

(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
2019 年 8 月 8 日 (08.08.2019)



(10) 国际公布号
WO 2019/148672 A1

(51) 国际专利分类号:
G06F 17/30 (2006.01)

(21) 国际申请号: PCT/CN2018/084322

(22) 国际申请日: 2018 年 4 月 25 日 (25.04.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201810093886.7 2018 年 1 月 31 日 (31.01.2018) CN

(71) 申请人: 平安科技(深圳)有限公司(PING AN TECHNOLOGY(SHENZHEN)CO.,LTD.) [CN/CN];
中国广东省深圳市福田区八卦岭八卦三路平安大厦六楼, Guangdong 518000 (CN)。

(72) 发明人: 邢方山(XING, Fangshan); 中国广东省深圳市福田区八卦岭八卦三路平安大厦六楼, Guangdong 518000 (CN)。

(74) 代理人: 深圳市精英专利事务所(SHENZHEN TALENT PATENT SERVICE); 中国广东省深圳市福田区深南中路 6009 号绿景广场 B 栋 20 层 B, Guangdong 518000 (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, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, 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,

(54) Title: METHOD, DEVICE, APPARATUS, AND COMPUTER READABLE STORAGE MEDIUM FOR PROCESSING DATA

(54) 发明名称: 一种数据处理方法、装置、设备及计算机可读存储介质

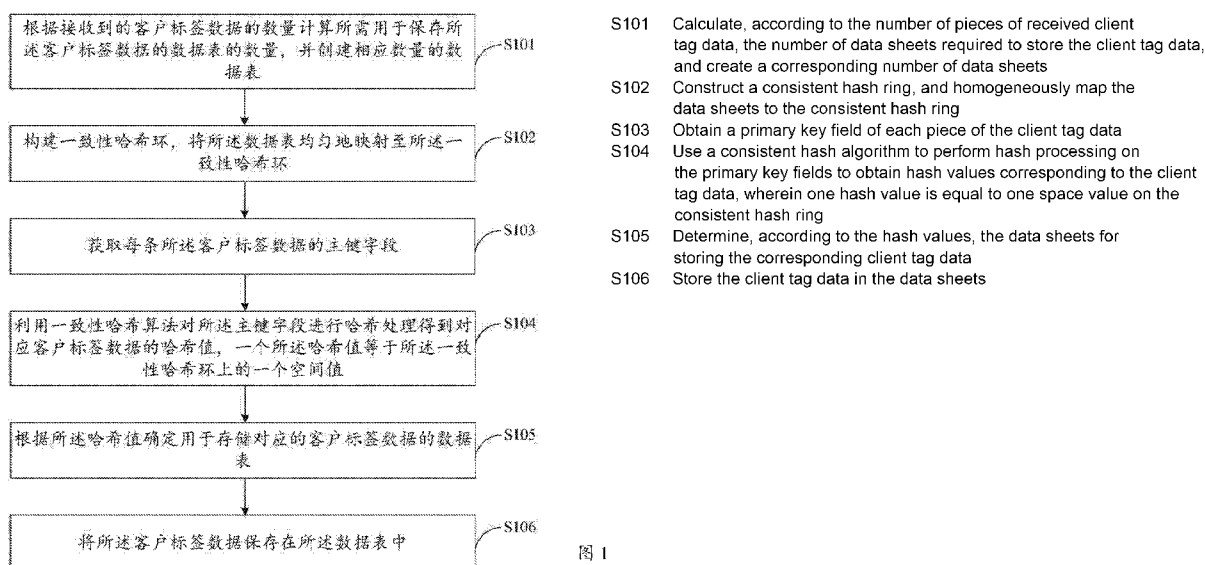


图 1

(57) Abstract: Embodiments of the present application provide a method, a device, an apparatus, and a computer readable storage medium for processing data. The method comprises: calculating, according to the number of pieces of received client tag data, the number of data sheets required to store the client tag data, and creating a corresponding number of data sheets; constructing a consistent hash ring, and homogeneously mapping the data sheets to the consistent hash ring; obtaining a primary key field of each piece of the client tag data; using a consistent hash algorithm to perform hash processing on the primary key fields to obtain hash values corresponding to the client tag data, wherein one hash value is equal to one space value on the consistent hash ring; determining, according to the hash values, the data sheets for storing the corresponding client tag data; and storing the client tag data in the data sheets. In the embodiments of the present application, hash sharding can be used to homogeneously store a large amount of client tag data in each data sheet.

SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, 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, GH, GM, KE, LR, LS, MW, MZ,
NA, RW, 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, 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))。

(57) 摘要: 本申请实施例提供一种数据处理方法、装置、设备及计算机可读存储介质。该方法包括: 根据接收到的客户标签数据的数量计算所需用于保存所述客户标签数据的数据表的数量, 并创建相应数量的数据表; 构建一致性哈希环, 将所述数据表均匀地映射至所述一致性哈希环; 获取每条所述客户标签数据的主键字段; 利用一致性哈希算法对所述主键字段进行哈希处理得到对应客户标签数据的哈希值, 一个所述哈希值等于所述一致性哈希环上的一个空间值; 根据所述哈希值确定用于存储对应的客户标签数据的数据表; 将所述客户标签数据保存在所述数据表中。本申请实施例通过哈希分表的方式可以将大量的客户标签数据均匀地存储到每个数据表中。

一种数据处理方法、装置、设备及计算机可读存储介质

本申请要求于 2018 年 01 月 31 日提交中国专利局、申请号为 201810093886.7、申请名称为“一种数据处理方法、装置、设备及计算机可读
5 存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及信息处理技术领域，尤其涉及一种数据处理方法、装置、设备及计算机可读存储介质。

10

背景技术

项目中经常遇到大批量数据（几亿至十几亿不等）需要写入数据库的情况，一般的方法是写存储过程。

目前存在一种批量存储大量数据的方法，通过一致性哈希算法将存储节点
15 映射到一致性哈希环上，根据每条数据的哈希值在一致性哈希环上按顺时针方向寻找存储该数据的存储节点，最终将数据写入到对应的存储节点。

每条数据哈希后的哈希值在一致性哈希环是具有随机分布的特性，也就是说，数据哈希后的哈希值在一致性哈希环上的分布可能会出现疏密情况，这样再按照顺时针方向寻找存储节点来保存数据的话，容易导致数据存储不均匀；
20 另外，当存储节点的数量不超过 10 时，根据一致性哈希算法将存储节点映射到一致性哈希环上，存储节点在一致性哈希环上的映射也不一定是均匀的，如此，即使数据的哈希值在一致性哈希环上是均匀分布的，最终保存数据的存储节点保存到的数据量也不一定是均匀的。

发明内容

本申请实施例提供了一种数据处理方法、装置、设备及计算机可读存储介质，能够均匀地将大批量数据存储到各个数据表中。

第一方面，本申请实施例提供了一种数据处理方法，该方法包括：

根据接收到的客户标签数据的数量计算所需用于保存所述客户标签数据的

数据表的数量，并创建相应数量的数据表；构建一致性哈希环，将所述数据表均匀地映射至所述一致性哈希环；获取每条所述客户标签数据的主键字段；利用一致性哈希算法对所述主键字段进行哈希处理得到对应客户标签数据的哈希值，一个所述哈希值等于所述一致性哈希环上的一个空间值；根据所述哈希值
5 确定用于存储对应的客户标签数据的数据表；将所述客户标签数据保存在所述数据表中。

第二方面，本申请实施例还提供了一种数据处理装置，所述数据处理装置包括用于实现第一方面所述的数据处理方法的单元。

第三方面，本申请实施例还提供了一种数据处理设备，所述数据处理设备
10 包括存储器，以及与所述存储器相连的处理器；所述存储器用于存储实现数据处理方法的计算机程序；所述处理器用于运行所述存储器中存储的计算机程序，以执行如上述第一方面所述的方法。

第四方面，本申请实施例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存储有一个或者一个以上计算机程序，所述一个或者一个以上计算机
15 机程序可被一个或者一个以上的处理器执行，以实现上述第一方面所述的方法。

本申请实施例通过哈希分表的方式可以将大量的客户标签数据均匀地存储到每个数据表中。

附图说明

20 图 1 是本申请实施例提供的一种数据处理方法的流程示意图；

图 2 是本申请实施例提供的一种数据处理方法的子流程示意图；

图 3 是本申请实施例提供的一种数据处理方法的另一子流程示意图；

图 4 是本申请实施例提供的客户标签数据和数据表在一致性哈希环的映射示意图；

25 图 5 是本申请实施例提供的一种数据处理方法的另一子流程示意图；

图 6 是本申请实施例提供的一种数据处理方法的另一流程示意图；

图 7 是本申请实施例提供的一种数据处理装置的示意性框图；

图 8 是本申请实施例提供的一种数据处理装置的子单元结构示意性框图；

图 9 是本申请实施例提供的一种数据处理装置的另一子单元结构示意性框
30 图；

图 10 是本申请实施例提供的一种数据处理装置的另一子单元结构示意性

框图；

图 11 是本申请实施例提供的一种数据处理装置的另一示意性框图；

图 12 是本申请实施例提供的一种数据处理设备的组成结构示意图。

5 具体实施方式

下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，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他实施例，都属于本申请保护的范围。

10 图 1 为本申请实施例提供的一种数据处理方法的流程示意图。该方法包括步骤 S101~S106。

S101，根据接收到的客户标签数据的数量计算所需用于保存所述客户标签数据的数据表的数量，并创建相应数量的数据表。

15 在本申请实施例中，客户标签数据来源于 Hadoop 平台，因也不需要，需
要将 Hadoop 平台中的客户标签数据需要导入到数据库中，在数据库中创建数据表来存储客户标签数据，每一个数据表存储的数据量有限，数据表的数量与客户标签数据的数量有关，例如，每个数据表可存储 6000 万条客户标签数据，则数据表的数量=客户标签数据的数量/6000 万，若有余数则向上取整为表个数。

S102，构建一致性哈希环，将所述数据表均匀地映射至所述一致性哈希环。

20 一致性哈希环为一个数字空间环，由数字 $0 \sim (2^{32})-1$ 首尾相连形成，每一个数字表示在该一致性哈希环的空间值，即空间位置。构建一致性哈希环，并建立数据表与一致性哈希环之间的映射关系，即将数据表与一致性哈希环上的空间值对应起来，使一致性哈希环上的某一个点或某一个数字代表一个数据表。

25 在本申请实施例中，如图 2 所示，步骤 S102 中的将所述数据表均匀地映射至所述一致性哈希环的步骤，包括子步骤 S1021-S1022。

S1021，根据所述数据表的数量将所述一致性哈希环进行等分，等分数量与
所述数据表的数量相等；S1022，获取所述一致性哈希环的所有等分点的空间值，
将各个空间值分别作为各个数据表映射在所述一致性哈希环的空间位置。

30 在本申请实施例中，不通过哈希算法将数据表映射至一致性哈希环上，是
为了避免当数据表数量较少时经哈希处理确定的映射关系无法使数据表在一致性哈希环上均匀分布，而直接将一致性哈希环进行等分，取等分点作为数据表

的映射，等分点的空间值即为各个数据表映射在一致性哈希环的空间位置，可以有效地保证所有数据表在一致性哈希环上均匀分布。

S103，获取每条所述客户标签数据的主键字段。

5 每条客户标签数据都有可以唯一标识该条客户标签数据的字段信息，该字段信息即为主键字段，在本申请实施例中，每一条客户标签数据包括的客户信息有客户基本信息、性格、爱好、婚姻状态、客户号、客户类型等字段信息，客户号可以唯一标识该条客户标签数据，因此客户号为该客户标签数据的主键字段。

10 S104，利用一致性哈希算法对所述主键字段进行哈希处理得到对应客户标签数据的哈希值，一个所述哈希值等于所述一致性哈希环上的一个空间值。

对每一条客户标签数据的主键字段，即客户号，利用一致性哈希算法进行哈希处理，将客户标签数据映射到该一致性哈希环上，每条客户标签数据的客户号是唯一的，因此对客户号进行哈希得到的客户标签数据的哈希值也是唯一，利用一致性哈希算法得到的哈希值的范围在一致性哈希环的空间 $0 \sim (2^{32})-1$ 范
15 围内，因此一个哈希值等于一致性哈希环上的一个空间值。

S105，根据所述哈希值确定用于存储对应的客户标签数据的数据表。

客户标签数据可以用一致性哈希环的空间值表示，数据表也用一致性哈希环的空间值表示，因此，可以通过这两个空间值的关系来确定保存客户标签数据的数据表为哪一个。

20 在本申请实施例中，如图3所示，步骤S105包括子步骤S1051a-S1053a。

S1051a，计算所述客户标签数据对应的哈希值与各个数据表的空间位置之间的相对位置比。

25 相对位置比 $gi(y) = ((hash(i) - 空间位置(y)) / 哈希环空间总值)^2$ ，其中， $gi(y)$ 表示第*i*条客户标签数据的哈希值与第*y*个数据表在一致性哈希环的空间位置的相对位置比， $hash(i)$ 表示第*i*条客户标签数据的哈希值（即该第*i*条客户标签数据映射在一致性哈希环中的空间值），空间位置（*y*）表示第*y*个数据表在哈希环中的空间位置（即该第*y*个数据表映射在一致性哈希环中的空间值），哈希环空间总值为 2^{23} 。

S1052a，获取最小的相对位置比对应的数据表。

30 S1053a，确定所述数据表为用于存储所述客户标签数据的数据表。

通过计算每条客户标签数据的哈希值与各个数据表在一致性哈希环上的空间位置的相对位置比，寻找最小相对位置比对应的数据表作为存储对应客户标

签数据的数据表。例如图 4，假设数据表为三个，分别为A1、A2 和A3，用黑点表示该三个数据表在一致性哈希环上的映射位置，白点表示客户标签数据在一致性哈希环上的映射位置，即获取白点最接近的黑点作为存储该白点对应的客户标签数据作为数据表。

5 通过一致性哈希算法将客户标签数据哈希映射在一致性哈希环，客户标签数据越多，客户标签数据在一致性哈希环上分布越均匀，本申请实施例中，数据表均匀映射在一致性哈希环上，将客户标签数据保存在距离最近的数据表中（距离最近，是针对映射在一致性哈希环的空间位置而言），这样有利于保证数据均匀的保存在每个数据列表中。

10 然而，即使客户标签数据在一致性哈希环上分布是接近均匀的，但是当数据量不足够大时，映射不可能完全均匀，那么会存在这样一种情况，在数据分布相对稀疏的地方，数据表存储的数据较少，而在数据分布相比密集的地方，数据表存储的数据则较多，造成每个数据存储的数据量差异大，因此，可以设置每个数据表存储的数据量具有一个预设值，当距离最近的数据表存储数据量
15 满的时候，存在另一次近而存储量未满足的数据表中，这样可以保证每个数据表最终存储的数据量基本是一样的。具体地，在其他实施例中，如图 5 所示，步骤S105 包括子步骤S1051b-S1055b。

 S1051b，计算所述客户标签数据对应的哈希值与各个数据表的空间位置之间的相对位置比。相对位置比的解释同上。

20 S1052b，获取最小的相对位置比对应的数据表。

 S1053b，判断所述数据表的存储数据量是否达到预设数量。

 S1054b，若否，则确定所述数据表为用于存储所述客户标签数据的数据表。

 S1055b，若是，则从存储数据量未达到预设数量的数据表中获取具有较小的相对位置比对应的数据表，确定所述数据表为用于存储所述客户标签数据的
25 数据表。

 通过以上子步骤可以实现每个数据表均匀地存储客户标签数据。

 S106，将所述客户标签数据保存在所述数据表中。

 在创建数据表后系统会给每个数据表命名，数据表的空间位置和表名是一一对应的，将空间位置和表名保存在treeMap中，则确定保存具体客户标签数据的
30 空间位置后，通过空间位置对应的表名几个查询到该数据表，从而将客户标

签数据保存在数据表中。

本申请实施例的数据处理方法，数据表是均匀映射在一致性哈希环中，根据数据与数据表的位置关系将客户标签数据保存在距离最近的数据表中，这样有利于保证数据均匀的保存在每个数据列表中；根据数据与数据表的位置关系，以及数据表的存储量关系，当数据表存储满数据之后，将所述存储到较近位置的数据表，有利于保证每个数据表存储的数据量基本是一致的。

图6是本申请实施例提供的一种数据处理方法的另一流程示意图，该方法包括步骤S201-S211。

S201，根据接收到的客户标签数据的数量计算所需用于保存所述客户标签数据的数据表的数量，并创建相应数量的数据表，利用序号为所述数据表命名。

在本申请实施例中，数据表的命名结合和序号，例如命名数据表的表名为表1、表2、表3，……，或A1、A2、A3，……，等。

S202，构建一致性哈希环，将所述数据表均匀地映射至所述一致性哈希环。S203，获取每条所述客户标签数据的主键字段。S204，利用一致性哈希算法对所述主键字段进行哈希处理得到对应客户标签数据的哈希值，一个所述哈希值等于所述一致性哈希环上的一个空间值。S205，根据所述哈希值确定用于存储对应的客户标签数据的数据表。S206，将所述客户标签数据保存在所述数据表中。

本申请实施例的步骤S202-S206与上述实施例的步骤S102-S106相同，具体描述可参考上述实施例。

S207，接收查询客户标签数据的指令信息，所述指令信息包括客户标签数据的主键字段。保存客户标签数据主要是为了后期的数据查询需要，用户可通过主键字段，即客户号来查询某个客户的客户标签数据。

S208，通过所述一致性哈希算法对所述主键字段进行哈希处理得到所述客户标签数据的哈希值。S209，从所有所述数据表的空间位置中搜索与所述哈希值最接近的空间位置。查询数据时对客户号进行的哈希处理与保存数据时进行的哈希处理的方法相同，两种情形得到的哈希值是一致的。

S210，获取与所述哈希值最接近的空间位置对应的数据表，在所述数据表中查询所述主键字段对应的客户标签数据。S211，若在所述数据表中未查询到所述客户标签数据，则根据所述数据表的表名中的序号依次查询剩余的数据表，直至查询到所述客户标签数据。因为客户标签数据是优先保存在与对应的哈希

值最接近的空间位置对应的数据表中，因此优先查询该数据表，当该数据表中不存在需要查询的客户标签数据时，以数据表的序号为标准从下一个数据表开始一个一个查询数据表，直至查询到该客户标签数据。例如，有五个数据表 A1、A2、A3、A4 和 A5，优先确定的数据表为 A3，但在 A3 中未查到对应的客户标签数据，因此，依次查询 A4、A5、A1 和 A2，直至查询到对应的客户标签数据。

本申请实施例的数据处理方法，对应哈希分表保存数据的方法来查询数据，防止在一个数据表中查询不到客户标签数据后而产生查询错误的情况。

图 7 为本申请实施例提供的一种数据处理装置 100 的示意性框图。该数据处理装置 100 包括计算单元 101、创建单元 102、构建单元 103、映射单元 104、获取单元 105、哈希单元 106、确定单元 107 和保存单元 108。

计算单元 101 用于根据接收到的客户标签数据的数量计算所需用于保存所述客户标签数据的数据表的数量；创建单元 102 用于创建相应数量的数据表；构建单元 103 用于构建一致性哈希环；映射单元 104 用于将所述数据表均匀地映射至所述一致性哈希环；第一获取单元 105 用于获取每条所述客户标签数据的主键字段；哈希单元 106 用于利用一致性哈希算法对所述主键字段进行哈希处理得到对应客户标签数据的哈希值，一个所述哈希值等于所述一致性哈希环上的一个空间值；确定单元 107 用于根据所述哈希值确定用于存储对应的客户标签数据的数据表；保存单元 108 用于将所述客户标签数据保存在所述数据表中。

在本申请实施例中，如图 8 所示，所述映射单元 104 包括：

等分子单元 1041，用于根据所述数据表的数量将所述一致性哈希环进行等分，等分数量与所述数据表的数量相等。

映射子单元 1042，用于获取所述一致性哈希环的所有等分点的空间值，将各个空间值分别作为各个数据表映射在所述一致性哈希环的空间位置。

在本申请实施例中，如图 9 所示，所述确定单元 107 包括：

计算子单元 1071，用于计算所述客户标签数据对应的哈希值与各个数据表的空间位置之间的相对位置比。

获取子单元 1072，用于获取最小的相对位置比对应的数据表。

第一确定子单元 1073，用于确定所述数据表为用于存储所述客户标签数据的数据表。

在其他申请实施例中，如图 10 所示，所述确定单元 107 包括：

计算子单元 1071，用于计算所述客户标签数据对应的哈希值与各个数据表的空间位置之间的相对位置比；

获取子单元 1072，用于获取最小的相对位置比对应的数据表。

5 判断子单元 1074，用于判断所述数据表的存储数据量是否达到预设数量。

第二确定子单元 1075，用于若所述数据表的存储数据量未达到预设数量，则确定所述数据表为用于存储所述客户标签数据的数据表，若所述数据表的存储数据量达到预设数量，则从存储数据量未达到预设数量的数据表中获取具有较小的相对位置比对应的数据表，确定所述数据表为用于存储所述客户标签数
10 据的数据表。

如图 11 所示，所述数据处理装置 100 还包括命名单元 109、接收单元 110、搜索单元 111、第二获取单元 112 和查询单元 113。

命名单元 109 用于在创建单元 102 创建相应数量的数据表之后利用序号为所述数据表命名；接收单元 110 用于接收查询客户标签数据的指令信息，所述
15 指令信息包括客户标签数据的主键字段。

所述哈希单元 106 还用于：通过所述一致性哈希算法对所述指令信息中的主键字段进行哈希处理得到所述客户标签数据的哈希值。

搜索单元 111 用于从所有所述数据表的空间位置中搜索与所述哈希值最接近的空间位置；第二获取单元 112 用于获取与所述哈希值最接近的空间位置对
20 应的数据表；查询单元 113 用于在所述数据表中查询所述主键字段对应的客户标签数据，若在所述数据表中未查询到所述客户标签数据，则根据所述数据表的表名中的序号依次查询剩余的数据表，直至查询到所述客户标签数据。

数据处理装置 100 各个单元的具体描述可参考上述方法实施例，本申请实施例不做赘述。上述数据处理装置 100 可以实现为一种计算机程序的形式，计
25 算机程序可以在如图 12 所示的计算机设备上运行。

图 12 为本申请实施例提供的一种数据处理设备的组成结构示意图。该设备可以是终端，也可以是服务器，其中，终端可以是智能手机、平板电脑、笔记本电脑、台式电脑、个人数字助理和穿戴式设备等具有通信功能的电子设备。服务器可以是独立的服务器，也可以是多个服务器组成的服务器集群。

30 该设备为一种计算机设备 200，包括通过系统总线 201 连接的处理器 202、存储器和网络接口 205，其中，存储器可以包括非易失性存储介质 203 和内存存储器 204。该计算机设备 200 的非易失性存储介质 203 可存储操作系统 2031 和

5 计算机程序 2032，该计算机程序 2032 被执行时，可使得处理器 202 执行一种数据处理方法。该计算机设备 200 的处理器 202 用于提供计算和控制能力，支撑整个计算机设备 200 的运行。该内存储器 204 为非易失性存储介质 203 中的计算机程序 2032 的运行提供环境。计算机设备 200 的网络接口 205 用于进行网络通信，如发送分配的任务、接收数据等。处理器 202 运行非易失性存储介质 203 中的计算机程序 2032 时可实现上述数据处理方法的所有实施例。

10 本领域技术人员可以理解，图 12 中示出的数据处理设备的实施例并不构成对数据处理设备具体构成的限定，在其他实施例中，数据处理设备可以包括比图示更多或更少的部件，或者组合某些部件，或者不同的部件布置。例如，在一些实施例中，数据处理设备可以仅包括存储器及处理器，在这样的实施例中，存储器及处理器的结构及功能与图 12 所示实施例一致，在此不再赘述。

15 本申请还提供一种计算机可读存储介质，所述计算机可读存储介质存储有一个或者一个以上计算机程序，所述一个或者一个以上计算机程序可被一个或者一个以上的处理器执行，当所述一个或者一个以上程序被一个或者一个以上的处理器执行，可实现上述数据处理方法的所有实施例。

20 以上所述，仅为本申请的具体实施方式，但本申请的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本申请揭露的技术范围内，可轻易想到各种等效的修改或替换，这些修改或替换都应涵盖在本申请的保护范围之内。因此，本申请的保护范围应以权利要求的保护范围为准。

权 利 要 求 书

1.一种数据处理方法，其特征在于，包括：

根据接收到的客户标签数据的数量计算所需用于保存所述客户标签数据的数据表的数量，并创建相应数量的数据表；

构建一致性哈希环，将所述数据表均匀地映射至所述一致性哈希环；

获取每条所述客户标签数据的主键字段；

利用一致性哈希算法对所述主键字段进行哈希处理得到对应客户标签数据的哈希值，一个所述哈希值等于所述一致性哈希环上的一个空间值；

根据所述哈希值确定用于存储对应的客户标签数据的数据表；

将所述客户标签数据保存在所述数据表中。

2.根据权利要求 1 所述的数据处理方法，其特征在于，所述将所述数据表均匀地映射至所述一致性哈希环，包括：

根据所述数据表的数量将所述一致性哈希环进行等分，等分数量与所述数据表的数量相等；

获取所述一致性哈希环的所有等分点的空间值，将各个空间值分别作为各个数据表映射在所述一致性哈希环的空间位置。

3.根据权利要求 2 所述的数据处理方法，其特征在于，所述根据所述哈希值确定用于存储对应的客户标签数据的数据表，包括：

计算所述客户标签数据对应的哈希值与各个数据表的空间位置之间的相对位置比；

获取最小的相对位置比对应的数据表；

确定所述数据表为用于存储所述客户标签数据的数据表。

4.根据权利要求 2 所述的数据处理方法，其特征在于，所述根据所述哈希值确定用于存储对应的客户标签数据的数据表，包括：

计算所述客户标签数据对应的哈希值与各个数据表的空间位置之间的相对位置比；

获取最小的相对位置比对应的数据表；

判断所述数据表的存储数据量是否达到预设数量；

若所述数据表的存储数据量未达到预设数量，则确定所述数据表为用于存储所述客户标签数据的数据表；

若所述数据表的存储数据量达到预设数量，则从存储数据量未达到预设数

量的数据表中获取具有较小的相对位置比对应的数据表，确定所述数据表为用于存储所述客户标签数据的数据表。

5.根据权利要求 1 所述的数据处理方法，其特征在于，所述创建相应数量的数据表之后，还包括：

5 利用序号为所述数据表命名；

所述将所述客户标签数据保存在所述数据表中之后，还包括：

接收查询客户标签数据的指令信息，所述指令信息包括客户标签数据的主键字段；

10 通过所述一致性哈希算法对所述主键字段进行哈希处理得到所述客户标签数据的哈希值；

从所有所述数据表的空间位置中搜索与所述哈希值最接近的空间位置；

获取与所述哈希值最接近的空间位置对应的数据表，在所述数据表中查询所述主键字段对应的客户标签数据；

15 若在所述数据表中未查询到所述客户标签数据，则根据所述数据表的表名中的序号依次查询剩余的数据表，直至查询到所述客户标签数据。

6.一种数据处理装置，其特征在于，包括：

计算单元，用于根据接收到的客户标签数据的数量计算所需用于保存所述客户标签数据的数据表的数量；

创建单元，用于创建相应数量的数据表；

20 构建单元，用于构建一致性哈希环；

映射单元，用于将所述数据表均匀地映射至所述一致性哈希环；

获取单元，用于获取每条所述客户标签数据的主键字段；

25 哈希单元，用于利用一致性哈希算法对所述主键字段进行哈希处理得到对应客户标签数据的哈希值，一个所述哈希值等于所述一致性哈希环上的一个空间值；

确定单元，用于根据所述哈希值确定用于存储对应的客户标签数据的数据表；

保存单元，用于将所述客户标签数据保存在所述数据表中。

7.根据权利要求 6 所述的数据处理装置，其特征在于，所述映射单元包括：

30 等分子单元，用于根据所述数据表的数量将所述一致性哈希环进行等分，等分数量与所述数据表的数量相等；

映射子单元，用于获取所述一致性哈希环的所有等分点的空间值，将各个

空间值分别作为各个数据表映射在所述一致性哈希环的空间位置。

8.根据权利要求7所述的数据处理装置，其特征在于，所述确定单元包括：

计算子单元，用于计算所述客户标签数据对应的哈希值与各个数据表的空间位置之间的相对位置比；

5 获取子单元，用于获取最小的相对位置比对应的数据表；

确定子单元，用于确定所述数据表为用于存储所述客户标签数据的数据表。

9.根据权利要求7所述的数据处理装置，其特征在于，所述确定单元包括：

计算子单元，用于计算所述客户标签数据对应的哈希值与各个数据表的空间位置之间的相对位置比；

10 获取子单元，用于获取最小的相对位置比对应的数据表；

判断子单元，用于判断所述数据表的存储数据量是否达到预设数量；

第二确定子单元，用于若所述数据表的存储数据量未达到预设数量，则确定所述数据表为用于存储所述客户标签数据的数据表，若所述数据表的存储数据量达到预设数量，则从存储数据量未达到预设数量的数据表中获取具有较小
15 的相对位置比对应的数据表，确定所述数据表为用于存储所述客户标签数据的数据表。

10.根据权利要求6所述的数据处理装置，其特征在于，还包括命名单元、接收单元、搜索单元、第二获取单元和查询单元；

所述命名单元用于在所述创建单元创建相应数量的数据表之后利用序号为
20 所述数据表命名；

所述接收单元用于接收查询客户标签数据的指令信息，所述指令信息包括客户标签数据的主键字段；

所述哈希单元还用于：通过所述一致性哈希算法对所述指令信息中的主键字段进行哈希处理得到所述客户标签数据的哈希值；

25 所述搜索单元用于从所有所述数据表的空间位置中搜索与所述哈希值最接近的空间位置；

所述第二获取单元用于获取与所述哈希值最接近的空间位置对应的数据表；

所述查询单元用于在所述数据表中查询所述主键字段对应的客户标签数据，
30 若在所述数据表中未查询到所述客户标签数据，则根据所述数据表的表名中的序号依次查询剩余的数据表，直至查询到所述客户标签数据。

11.一种数据处理设备，其特征在于，所述数据处理设备包括存储器，以及

与所述存储器相连的处理器；

所述存储器，用于存储实现数据处理方法的计算机程序；

所述处理器，用于运行所述存储器中存储的计算机程序，以执行以下步骤：

根据接收到的客户标签数据的数量计算所需用于保存所述客户标签数据的

5 数据表的数量，并创建相应数量的数据表；

构建一致性哈希环，将所述数据表均匀地映射至所述一致性哈希环；

获取每条所述客户标签数据的主键字段；

利用一致性哈希算法对所述主键字段进行哈希处理得到对应客户标签数据的哈希值，一个所述哈希值等于所述一致性哈希环上的一个空间值；

10 根据所述哈希值确定用于存储对应的客户标签数据的数据表；

将所述客户标签数据保存在所述数据表中。

12.根据权利要求 11 所述的数据处理设备，其特征在于，所述处理器在执行所述将所述数据表均匀地映射至所述一致性哈希环的步骤时，具体执行以下步骤：

15 根据所述数据表的数量将所述一致性哈希环进行等分，等分数量与所述数据表的数量相等；

获取所述一致性哈希环的所有等分点的空间值，将各个空间值分别作为各个数据表映射在所述一致性哈希环的空间位置。

13.根据权利要求 12 所述的数据处理设备，其特征在于，所述处理器在执行所述根据所述哈希值确定用于存储对应的客户标签数据的数据表的步骤时，具体执行以下步骤：

计算所述客户标签数据对应的哈希值与各个数据表的空间位置之间的相对位置比；

获取最小的相对位置比对应的数据表；

25 确定所述数据表为用于存储所述客户标签数据的数据表。

14.根据权利要求 12 所述的数据处理设备，其特征在于，所述处理器在执行所述根据所述哈希值确定用于存储对应的客户标签数据的数据表的步骤时，具体执行以下步骤：

30 计算所述客户标签数据对应的哈希值与各个数据表的空间位置之间的相对位置比；

获取最小的相对位置比对应的数据表；

判断所述数据表的存储数据量是否达到预设数量；

若所述数据表的存储数据量未达到预设数量，则确定所述数据表为用于存储所述客户标签数据的数据表；

若所述数据表的存储数据量达到预设数量，则从存储数据量未达到预设数量的数据表中获取具有较小的相对位置比对应的数据表，确定所述数据表为用于存储所述客户标签数据的数据表。

15.根据权利要求 11 所述的数据处理设备，其特征在于，所述处理器在执行所述创建相应数量的数据表的步骤之后，还执行以下步骤：

利用序号为所述数据表命名；

所述处理器在执行所述将所述客户标签数据保存在所述数据表中的步骤之后，还执行以下步骤：

接收查询客户标签数据的指令信息，所述指令信息包括客户标签数据的主键字段；

通过所述一致性哈希算法对所述主键字段进行哈希处理得到所述客户标签数据的哈希值；

15 从所有所述数据表的空间位置中搜索与所述哈希值最接近的空间位置；

获取与所述哈希值最接近的空间位置对应的数据表，在所述数据表中查询所述主键字段对应的客户标签数据；

若在所述数据表中未查询到所述客户标签数据，则根据所述数据表的表名中的序号依次查询剩余的数据表，直至查询到所述客户标签数据。

20 16.一种计算机可读存储介质，其特征在于，所述计算机可读存储介质存储有一个或者一个以上计算机程序，所述一个或者一个以上计算机程序可被一个或者一个以上的处理器执行，以实现以下步骤：

根据接收到的客户标签数据的数量计算所需用于保存所述客户标签数据的数据表的数量，并创建相应数量的数据表；

25 构建一致性哈希环，将所述数据表均匀地映射至所述一致性哈希环；

获取每条所述客户标签数据的主键字段；

利用一致性哈希算法对所述主键字段进行哈希处理得到对应客户标签数据的哈希值，一个所述哈希值等于所述一致性哈希环上的一个空间值；

根据所述哈希值确定用于存储对应的客户标签数据的数据表；

30 将所述客户标签数据保存在所述数据表中。

17.根据权利要求 16 所述的计算机可读存储介质，其特征在于，所述将所述数据表均匀地映射至所述一致性哈希环的步骤包括：

根据所述数据表的数量将所述一致性哈希环进行等分，等分数量与所述数据表的数量相等；

获取所述一致性哈希环的所有等分点的空间值，将各个空间值分别作为各个数据表映射在所述一致性哈希环的空间位置。

5 18.根据权利要求 17 所述的计算机可读存储介质，其特征在于，所述根据所述哈希值确定用于存储对应的客户标签数据的数据表的步骤包括：

计算所述客户标签数据对应的哈希值与各个数据表的空间位置之间的相对位置比；

获取最小的相对位置比对应的数据表；

10 确定所述数据表为用于存储所述客户标签数据的数据表。

19.根据权利要求 17 所述的计算机可读存储介质，其特征在于，所述根据所述哈希值确定用于存储对应的客户标签数据的数据表的步骤包括：

计算所述客户标签数据对应的哈希值与各个数据表的空间位置之间的相对位置比；

15 获取最小的相对位置比对应的数据表；

判断所述数据表的存储数据量是否达到预设数量；

若所述数据表的存储数据量未达到预设数量，则确定所述数据表为用于存储所述客户标签数据的数据表；

20 若所述数据表的存储数据量达到预设数量，则从存储数据量未达到预设数量的数据表中获取具有较小的相对位置比对应的数据表，确定所述数据表为用于存储所述客户标签数据的数据表。

20.根据权利要求 16 所述的计算机可读存储介质，其特征在于，在实现所述创建相应数量的数据表的步骤之后，还实现以下步骤：

利用序号为所述数据表命名；

25 在实现所述将所述客户标签数据保存在所述数据表中的步骤之后，还实现以下步骤：

接收查询客户标签数据的指令信息，所述指令信息包括客户标签数据的主键字段；

30 通过所述一致性哈希算法对所述主键字段进行哈希处理得到所述客户标签数据的哈希值；

从所有所述数据表的空间位置中搜索与所述哈希值最接近的空间位置；

获取与所述哈希值最接近的空间位置对应的数据表，在所述数据表中查询

所述主键字段对应的客户标签数据；

若在所述数据表中未查询到所述客户标签数据，则根据所述数据表的表名中的序号依次查询剩余的数据表，直至查询到所述客户标签数据。

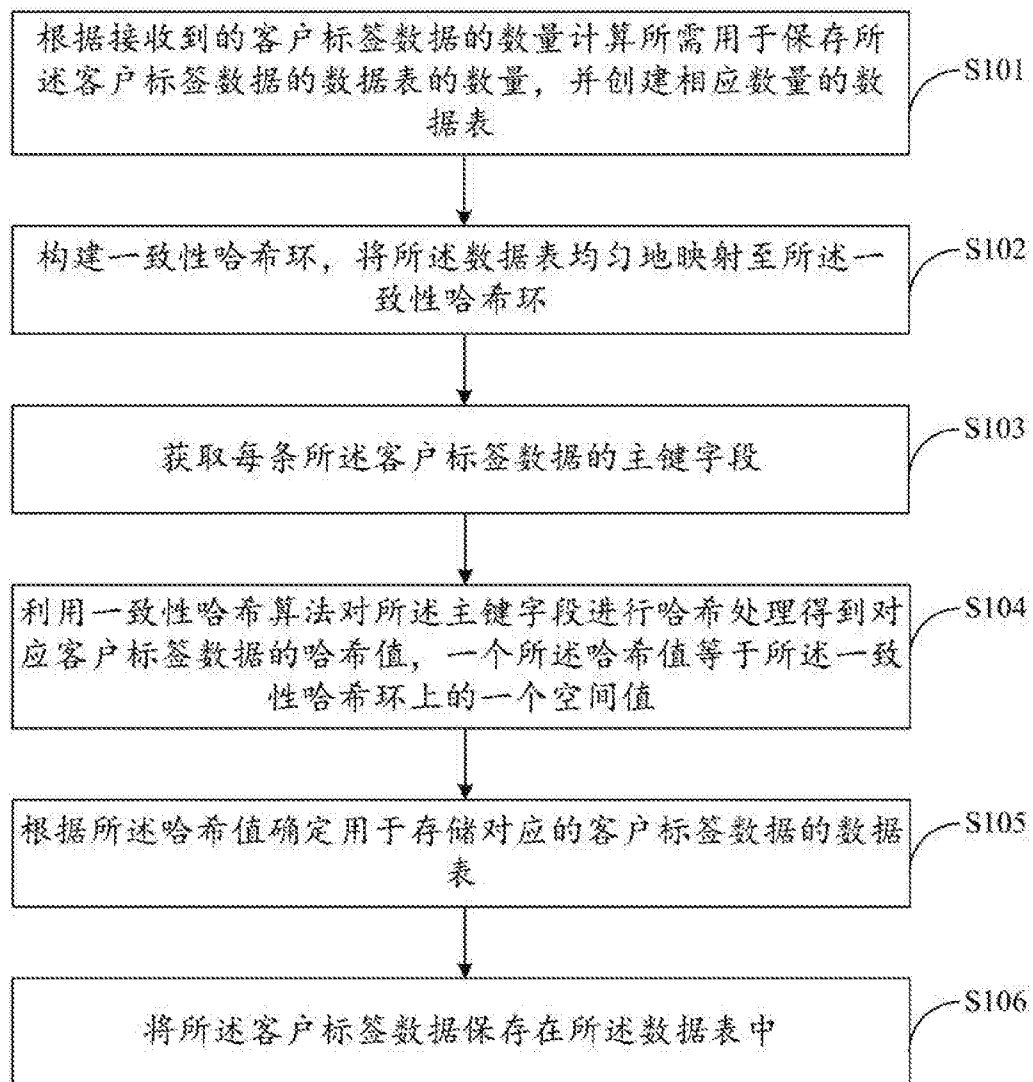


图 1

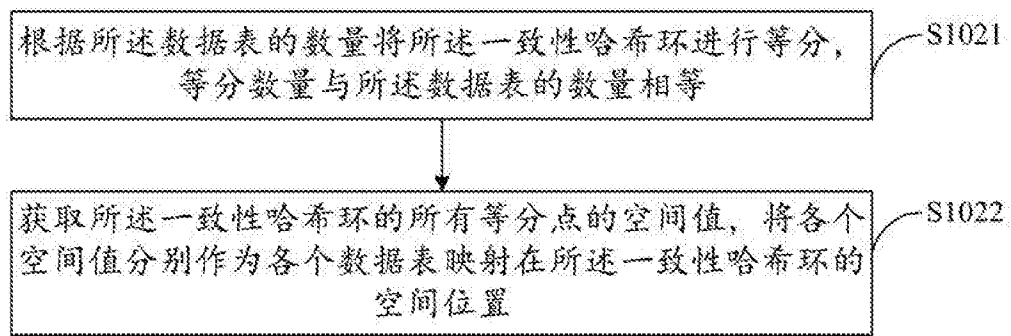


图 2

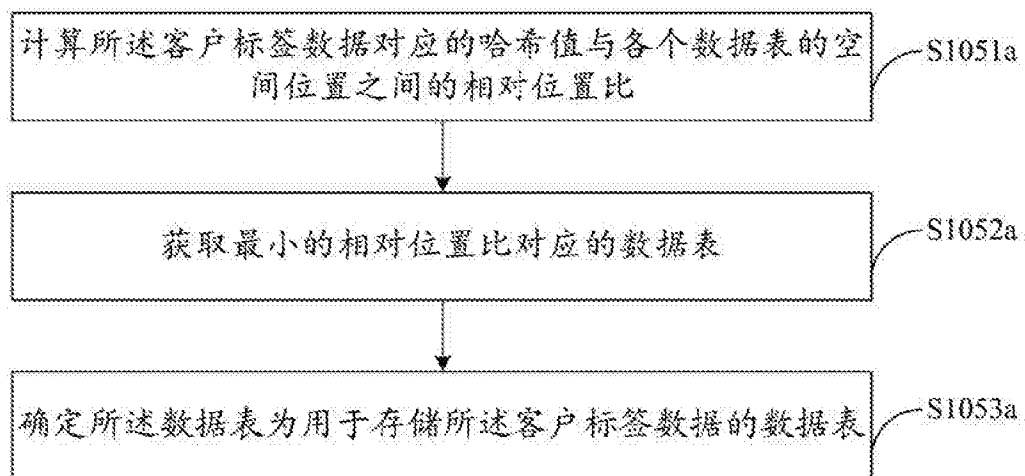


图 3

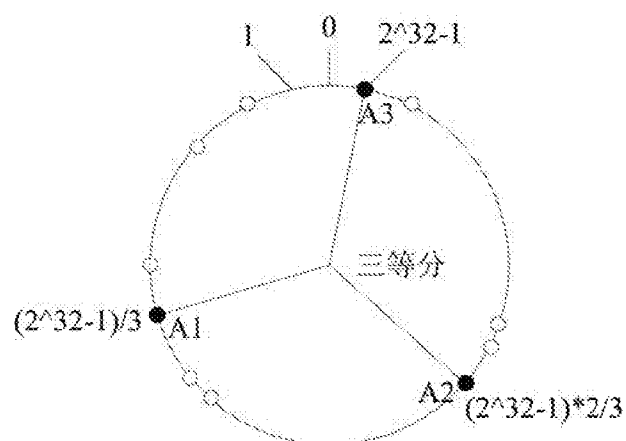


图 4

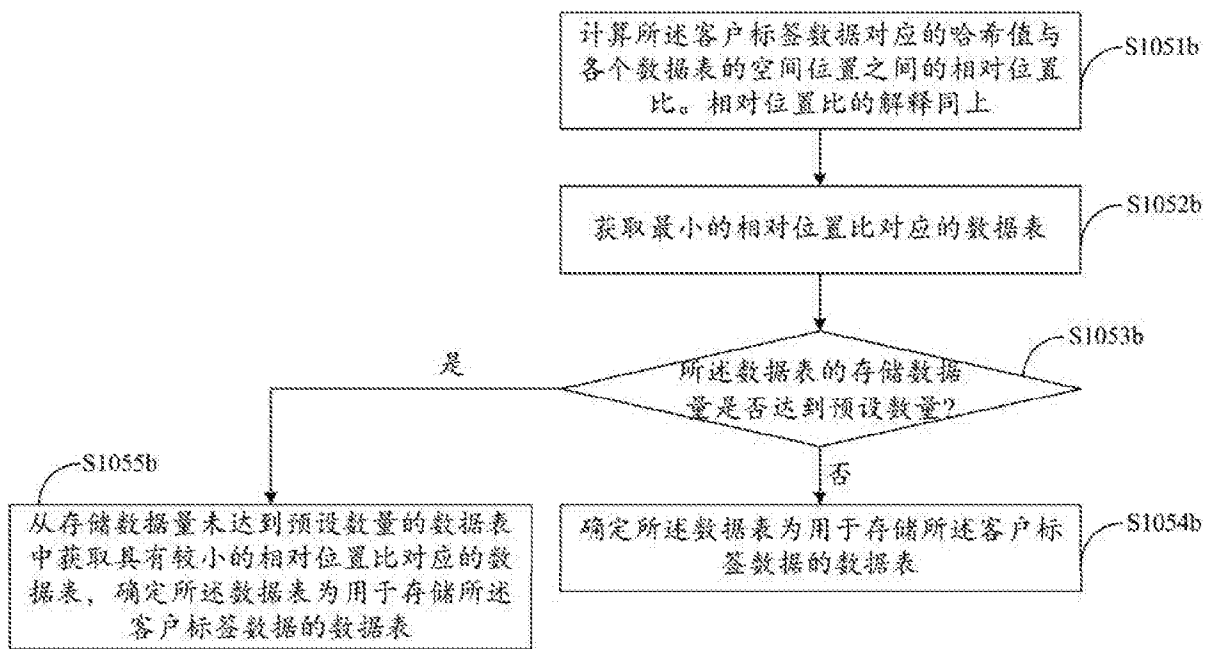


图 5

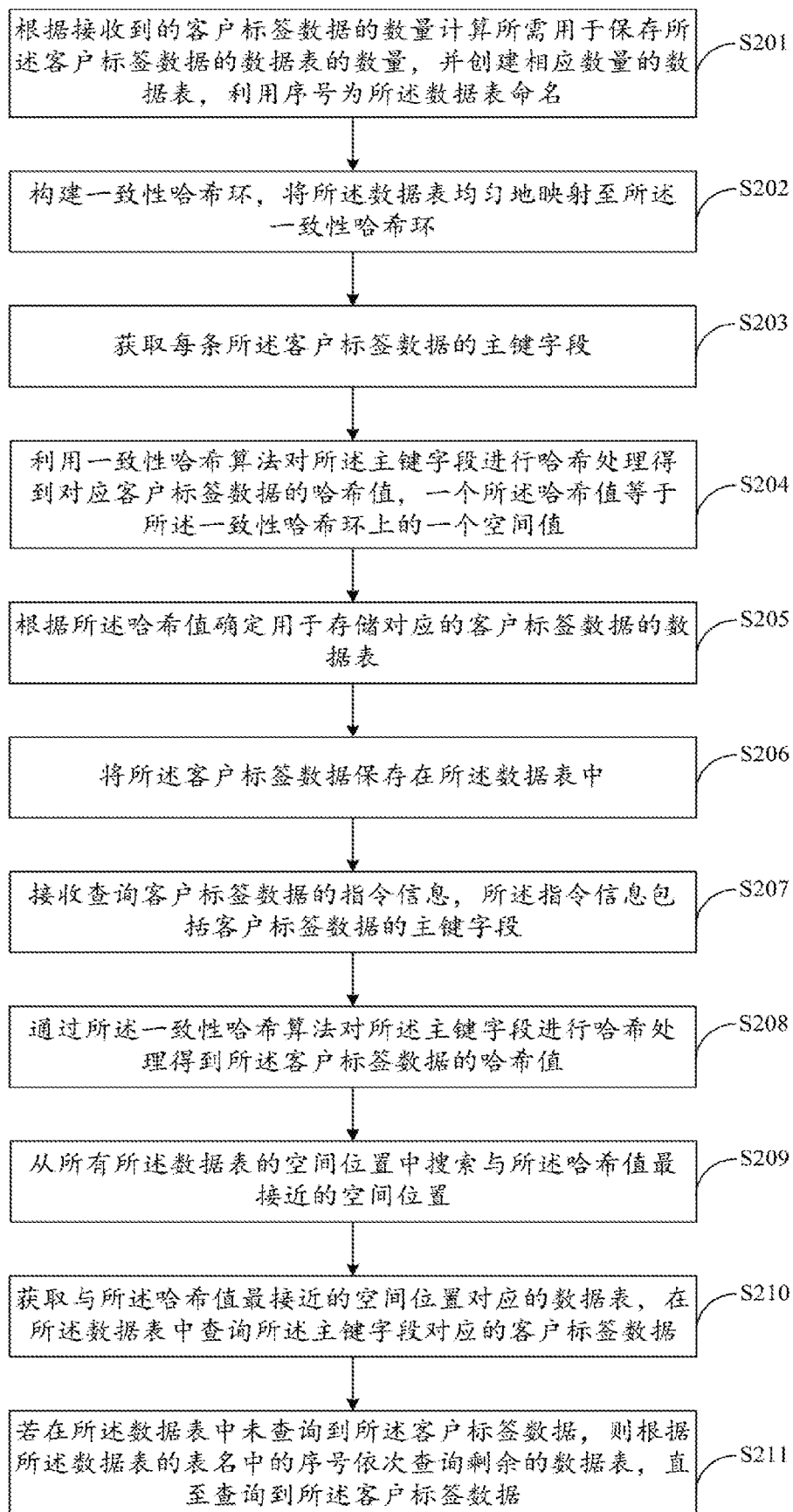


图 6

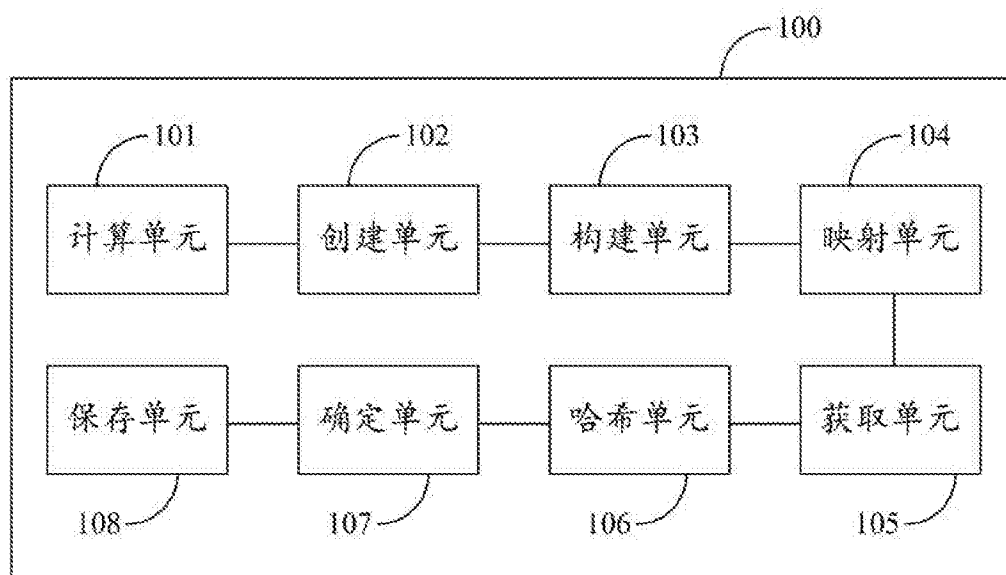


图 7

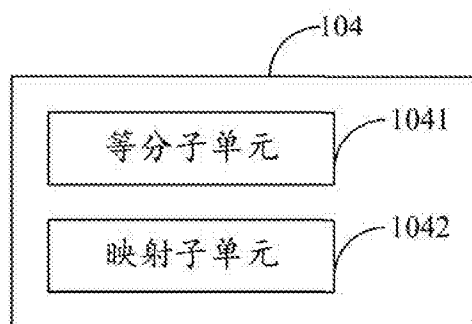


图 8

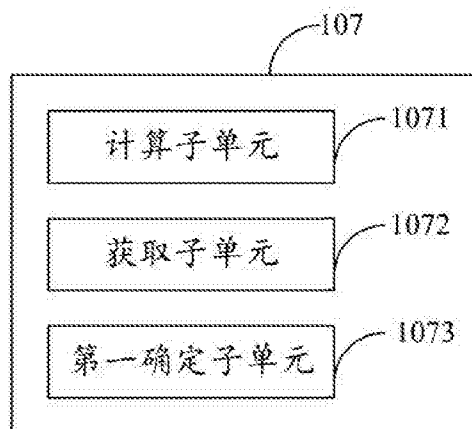


图 9

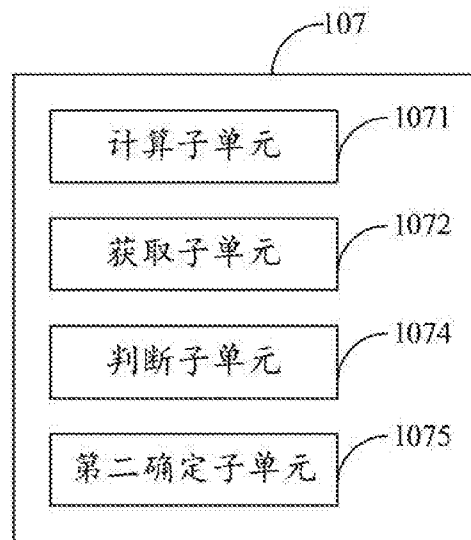


图 10

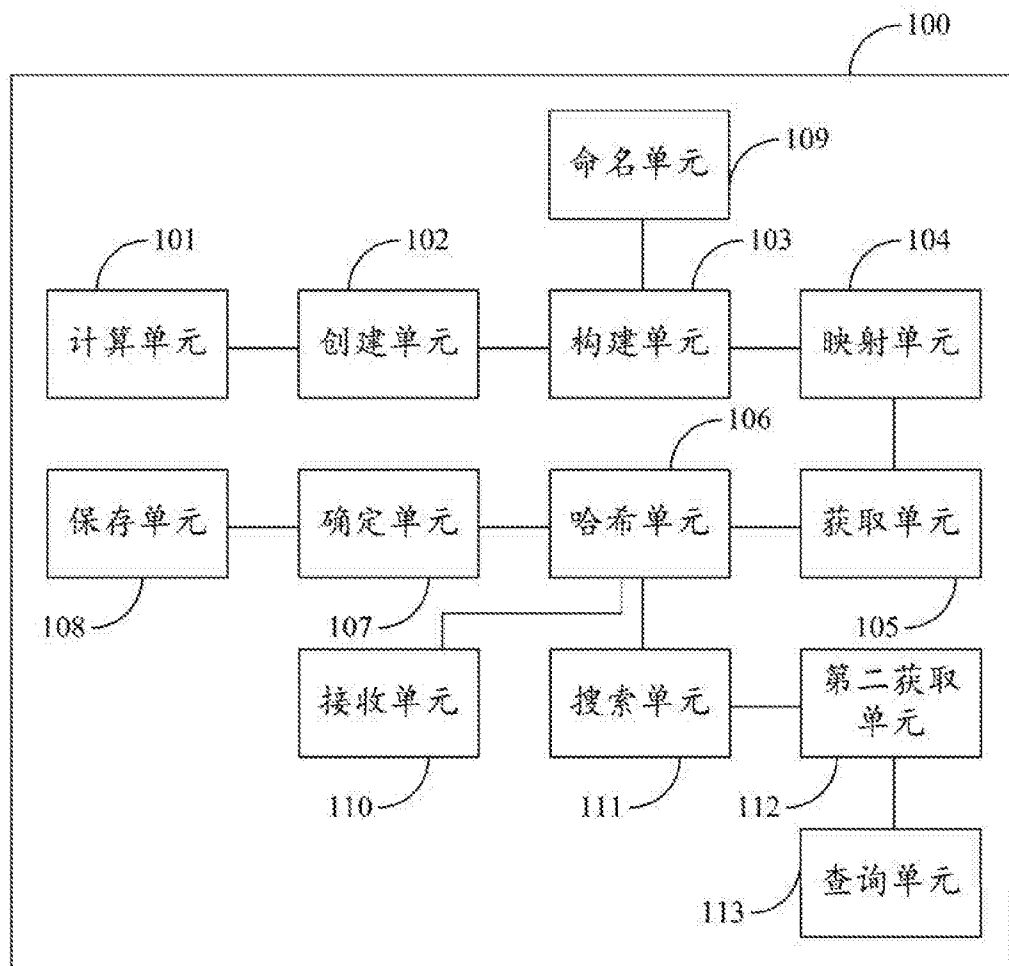


图 11

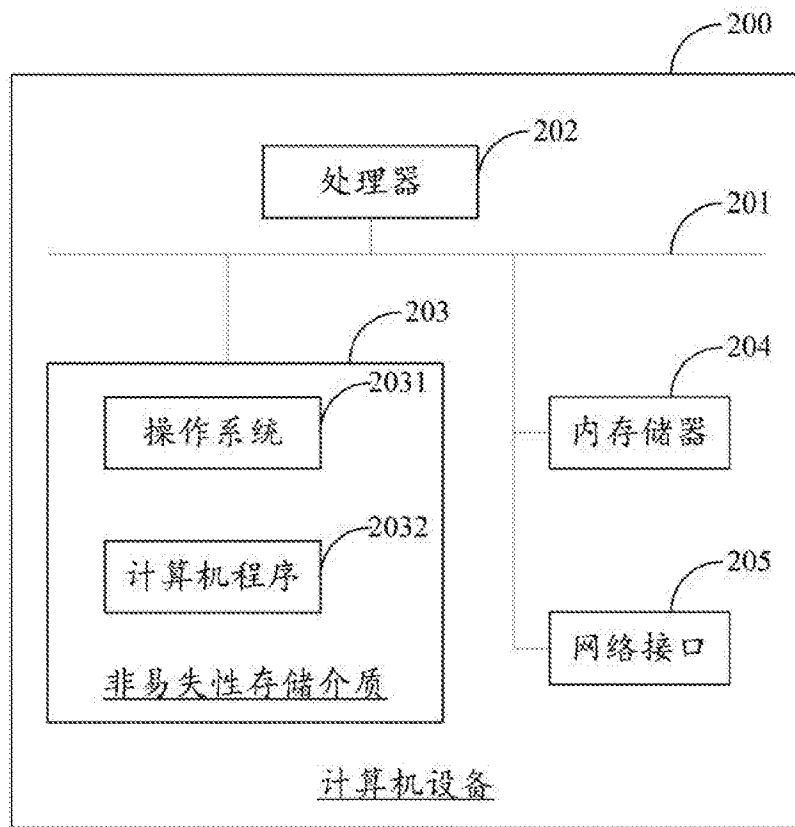


图 12

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/084322

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/30(2006.01)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)

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)

CNTXT; CNABS; DWPI; SIPOABS; CNKI: 主键, 映射, 哈希, 数据, 标签, 客户, 空间, 位置, primary, mapping, hash, data, tag, customer, interspace, location

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 105893466 A (CHINA CONSTRUCTION BANK CORPORATION) 24 August 2016 (2016-08-24) entire document	1-20
A	CN 106897409 A (BEIJING SEEYON SOFTWARE CO., LTD.) 27 June 2017 (2017-06-27) entire document	1-20
A	US 2017206254 A1 (DENSO CORP.) 20 July 2017 (2017-07-20) entire document	1-20
A	US 2013117273 A1 (ELECTRONICS AND TELECOMMUNICATIONS RESEARCH INSTITUTE) 09 May 2013 (2013-05-09) entire document	1-20

☐ Further documents are listed in the continuation of Box C.☒ 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

“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

26 October 2018

Date of mailing of the international search report

02 November 2018

Name and mailing address of the ISA/CN

State Intellectual Property Office of the P. R. China
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2018/084322

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	105893466	A	24 August 2016	None			
CN	106897409	A	27 June 2017	None			
US	2017206254	A1	20 July 2017	WO	2014174744	A1	30 October 2014
				JP	2014215803	A	17 November 2014
				JP	6221325	B2	01 November 2017
US	2013117273	A1	09 May 2013	KR	20130049111	A	13 May 2013
				US	8799291	B2	05 August 2014

国际检索报告

国际申请号

PCT/CN2018/084322

A. 主题的分类

G06F 17/30(2006.01)i

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

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06F

包含在检索领域中的除最低限度文献以外的检索文献

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))

CNTXT;CNABS;DWPI;SIPOABS;CNKI:主键, 映射, 哈希, 数据, 标签, 客户, 空间, 位置, primary, mapping, hash, data, tag, customer, interspace, location

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
A	CN 105893466 A (中国建设银行股份有限公司) 2016年 8月 24日 (2016 - 08 - 24) 全文	1-20
A	CN 106897409 A (北京致远互联软件股份有限公司) 2017年 6月 27日 (2017 - 06 - 27) 全文	1-20
A	US 2017206254 A1 (DENSO CORP) 2017年 7月 20日 (2017 - 07 - 20) 全文	1-20
A	US 2013117273 A1 (ELECTRONICS AND TELECOMMUNICATIONS RESEARCH INSTITUTE) 2013年 5月 9日 (2013 - 05 - 09) 全文	1-20

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 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

“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

国际检索实际完成的日期

2018年 10月 26日

国际检索报告邮寄日期

2018年 11月 2日

ISA/CN的名称和邮寄地址

中华人民共和国国家知识产权局(ISA/CN)
中国北京市海淀区蓟门桥西土城路6号 100088

受权官员

王燕

传真号 (86-10)62019451

电话号码 62089117

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/084322

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	105893466	A	2016年 8月 24日	无			
CN	106897409	A	2017年 6月 27日	无			
US	2017206254	A1	2017年 7月 20日	WO	2014174744	A1	2014年 10月 30日
				JP	2014215803	A	2014年 11月 17日
				JP	6221325	B2	2017年 11月 1日
US	2013117273	A1	2013年 5月 9日	KR	20130049111	A	2013年 5月 13日
				US	8799291	B2	2014年 8月 5日

表 PCT/ISA/210 (同族专利附件) (2015年1月)