾讯科技(深圳)有限公司) 2021年6月4日 (2021 - 06 - 04) 权利要求1-10, 说明书第[0027]-[0126]段</td> <td>1-14</td> </tr> <tr> <td>A</td> <td>CN 112966004 A (北京百度网讯科技有限公司) 2021年6月15日 (2021 - 06 - 15) 全文</td> <td>1-14</td> </tr> <tr> <td>A</td> <td>CN 1790322 A (国际商业机器公司) 2006年6月21日 (2006 - 06 - 21) 全文</td> <td>1-14</td> </tr> <tr> <td>A</td> <td>US 2012124081 A1 (VERIZON PATENT AND LICENSING INC.) 2012年5月17日 (2012 - 05 - 17) 全文</td> <td>1-14</td> </tr> </tbody> </table> <p><input type="checkbox"/> 其余文件在C栏的续页中列出。 <input checked="" type="checkbox"/> 见同族专利附件。</p> <p>* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “D” 申请人在国际申请中引证的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件</p>			类型*	引用文件, 必要时, 指明相关段落	相关的权利要求	Y	CN 114647716 A (天津南大通用数据技术股份有限公司) 2022年6月21日 (2022 - 06 - 21) 权利要求1-10, 说明书第[0025]-[0067]段	1-14	Y	CN 112905595 A (腾讯科技(深圳)有限公司) 2021年6月4日 (2021 - 06 - 04) 权利要求1-10, 说明书第[0027]-[0126]段	1-14	A	CN 112966004 A (北京百度网讯科技有限公司) 2021年6月15日 (2021 - 06 - 15) 全文	1-14	A	CN 1790322 A (国际商业机器公司) 2006年6月21日 (2006 - 06 - 21) 全文	1-14	A	US 2012124081 A1 (VERIZON PATENT AND LICENSING INC.) 2012年5月17日 (2012 - 05 - 17) 全文	1-14
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求																		
Y	CN 114647716 A (天津南大通用数据技术股份有限公司) 2022年6月21日 (2022 - 06 - 21) 权利要求1-10, 说明书第[0025]-[0067]段	1-14																		
Y	CN 112905595 A (腾讯科技(深圳)有限公司) 2021年6月4日 (2021 - 06 - 04) 权利要求1-10, 说明书第[0027]-[0126]段	1-14																		
A	CN 112966004 A (北京百度网讯科技有限公司) 2021年6月15日 (2021 - 06 - 15) 全文	1-14																		
A	CN 1790322 A (国际商业机器公司) 2006年6月21日 (2006 - 06 - 21) 全文	1-14																		
A	US 2012124081 A1 (VERIZON PATENT AND LICENSING INC.) 2012年5月17日 (2012 - 05 - 17) 全文	1-14																		
国际检索实际完成的日期 2023年6月9日	国际检索报告邮寄日期 2023年6月19日																			
ISA/CN的名称和邮寄地址 中国国家知识产权局 中国北京市海淀区蓟门桥西土城路6号 100088	授权官员 胡贝贝 电话号码 (+86) 010-53961415																			

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2022/135215

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	114647716	A	2022年6月21日	无			
CN	112905595	A	2021年6月4日	无			
CN	112966004	A	2021年6月15日	无			
CN	1790322	A	2006年6月21日	US	2006136469	A1	2006年6月22日
				US	8112459	B2	2012年2月7日
US	2012124081	A1	2012年5月17日	无			