

# (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2月7日 (07.02.2019)



(10) 国际公布号  
**WO 2019/024060 A1**

- (51) 国际专利分类号:  
**G06F 17/30** (2006.01)
- (21) 国际申请号: PCT/CN2017/095893
- (22) 国际申请日: 2017年8月3日 (03.08.2017)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (71) 申请人: 华为技术有限公司 (HUAWEI TECHNOLOGIES CO., LTD.) [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。
- (72) 发明人: 毕杰山 (BI, Jieshan); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 钟超强 (ZHONG, Chaoqiang); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 林铭 (LIN, Ming); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 季振峰 (JI, Zhenfeng);
- (74) 代理人: 北京三高永信知识产权代理有限责任公司 (BEIJING SAN GAO YONG XIN INTELLECTUAL PROPERTY AGENCY CO., LTD.); 中国北京市海淀区学院路蓟门里和景园A座1单元102室, Beijing 100088 (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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(54) Title: DATA STORAGE METHOD AND DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 数据存储方法、装置和存储介质

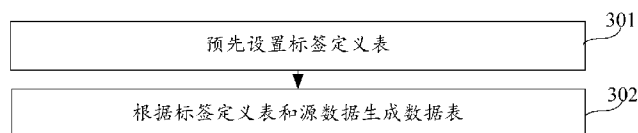


图 3

301 Pre-configure a label definition table  
302 Generate a data table according to the label definition table and source data

(57) Abstract: A data storage method and device, and a storage medium, relating to the field of optical information processing technology, the method comprising: acquiring at least one data record (401), each data record comprising a bearer identifier and at least one label value, the at least one data record being stored in a data table, the data table being used for recording a correlation between the bearer identifier and the label value; establishing a bitmap index corresponding to the at least one data record (402), the bitmap index comprising at least one bitmap, each bitmap corresponding to one label value, each bitmap comprising at least one bitmap bit, each bitmap bit being used for recording whether a bearer corresponding to one bearer identifier has a label value corresponding to the current bitmap. The present invention solves the problem in the related art of low data query efficiency, and achieves the effect of improving data query efficiency.

(57) 摘要: 一种数据存储方法、装置和存储介质, 涉及光信息处理技术领域, 所述方法包括: 获取至少一条数据记录(401), 每条数据记录包括一个载体标识和至少一个标签值, 所述至少一条数据记录存储于数据表中, 所述数据表用于记录载体标识与标签值的对应关系; 建立所述至少一条数据记录对应的位图索引(402), 所述位图索引包括至少一个位图, 每个位图对应于一个标签值, 每个位图包括至少一个位图位, 每个位图位用于记录一个载体标识所对应的载体是否具备当前位图所对应的标签值。解决了相关技术中数据查询效率低的问题; 达到了可以提高数据查询效率的效果。

WO 2019/024060 A1

(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))。

## 数据存储方法、装置和存储介质

## 技术领域

5 本申请涉及信息处理技术领域，特别涉及一种数据存储方法、装置和存储介质。

## 背景技术

HBase (Hadoop Database, Hadoop 数据库) 具有分布式、高可靠、高性能、基于 Key Value 存储等特点，因此越来越多的企业和用户使用 HBase 来存储构建数据表。

10 通常情况下，数据表包括多行数据记录，每一行数据记录包括承载体的标识和该承载体具有的各个标签的标签值。比如，对于用户 A 来说，其具有性别“女”和职业“工程师”两个标签值，则数据表中用户 A 所对应的行中包括用户 A 的标识、标签值“女”和标签值“工程师”。也就是说，相关技术中，数据表中记录了承载体的标识和其所具有的标签值的对应关系。

15 基于数据表的存储方式，当需要在数据表中查询数据时，根据承载体的标识进行数据查询时查询效率高，而在根据某一标签值或者标签值组合查询时，相关技术中只能根据 column value filter (行值过滤器) 按照承载体的标识逐行查询各个承载体的标签值，并且由于数据表的行数通常成千上万个，因此相关方案中，基于标签值进行数据查询时，其数据查询效率较低。

20

## 发明内容

为了解决相关技术中数据查询效率低的问题，本申请实施例提供了一种数据存储方法、装置和存储介质。所述技术方案如下：

第一方面，提供了一种数据存储方法，该数据存储方法包括：

25 获取至少一条数据记录，每条数据记录包括一个承载体标识和至少一个标签值。其中，至少一条数据记录可以存储在数据表中，该数据表用于记录承载体标识与标签值的对应关系。

在获取到至少一条数据记录之后，可以建立至少一条数据记录对应的位图索引。其中，位图索引包括至少一个位图，每个位图对应于一个标签值，每个位图包括至少一个位图位，  
30 每个位图位用于记录一个承载体标识所对应的承载体是否具备当前位图所对应的标签值。比如，某一承载体具备当前位图所对应的标签值，则其对应的位图位中可以设置有“1”，反之，若不具备当前位图所对应的标签值，则其对应的位图位中可以设置有“0”。其中，位图位为每个承载体的标识在位图中对应有的一个位置。

通过在获取到至少一条数据记录之后，建立该至少一条数据记录对应的位图索引，且  
35 位图索引中包括标签值和其对应的位图的对应关系，位图中包括用于记录承载体标识所对应的承载体是否具备对应的标签值的至少一个位图位；使得在后续基于标签值查询时，可以根据位图索引直接确定对应的数据，提高了数据查询效率。

在第一方面的第一种可能的实现方式中，在获取至少一条数据记录之前，还可以预先设置标签定义表，该预先设置的标签定义表包括预设的多个标签值；通常情况下，该标签

定义表为设计人员预先设置并保存的定义表，并且，该标签定义表中包括数据表中可能包括的所有标签值；

相应的，上述获取至少一条数据记录的步骤可以包括：

根据待存储的源数据和预先设置的标签定义表，生成该数据表。

5 通过预先设置标签定义表，使得在存储待存储的源数据时，可以根据标签定义表中定义的各个标签值确定源数据中的标签值，进而生成得到数据表。

结合第一种可能的实现方式，在第二种可能的实现方式中，实际实现时，标签定义表中还可以包括标签配置信息，该标签配置信息还包括用于表示每个标签值所对应的位图是否常驻内存的信息。可选地，每个标签值可以设置有对应的一条标签配置信息，该标签配置信息用于表示该标签值对应的位图是否常驻内存。对于标签配置信息中用于表示常驻内存的信息所对应的标签值，可以将其对应的位图加载至内存。一种实现中，可以仅包括用于表示需要常驻内存的标签配置信息，也可以仅包括用于表示不需要常驻内存的标签配置信息，还可以同时包括两者。

15 通常情况下，需要常驻内存的位图所对应的标签值是热查询的标签值，也即通常为查询频率高于预设阈值的标签值，因此，通过将此类标签值对应的位图常驻内存使得后续数据查询时，可以根据内存中的位图直接查询得到标签值对应的数据，提高了数据查询效率。

结合第一方面、第一方面的第一种可能的实现方式或者第一方面的第二种可能的实现方式，在第三种可能的实现方式中，位图索引可以划分为多个位图索引分区，每个位图索引分区中可以包括至少一个子位图，并且，每个位图索引分区对应于一个承载体标识的集合，不同位图索引分区所对应的承载体标识的集合不存在交集。其中，位图索引分区中的每个子位图对应于一个标签值。

通过将位图索引在多个位图索引分区中分布式存储，避免了受存储空间的限制而导致的所能存储的数据量有限的问题。

25 结合第一方面、第一方面的第一种可能的实现方式或者第一方面的第二种可能的实现方式，在第四种可能的实现方式中，与位图索引类似的，数据表也可以划分为多个数据分区，进而在多个数据分区中分布式存储数据表；可选地，可以根据承载体标识的第一范围将数据表划分为  $M$  个数据分区，每个数据分区中包括至少一个子数据表，并且每个数据分区所对应的承载体标识的范围为第一范围，且不同数据分区所对应的承载体标识的集合不存在交集。在数据表划分  $M$  个数据分区时，上述所说的位图索引分区可以根据  $M$  个数据分区确定得到  $N$  个位图索引分区，并且，与第三种可能的实现方式不同的是，在本实现方式中，每个位图索引分区所对应的承载体标识的范围大于等于第一范围，并且不同位图索引分区所对应的承载体的标识的集合不存在交集。其中， $M$  和  $N$  为正整数，且  $N$  小于或等于  $M$ ， $N$  大于或等于 2。

35 在本实现方式中，位图索引分区根据数据分区来确定，而实际实现时，位图索引分区还可以根据预先设置的分区方式分区，对此并不做限定。

通过将位图索引和数据表均划分分区，进而分布式存储，避免了受存储空间的限制而导致的所能存储的数据量有限的问题。

结合第三种可能的实现方式或者第四种可能的实现方式，在第五种可能的实现方式中，上述所说的每个位图索引分区中的子位图的数量为预先定义的标签定义表中设置的全部标

签值的数量。比如，标签定义表中的标签值共 10 个，则每个位图索引分区中的子位图的总数量也是 10 个。对于标签定义表中的某一标签值，若属于某一位图索引分区的各个承载体的标识所对应的承载体均不具备该标签值，则该标签值对应的位图中的内容均表示不具备标签值。

5 通过在每个位图索引分区中均设置标签定义表中的各个标签值所对应的位图，达到了在基于某一标签值查询时，可以根据该标签值所对应的位图索引确定得到对应的数据，提高了数据查询效率的效果。

结合第三种可能的实现方式、第四种可能的实现方式或者第五种可能的实现方式，在第六种可能的实现方式中，位图索引中的不同位图中相同的位图位对应于相同的承载体标识；每个位图索引分区中的不同子位图的相同的位图位对应于相同的承载体标识，不同位图索引分区中的不同子位图的相同的位图位对应于不相同的承载体标识。

通过设置位图索引或者位图索引分区中的位图位和承载体标识之间的对应关系，使得在根据位图索引或者位图索引分区进行数据查询时，可以快速的查询到对应的承载体标识，进而后续查询到对应的数据，提高了数据查询效率。

15 结合第一方面的第三种可能的实现方式、第四种可能的实现方式、第五种可能的实现方式或者第六种可能的实现方式，在第七种可能的实现方式中，当用户需要基于标签值查询时，可以发送数据查询请求，相应的，在接收到数据查询请求之后，执行数据查询。也即上述数据存储方法还包括：

接收数据查询请求，数据查询请求携带有至少一个目标标签值；

20 在多个位图索引分区中，分别确定至少一个目标标签值对应的至少一个目标子位图，分别确定至少一个目标子位图均包含的目标位图位所对应的目标承载体的标识；

查询数据表中目标承载体具有的标签值；

反馈目标承载体具有的标签值。

25 通过在基于标签值进行数据查询时，可以根据之前建立的位图索引直接查询该目标标签值对应的目标子位图，进而查询得到目标承载体的标签值；解决了相关技术中根据 column value filter 进行数据查询时数据查询效率低的问题，达到了可以提高数据查询效率的效果。

结合第一方面或者第一方面的上述任一种可能的实现方式，在第八种可能的实现方式中，在终端需要存储新数据时，服务器可以相应的获取到该新数据。其中，新数据中包括的新对应关系中承载体的标识和/或标签值是新的。这也就是说，该新数据可能是更新已经存储的数据的标签值的数据，也可能是之前未存储过的数据。

30 在获取到新数据之后，可以根据新对应关系更新位图索引。

通过在获取到新数据之后，更新位图索引，使得在接收到数据查询请求之后，可以根据最新的位图索引查询到所需要的数据，提高了数据查询效率。

结合第八种可能的实现方式，在第九种可能的实现方式中，上述所说的根据新对应关系更新位图索引的步骤可以包括：

确定新对应关系中的承载体的标识所属的位图索引分区；

在确定的位图索引分区中，更新新对应关系中的标签值，和/或更新新对应关系中的标签值所对应的位图。

第二方面，提供了一种数据存储装置，所述数据存储装置包括：存储器和处理器；所

述存储器中存储有指令；数据存储装置（可以是计算机设备，例如服务器）执行该指令，例如数据存储装置中的处理器执行该指令，使得该数据存储装置实现上述第一方面所述的数据存储方法。

5 第三方面，提供了一种数据存储装置，该装置包括至少一个单元，该至少一个单元用于实现上述第一方面所提供的数据存储方法。

第四方面，提供了一种计算机可读存储介质，该计算机可读存储介质中存储有指令；数据存储装置（可以是计算机设备，例如服务器）执行该指令，例如数据存储装置中的处理器执行该指令，使得该数据存储装置实现第一方面所提供的数据存储方法。

## 10 附图说明

图 1 是本申请各个实施例所涉及的位图索引的示意图。

图 2 是本申请各个实施例所涉及的实施环境的示意图。

图 3 是本申请一个实施例提供的生成数据表的方法流程图。

图 4 是本申请一个实施例提供的建立位图索引的方法流程图。

15 图 5 是本申请一个实施例提供的在底层数据库中位图索引的存储示意图。

图 6 和图 7 是本申请一个实施例提供的数据查询方法的方法流程图。

图 8 和图 9 是本申请一个实施例提供的数据存储装置的示意图。

## 具体实施方式

20

为了便于理解，首先对下述各个实施例所涉及的相关术语做简单介绍。

标签是一种内容组织方式，用于表征数据的某一特征进而帮助人们描述和分类内容。比如，常见的标签有性别、学历、职业、颜色等等。可选地，标签是人为规定的。

25 一种可能实现，标签可以包括枚举标签和布尔标签两种。枚举标签是指包括多个枚举值的标签，比如，学历包括专科、本科、研究生、博士等等，又比如，性别包括男或者女；而布尔标签只用于表示是否具备该标签，比如，是否有房、是否吸毒、是否有过犯罪记录等等。

30 在标签为枚举标签时，标签的标签值是指标签的具体取值。比如，以标签为学历为例，当学历是本科时其标签值为本科，当学历是研究生时其标签值为研究生。而在标签为布尔标签时，标签的标签值为标签本身。比如，当用户有房时其标签值为有房，再比如，当用户没有犯罪记录时，其对应的标签值为无犯罪记录。

35 承载体：是各个标签所描述的对象。可选地，承载体可以是人、车、电话号码或者虚拟用户帐号等等。一个承载体可以具有一个标签，也可以具有多个标签。比如，以承载体是人为例，描述人的标签可以有性别、学历、是否有房、是否有过犯罪记录等等。又比如，以承载体为车为例，描述车的标签可以有颜色、是否有违规记录等等。

数据表：为数据库中以承载体为索引建立的数据记录。数据表中的每条数据记录，记录一个承载体的标识，记录该个承载体具有的所有标签值，以及记录该个承载体的标识与该承载体具有的标签值之间的对应关系。

位图索引：为数据库中以数据表中的标签值为索引而建立的二级索引。可选地，位图

索引记录标签值和位图，还记录标签值和位图之间的一一对应关系。其中，位图中的每一个位图位对应一个承载体的标识，但位图中的不同位图位对应不同承载体的标识，即位图中的所有位图位与所有承载体的标识的一一对应。位图中的每个位图位记录一个承载体的标识所对应的承载体是否具备当前位图（该位图位所在的位图）所对应的标签值；比如，  
5 若一个标签值的位图中的某个位图位为 1，则代表该位图位对应的承载体具有该个标签值，反之，若该位图位为 0，则代表该位图位对应的承载体不具有该个标签值。不同位图中相同的位图位对应于相同的承载体的标识。

以承载体为虚拟用户账号为例，假设共 8 个虚拟用户帐号，每个虚拟用户帐号分别为 user1、user2、...、user8。具有标签值“网购达人”的集合为：user1、user4、user8，具有  
10 标签值“论坛活跃分子”的集合为：user1、user2、user8。在位图中为 8 个虚拟用户帐号分配的位图位依次为 1、2、3...8，如图 1 所示；对于标签值“网购达人”，其对应的位图为“10010001”；对于标签值“论坛活跃分子”，其对应的位图包括 11000001。以“网购达人”对应的位图“10010001”为例，位图中的第一个“1”表示位图位为 1 的虚拟用户帐号是网络达人，类似的，位图中的第二个“1”表示位图位为 4 的虚拟用户帐号也是网购达人，位图中的  
15 第三个“1”表示位图位为 8 的虚拟用户帐号也是网购达人；标签值“论坛活跃分子”对应的位图“11000001”表达的意思类似。由图 1 可知，user1 和 user8 同时具有“网络达人”和“论坛活跃分子”两个标签值。

请参考图 2，其示出了本申请各个实施例的实施环境的示意图。如图 2 所示，该实施环  
20 境中包括终端 210 和服务器 220。其中：

终端 210 中安装有用于请求数据存储或者数据查询的客户端。可选地，该客户端可以为浏览器。一种可能实现，终端 210 可以为诸如手机、平板电脑、电子阅读器、台式电脑之类的设备。终端 210 可以通过有线或者无线网络与服务器 220 连接。

服务器 220 可以为一台或多台服务器；可选地，多台服务器可以以服务器集群的方式  
25 为终端 210 提供数据库服务。一种可能实现，服务器 220 中设置有数据库，该数据库可以为 HBase、Mongo 数据库 (Mongo Database, MongoDB)、分布型关系数据库服务 (Distribute Relational Database Service, DRDS)、Volt 数据库 (Volt Database, VoltDB)、和 ScaleBase 等分布式数据库。

30 下面介绍数据表的生成方法

请参考图 3，其示出了本申请一个实施例提供的生成数据表的方法的流程图，本实施例以该数据存储方法用于图 2 所示的服务器中来举例说明。如图 3 所示，该数据存储方法包括步骤 301 和步骤 302。

步骤 301，预先设置标签定义表。

35 标签定义表可以为服务器预先获取并存储的信息。可选地，标签定义表可以以独立的文件的形式存储，如以可扩展标记语言 (Extensible Markup Language, XML) 文件的形式存储，也可以在第三方分布式存储系统中存储，如存储至 ZooKeeper。

该标签定义表记录预设的多个标签值。一种可选的预设方式，根据历史数据设置标签包含的标签值，或者人为定义标签包含的标签值。

表 1 示出了一种可能的标签定义表。当然，表 1 还可能会包括更多或者更少的标签，对此并不做限定。

标签	标签值	标签配置信息
性别	男，女	常驻内存
学历	专科，本科，研究生，博士	不常驻内存
职业	学生，教师，个体，企业员工	常驻内存
网购狂人	网购狂人	常驻内存
吸毒者	吸毒者	不常驻内存

表 1

5 可选地，如表 1 所示，标签定义表还可以包括标签配置信息，该标签配置信息包括表示标签值是否需要常驻内存，需要常驻内存的标签值所对应的位图也需要是否常驻在内存，不需要常驻内存的标签值所对应的位图也不需要是否常驻在内存。

10 在表 1 中，对需要常驻内存的标签值设置了标识“常驻内存”，对不需要常驻内存的标签值设置了标识“不常驻内存”。应当理解，也可以对需要常驻内存的标签值设置标识“常驻内存”，可以对不需要常驻内存的标签值不设置标识，表 1 对不需要常驻内存的标签值设置标识“不常驻内存”仅是一种示例。

数据库存储在服务器的存储介质（例如硬盘）中，在初始使用数据库时便将标签值和该标签值对应的位图加载至内存，并且保持内存中的位图与存储介质中的位图同步更新，使得每次查询该标签值时可以直接使用内存中该标签值的位图而不需要访问存储介质中的位图，提高了查询效率，节省了查询时间。

15 可选地，标签定义表还可以包括每个标签值的生命周期，该生命周期是指该标签值为有效的时间段；即不属于该生命周期的其他时间，该标签值为无效。

20 可选地，服务器还可以为表 1 中的每个标签值分配一个标签号。在存储标签值与位图的映射关系时可以用该标签号替代标签值，存储标签号相对于存储标签值可以节省存储空间。另外，在中，可以根据标签号查询到对应的标签值，可以根据标签值查询对应的标签号。

步骤 302，服务器根据标签定义表和源数据生成数据表。

25 服务器首先获取待存储的源数据。一种可选实现，服务器接收终端发送的数据存储请求，该数据存储请求中携带有待存储的源数据。又一种可选实现，服务器可以主动从终端中获取该源数据，或者，服务器从存储该源数据的数据库中获取该源数据。因此，本发明实施例对服务器如何获取源数据并不做限定。

待存储的源数据的数量可以是一条或者多条。可选地，参考表 2，多条源数据可以以表格的形式存在，表格中的每一行表示一条源数据，每条源数据具有唯一的承载体的标识。

承载体的标识	
A01	性别:男，学历:本科，职业:学生
A02	性别:女，学历:专科，职业:个体，网购达人

B01	性别:男, 学历:专科, 职业:企业员工
B02	性别:女, 学历:本科, 职业:学生
C01	性别:男, 学历:研究生,
C02	性别:女, 学历:专科, 职业:企业员工, 网购达人
D01	性别:男, 学历:研究生, 职业:企业员工, 网购达人
D02	性别:女, 学历:本科, 职业:学生
D03	性别:男, 学历:专科, 职业:企业员工
E01	性别:男, 学历:研究生, 职业:个体, 吸毒者
E02	性别:女, 学历:专科, 职业:个体
E03	性别:男, 学历:本科, 职业:学生
F01	性别:男, 学历:研究生, 职业:教师,
F02	性别:女, 学历:专科, 职业:企业员工, 网购达人
F03	性别:男, 学历:本科, 职业:学生
F04	性别:女, 学历:研究生, 职业:个体, 吸毒者
.....	.....

表 2

服务器获取到源数据之后, 服务器可以根据标签定义表确定每条源数据具有的所有标签值。比如, 对于表 2 所示的源数据, 表 3 示出了服务器根据表 1 确定得到的各条源数据的标签值; 其中, 表 3 中每个[]中的内容表示一个标签值。

承载体的标识	标签值
A01	[性别:男][学历:本科][职业:学生]
A02	[性别:女][学历:专科][职业:个体][网购达人]
B01	[性别:男][学历:专科][职业:企业员工]
B02	[性别:女][学历:本科][职业:学生]
C01	[性别:男][学历:研究生]
C02	[性别:女][学历:专科][职业:企业员工][网购达人]
D01	[性别:男][学历:研究生][职业:企业员工][网购达人]
D02	[性别:女][学历:本科][职业:学生]
D03	[性别:男][学历:专科][职业:企业员工]
E01	[性别:男][学历:研究生][职业:个体][吸毒者]
E02	[性别:女][学历:专科][职业:个体]
E03	[性别:男][学历:本科][职业:学生]

F01	[性别:男][学历:研究生][职业:教师]
F02	[性别:女][学历:专科][职业:个体][网购达人]
F03	[性别:男][学历:专科][职业:企业员工]
F04	[性别:女][学历:本科][职业:学生]
.....	.....

表 3

如表 3 所示的数据表，该数据表包括多条数据记录。每条数据记录包括一个承载体的标识，例如第一条数据记录包括标识 A01 和第九条数据记录包括标识 D03；每条数据记录还包括承载体的标识对应的标签值，例如第一条数据记录包括标识 A01 对应的标签值“[性别:男][学历:本科][职业:学生]”，再例如第九条数据记录包括标识 D03 对应的标签值“[性别:男][学历:专科][职业:企业员工]”。如表 3 所示，所述数据表中的每一条数据记录包括一个承载体的标识与该承载体的所有标签值的对应关系，例如该数据表的第一条数据记录包括标识 A01 与标签值“[性别:男][学历:本科][职业:学生]”的对应关系。

对于数据表，可以根据承载体的标识，查询数据表中该承载体的标识对应的所有标签值，即查询得到该载体具有的所有标签值。

可选地，可以将数据表划分为多个数据分区，分布式地存储数据表的多个数据分区。划分数据分区的可选方式，可以通过指定数据分区的数量来划分数据分区，或者可以直接定义每个数据分区的分区区间。

举例说明，为数据表设置如下分区区间：

- 15 数据分区 1: [ , A)
- 数据分区 2: [ , B)
- 数据分区 3: [B , C)
- 数据分区 4: [C , D)
- 数据分区 5: [D , E)
- 20 数据分区 6: [E , F)

.....

表 3 按照上述分区区间划分后得到表 4 所示的数据分区。

数据分区	承载体的标识	标签值
数据分区 1	A01	[性别:男][学历:本科][职业:学生]
	A02	[性别:女][学历:专科][职业:个体][网购达人]
数据分区 2	B01	[性别:男][学历:专科][职业:企业员工]
	B02	[性别:女][学历:本科][职业:学生]
数据分区 3	C01	[性别:男][学历:研究生]
	C02	[性别:女][学历:专科][职业:企业员工][网购达人]
数据分区 4	D01	[性别:男][学历:研究生][职业:企业员工][网购达人]
	D02	[性别:女][学历:本科][职业:学生]

	D03	[性别:男][学历:专科][职业:企业员工]
数据分区5	E01	[性别:男][学历:研究生][职业:个体][吸毒者]
	E02	[性别:女][学历:专科][职业:个体]
	E03	[性别:男][学历:本科][职业:学生]
数据分区6	F01	[性别:男][学历:研究生][职业:教师]
	F02	[性别:女][学历:专科][职业:个体][网购达人]
	F03	[性别:男][学历:专科][职业:企业员工]
	F04	[性别:女][学历:本科][职业:学生]
....	.....	.....

表 4

数据表的每个数据分区可以自动裂变或者扩展。比如，随着时间推移，某个数据分区的数据（例如承载体的标识，或者标签值）越来越多，在该个数据分区的数据量达到分裂阈值，服务器可以将该个数据分区进行分裂成两个数据分区，从而避免由于该个数据分区的存储空间被存满之后无法继续向该个数据分区写入新数据。

每个数据分区包括至少一个子数据表。以表 4 的数据分区 1 为例，如果数据分区 1 具有一个子数据表，该子数据表包括第一条数据记录（包括标识 A01 和标识 A01 对应的数据标签）、第二条数据记录（包括标识 A02 和标识 A03 对应的数据标签）；当然，数据分区 1 还可以具有两个子数据表，其中一个子数据表记录该第一条数据记录，另一个子数据表记录该第二条数据记录。

每个数据分区中的子数据表所对应的承载体的标识的集合不存在交集。以表 4 为例，数据分区 1 的子数据表所对应的承载体的标识的集合为区间[ ,B)，数据分区 2 的子数据表所对应的承载体的标识的集合为区间[B ,C)，数据分区 1 的子数据表所对应的集合为区间[ ,B) 与数据分区 2 的子数据表所对应的集合为区间[B ,C)不存在交集，以此类推，表 4 中不同数据分区 1 的子数据表所对应的集合都不存在交集。

下面介绍数据表的更新

(1)、更新标签值。

若数据表中已存储有目标承载体的标识，服务器使用待更新的标签值更新该目标承载体的标识所具有的标签值，更新标签值可以包括增加标签值或者删除标签值。

举例说明，在表 4 中，标识 A01 的数据记录具有三个标签值 “[性别:男][学历:本科][职业:学生]”。参见表 5，如果为标识 A01 的数据记录增加标签值 “[网购达人]”，标签值更新后，标识 A01 的数据记录具有四个标签值 “[性别:男][学历:本科][职业:学生][网购达人]”。参见表 6，如果期望删除标识 A01 的数据记录所具有的标签值 “[职业:学生]”，标签值更新后，标识 A01 的数据记录具有二个标签值 “[性别:男][学历:本科]”。

分区	承载体的标识	标签值
数据分区1	A01	[性别:男][学历:本科][职业:学生][网购达人]
	A02	[性别:女][学历:专科][职业:个体][网购达人]

表 5

分区	承载体的标识	标签值
数据分区1	A01	[性别:男][学历:本科]
	A02	[性别:女][学历:专科][职业:个体][网购达人]

表 6

上述举例仅以更新数据记录中的单个标签值为例，当然也可以同时更新数据记录中的多个标签值，还可以同时更新多条数据记录中的标签值，在此不再赘述。

5 (2)、增加承载体。

服务器在数据表中增加数据记录，包括增加目标承载体的标识、并在数据表中同时增加目标承载体的标识具有的所有标签值。

举例说明，在表 4 中增加标识 A03 的数据记录，参见表 7，具体是在数据分区 1 的承载体的标识这一列中，标识 A02 的下一行记录标识 A03；相应地，在标签值这一列记录标识 A03 对应的“标签值[性别:女][学历:专科][职业:企业员工][网购达人]”。

10

数据分区	承载体的标识	标签值
数据分区1	A01	[性别:男][学历:本科][职业:学生]
	A02	[性别:女][学历:专科][职业:个体][网购达人]
	A03	[性别:女][学历:专科][职业:企业员工][网购达人]

表 7

上述举例仅以增加标识 A03 的数据记录为例，当然，服务器也可以在数据分区 1 中增加多条数据记录，还可以在多个数据分区分别增加多条数据记录，在此不再赘述。

15 另外，服务器还可以删除数据表的数据分区中的数据记录，包括删除承载体的标识和删除该承载体具有的标签值。例如在删除标识 A02 的数据记录时，会删除标识 A02，以及删除标识 A02 对应的标签值“[性别:女][学历:专科][职业:个体][网购达人]”。

下面介绍位图索引的生成方法

20 服务器可以为该每个承载体的标识分配一个在位图中的位图位，即位图位为每个承载体的标识在位图中对应有的一个位置。如表 8 所示的一种位图位分配举例，对于标识 A01 的承载体，为该承载体分配的位图位为 1；对于标识 A02 的承载体，为该承载体分配的位图位为 2，以此类推。

承载体的标识	位图位
A01	1
A02	2
B01	3
B02	4
C01	5
C02	6

D01	7
D02	8
D03	9
E01	10
E02	11
E03	12
F01	13
F02	14
F03	15
F04	16
.....	.....

表 8

请参考图 4，其示出了本申请一个实施例提供的建立位图索引方法的流程图，本实施例以该数据存储方法用于图 2 所示的服务器中来举例说明。如图 4 所示，该数据存储方法包括步骤 401 和步骤 402。

5 步骤 401，服务器获取至少一条数据记录。

所述至少一条数据记录存储于数据表中。所述数据记录包括一个承载体的标识和至少一个标签值；所述数据表中的每条数据，记录一个承载体的标识，和记录该个该承载体具有的所有标签值。对数据表和数据记录的详细描述，可以参见上述以表 3 或表 4 为例的相关描述，在此不再赘述。

10 步骤 402，服务器建立所述至少一条数据记录对应的位图索引。

所述数据表具有至少一条数据记录，服务器为每条数据记录建立一个位图；因此，位图索引包括至少一个位图，每条数据记录对应该位图索引的一个位图。

每个位图对应于一个标签值，每个位图包括至少一个位图位。每个位图位记录一个承载体的标识所对应的承载体是否具备当前位图（该位图位所在的位图）所对应的标签值，位图索引中的不同位图中的相同位图位对应于相同的承载体的标识。对位图索引的结构描述，可以上述介绍位图索引的相关描述，在此不再赘述。

15 举例说明，以表 3 或表 4 所示的数据表为基础、并按照表 8 所示承载体的标识与位图位的对应关系，得到表 9 所示的位图索引。在表 9 所示的位图索引中，标签值“吸毒者”对应一个位图 “[0000000001000000....]”；该位图中的第 10 个位图位为“1”，代表标识 E01 的承载体具有标签值“吸毒者”。在表 9 所示的位图索引中，标签值“网购达人”对应一个位图 “[0100011000000100....]”；该位图中的第 10 个位图位为“0”，代表标识 E01 的承载体不具有标签值“网购达人”。

标签值	位图
性别:男	[1010101011011010....]
性别:女	[0101010100100101....]

学历:专科	[0110010010100100....]
学历:本科	[1001000100010010....]
.....	.....
网购达人	[0100011000000100....]
吸毒者	[0000000001000000....]
.....	.....

表 9

可选地，用数组表示位图中为“1”的位图位。例如在表 9 所示的位图索引中，标签值“吸毒者”对应一个位图“[0000000001000000....]”，则该位图可以表示为数组[10]；标签值“网购达人”对应一个位图“[0100011000000100....]”，则该位图可以表示为数组[2 6 7 14]。

5 采用数组替换位图在数据库中存储，可以节省存储空间。可选地，在内存中仍采用位图存储，这样可以在标签值查询时按照位图位对目标位图进行运算（例如与运算、或运算）来确定目标位图位。

可选地，用数组表示位图中占比较小的位图位。例如在表 9 所示的位图索引中，标签值“吸毒者”对应一个位图“[0000000001000000....]”，该位图中“1”占比较小，则该位图可以表示为包含“1”的数组[10]；标签值“网购达人”对应一个位图“[0100011000000100....]”，该位图中“1”占比较小，则该位图可以表示为包含“1”的数组[2 6 7 14]。采用数组替换位图在数据库中存储，可以节省存储空间。可选地，在内存中仍采用位图存储，这样可以在标签值查询时按照位图位对目标位图进行运算（例如与运算、或运算）来确定目标位图位。

10

可选地，表 9 所示结构是位图索引在内存中的存储结构，而在数据库中位图索引采用 Base+Delta 进行存储。其中，Base 为位图索引的具体内容，其实际存储结构与内存中的存储结构类似，如在位图中表示具备标签值或者不具备标签值的索引所占的比例低于预设比例，则可以通过数组表示该类索引，反之则按位存储；请参考图 5，Delta 包括至少一个 Key Value，每个 Key Value 对应于一个变更位图中的内容的变更操作，该变更操作包括更新某一位中已有内容的操作，比如，Delta 包括“将位图中的第 10 位更改为 1”以及“将位图中的第 15 位更改为 0”；或者，该变更操作包括在某一位新增预设内容的操作，比如，包括“在位图的第 21 位新增 1”。可选地，当 Base 的数据大小达到第一大小小时，则合并 Base 和 Delta，合并后的 Base 为合并后的文件的存储路径，合并后的 Delta 中存储有合并后的文件。当然一种可能实现，位图索引在数据库中还可以以文件的形式存储，对此并不做限定。

15

20

可选地，在初始使用数据库时，根据表 1 中的标签定义表的标签配置信息加载需要常驻内存的标签值，以及加载该需要常驻内存的标签值所对应的位图。保持内存中的位图与存储介质中的位图同步更新，使得每次查询该标签值时可以直接使用内存中该标签值的位图而不需要访问存储介质中的位图，提高了查询效率，节省了查询时间。

25

在本发明实施例提供的数据存储方法中，可以将位图索引划分为多个位图索引分区。

30 将位图索引划分为多个位图索引分区，与将数据表划分为数据分区类似，位图索引分区的划分方式可以类似参考数据分区的划分方式。比如，表 10 示出一种位图索引分区的划分结果，位图索引分区 1 为区间[ , E)，位图索引分区 2 为区间[E, G)，以此类推。在每个

位图索引分区中，每个承载体的标识的位图位是唯一的；在不同位图索引分区中，同一位图位对应不同承载体的标识。

位图索引分区	承载体的标识	位图位
位图索引分区1	A01	1
	A02	2
	B01	3
	B02	4
	C01	5
	C02	6
	D01	7
	D02	8
	D03	9
位图索引分区2	E01	1
	E02	2
	E03	3
	F01	4
	F02	5
	F03	6
	F04	7
....	.....	.....

表 10

5 可选地，每个位图索引分区所对应的区间为用户预先设置的，服务器按照预先设置的区间划分出位图索引分区。例如，预先设置不重叠的三个区间[ , E), [E , G)和[G , )，根据该三个区间划分出位图表 10 所示的位图索引分区 1，位图索引分区 2 和位图索引分区 3（未示意出）。

10 可选地，服务器可以将 M 个数据分区划入 N 个位图索引分区；M、N 为正整数，N 小于或等于 M，N 大于或等于 2。M 个数据分区中的每个数据分区属于唯一的位图索引分区，N 个位图索引分区中的每个位图索引分区包含至少一个数据分区。

确定位图索引分区后，数据表中的承载体的标识所属的位图索引分区也确定了。相应地，一个位图索引分区具有位图索引中的一个子位图索引，该个子位图索引包括该个位图索引分区对应的所有承载体的标识所对应的所有子位图，每个子位图包括该所有承载体的标识所对应的位图位；以此类推，每个位图索引分区中包括至少一个子位图。

15 表 11 示意了划分位图索引分区后的位图索引。以位图索引分区 1 为例，位图索引分区 1 对应一个子位图索引，该个子位图索引包括多个子位图，例如标签值“性别:男”对应的子位图“[101010101]”，该子位图包括表 9 中“性别:男”对应的位图“[1010101011011010....]”

中的前九个位图位（即第一个位图位至第九个位图位）。以此类推，得到位图索引分区 1 中每个标签值对应的子位图。

分区	标签值	子位图
位图索引 分区1	性别:男	[101010101]
	性别:女	[010101010]
	学历:专科	[011001001]
	学历:本科	[100100010]
	.....	.....
	网购达人	[010001100]
	吸毒者	[000000000]
位图索引 分区2	性别:男	[1011010]
	性别:女	[0100101]
	学历:专科	[0100100]
	学历:本科	[0010010]
	.....	.....
	网购达人	[0000100]
	吸毒者	[1000000]
	.....	

表 11

5 应注意，对于表 11 中的位图索引分区 2，该位图索引分区 2 中的每个位图位与承载体标识的对应关系，与表 9 中每个位图位与承载体标识的对应关系是不同的。例如表 11 所示位图索引分区 2 中的每个子位图的第一个位图位对应承载体的标识 E01，而表 9 所示位图索引的位图中的第十个位图位对应承载体的标识 E01。

10 另外，与表 9 的位图类似，表 11 所示位图索引分区的子位图也是记录标签值与位图位（对应承载体的标识）的对应关系，区别在于表 9 的位图记录的位图位的数量多于表 11 所示子位图记录的位图位的数量。

结合上述分析可知，位图索引中的不同位图中相同的位图位对应于相同的承载体的标识；每个位图索引分区中的不同子位图的相同的位图位对应于相同的承载体的标识，不同位图索引分区中的不同子位图的相同的位图位对应于不相同的承载体的标识。

15 可选地，每个位图索引分区中的子位图所对应的承载体的标识的集合相互独立，不存在交集。比如在表 10 和表 11 中，位图索引分区 1 所对应的承载体的标识的集合为区间[ , E)，位图索引分区 2 所对应的承载体的标识的集合为区间[E , G)；区间[ , E)与区间[E , G)没有交集。更具体地，每个承载体的标识属于唯一的集合（即区间），因此该个承载体的标识对应的位图位也唯一属于该集合对应的位图索引分区。举例说明，承载体的标识 F04 属

于区间[E, G), 该区间[E, G)对应位图索引分区 2, 则承载体的标识 F04 的位图位在位图索引分区 2 中, 即包含承载体的标识 F04 的位图位的所有子位图都在位图索引分区 2 中。

5 可选的, 每个位图索引分区中的子位图所对应的承载体的标识的集合大于等于第一范围, 该第一范围为数据分区中的子数据表所对应的承载体的标识的集合。具体地, 每个位图索引分区包含至少一个数据分区, 数据表中的每个数据分区属于唯一的位图索引分区; 因此, 每个数据分区的第一范围 (即该数据分区中的子数据表所对应的承载体的标识的集合) 小于或等于该数据分区所属索引分区对应的标识集合 (即该索引分区的子位图所对应的承载体的标识的集合)。

10 可选地, 在每个位图索引分区中, 为标签定义表中的所有标签值分别建立一个子位图。具体地, 无论子位图索引中的承载体是否具有某个标签值, 在该子位图索引中为该标签值建立一个子位图。以表 11 为例, 位图索引分区 1 中的所有承载体均不具有标签值“吸毒者”, 在位图索引分区 1 的子位图索引中, 仍为该标签值“吸毒者”建立子位图“[000000000]”。这样, 如果后续位图索引分区 1 中的某个承载体更新为具有该标签值“吸毒者”, 直接更新该子位图“[000000000]”, 不需要增加建立该标签值“吸毒者”对应的子位图, 由于仅更改  
15 子位图的位图位相对于增加子位图不需要额外的存储资源, 因此不会导致位图索引分区 1 因数据量增大而分裂。

举例说明, 位图索引分区一旦确定存储容量大小, 如果该位图索引分区因增加子位图而导致所需存储容量大幅度增加, 当增加到预设的存储容量阈值时, 该位图索引分区不能存储新数据, 需要对该位图索引分区分裂。

20 另外, 以表 11 为例, 即使位图索引分区 1 增加新的承载体, 该承载体具有标签值“吸毒者”, 为该承载体的标识在子位图中增加位图位所占用的存储资源较小, 例如 1 个比特。因此, 增加承载体而引发的增加位图位通常不会导致图索引分区 1 分裂。

由于在每个位图索引分区中为标签定义表中的所有标签值分别建立一个子位图, 因此每个位图索引分区中的子位图的数量为标签定义表中的全部标签的数量。比如, 假设标签  
25 定义表中设置的标签值的总个数为 10 个, 则在每个位图索引分区分别为标签定义表中的所有标签值分别建立子位图之后, 每个位图索引分区的子位图的数量也为 10 个。

可选地, 由于位图索引分区扩展或者分裂会导致位图索引中的所有位图索引分区需要重建, 代价较高, 在本实施例中服务器将位图索引分区设置为不可分裂或者扩展。

30 然而, 随着时间的推移, 位图索引分区中的容量会不断增加或者变化; 比如, 大数据场景中承载体的个数不断增多, 或者, 因标签定义表中新增标签值而导致每个位图索引分区相应增加子位图, 又或者, 大量承载体具有的标签值失效。因此, 为了避免因存储容量不够而导致位图索引分区中无法存储新增数据, 服务器可以重建位图索引分区。可选地, 服务器可以基于 MapReduce 机制, 通过分布式读取数据表中存储的数据记录, 并且获取更新后的位图索引分区, 根据更新后的位图索引分区重建每个位图索引分区中的子位图索引。  
35 具体地, 服务器可以为每个位图索引分区的承载体重新分配位图位, 并且根据分配的位图位生成各个标签值所对应的子位图。

下面介绍位图索引的更新方法

服务器存储新数据, 该新数据可以是数据记录; 该新数据包括的对应关系中承载体的

标识和/或标签值是新的，则服务器根据新对应关系更新位图索引。

其中，根据新对应关系更新位图索引的步骤包括：

第一步，确定新对应关系中的承载体的标识所属的位图索引分区。

5 参考上述关于位图索引分区中的子位图所对应的承载体的标识的集合的介绍，由于每个承载体的标识唯一属于一个位图索引分区，因此可以确定新对应关系中的承载体的标识所属的位图索引分区。

举例说明，结合表 10，若新对应关系中的承载体的标识为 F03，则可以确定该承载体的标识所属的位图索引分区为位图索引分区 2；而若新对应关系中的承载体的标识为 B02，则可以确定其所属的位图索引分区为位图索引分区 1。

10 第二步，在确定的位图索引分区中，更新新对应关系中的标签值，和/或更新新对应关系中的标签值所对应的位图。

#### (1)、更新标签值。

15 若数据表中已存储有新数据中的目标承载体的标识，更新确定的位图索引分区中更新的标签值所对应的子位图。

可选地，若承载体在更新之前具备该标签值而更新之后不具备该标签值，则服务器在确定的位图索引分区中，将该标签值的子位图中该承载体的标识所对应的位图位由“1”更新为“0”。举例说明，承载体的标识在子位图中所对应的位图位为 3，更新前的子位图为 1011，则更新后的子位图为 1001。而若承载体在更新之前不具备该标签值而更新之后具备该标签值，则服务器在确定的位图索引分区中，将该标签值的子位图中该承载体的标识所对应的位图位由“0”更新为“1”。举例说明，承载体的标识在子位图中所对应的位图位为 5，更新前的子位图为 10110，则更新后的子位图为 10111。

20 可选地，用数组表示子位图中为“1”的位图位，则若承载体在更新之前不具备该标签值而更新之后具备该标签值，则服务器在确定的位图索引分区中，该标签值的数组中新增该承载体的标识的位图位。比如，承载体的标识在子位图中所对应的位图位为 3，更新的标签值所对应的初始子位图为 (1, 7)，则更新后的子位图为 (1,3,7)；而若承载体在更新之前具备该标签值而更新之后不具备该标签值，则服务器删除确定的位图索引中，标签值的数组中该承载体的标识所对应的位图位；比如，若源数据在子位图中所对应的位图位为 3，更新的标签值所对应的初始子位图为 (1,3,7)，则更新后的子位图为 (1, 7)。类似的，若  
30 用数组中为“0”的位图位，则服务器可以采用类似的更新方法更新子位图中的内容，在此不再赘述。

可选地，在某种情况下，若新数据中的某一标签值不是标签定义表中预先定义的标签值，则服务器在确定的位图索引分区中增加新增的标签值所对应的子位图。其中，属于该位图索引分区中的各条数据在新增的子位图中的位图位与在该位图索引分区中的其他子位图  
35 图中的位图位一致。

#### (2)、更新承载体。

若新数据（例如数据表中的新数据记录）中增加承载体的标识，则在确定的位图索引分区中为该承载体分配位图位，根据该新数据的承载体的标识具有的所有标签值更新子位图。

可选地，对于确定的位图索引分区中的各个标签值所对应的子位图，若新数据的载体具备该标签值，则在子位图中分配的位图位处增加用于表示具备该标签值的信息；举例说明，假设为新数据的载体分配的位图位为 5，某一标签值对应的初始子位图为 0101，则更新后的子位图为 01011；而若新数据的载体不具备该标签值，则在子位图中分配的位图位处增加用于表示不具备该标签值的信息；举例说明，仍然以为新数据的载体分配的位图位为 5 来举例，某一标签值对应的初始子位图为 0101，则更新后的子位图为 01010。

可选地，若用数组表示位图位中为“1”的位图位，则对于确定的位图索引分区中各个标签值所对应的子位图，若新数据的载体具备该标签值，则在子位图中增加新分配的位图位；举例说明，假设为该新数据的载体分配的位图位为 5，该载体具有某一标签值，该标签值对应的初始子位图为 (3)。类似的，若用数组表示位图位中为“0”的位图位，则服务器可以采用类似方法更新，在此不做赘述。

可选地，若位图索引不划分索引分区，则服务器根据新对应关系直接更新位图索引。其具体更新方法与上述类似，不同的是，在划分索引分区时，需要先确定新数据中的载体的标识所属的位图索引分区，然后再更新；而此处无需确定直接更新，在此不再赘述。

请参考图 6，其示出了本申请一个实施例提供的数据查询方法的方法流程图，本实施例以该数据查询方法用于图 2 所示的服务器中来举例说明，如图 6 所示，该数据查询方法包括：

步骤 601，接收数据查询请求，数据查询请求中携带有至少一个目标标签值。

终端需要查询具备某个标签值的数据时，终端可以发送数据查询请求至服务器，相应的，服务器可以接收到该数据查询请求。其中，数据查询请求中携带有需要查询的至少一个目标标签值。

可选地，终端需要查询具备标签值 A 的数据时，终端可以发送携带有标签值 A 的数据查询请求至服务器；在终端需要查询不具备标签值 A 的数据时，终端可以发送携带有“not A”的数据查询请求至服务器。

可选地，当数据查询请求中携带有至少两个目标标签值时，该至少两个目标标签值之间可以是“和”的关系，也可以为“或”的关系，还可以为“非”的关系。

比如，当终端需要查询同时满足标签值 A 和标签值 B 的数据时，终端可以发送携带有“A and B”的数据查询请求至服务器。又比如，当终端需要查询标签值为 A 或者标签值 B 的数据时，终端可以发送携带有“A or B”的数据查询请求至服务器。再比如，当需要查询标签值为 A 非 B 的数据时，终端可以发送携带有“A not B”的数据查询请求至服务器。

步骤 602，根据位图索引，确定每个目标标签值对应的目标位图。

位图索引包括至少一个位图，每个位图对应于一个标签值，每个位图包括至少一个位图位，每个位图位用于记录一个载体的标识对应载体是否具备当前位图所对应的标签值，不同位图中相同的位图位对应于相同的载体的标识；位图索引为数据表对应的索引。其中，该位图索引为上述实施例中创建的索引，对于其具体实现详见上述实施例，在此不再赘述。

服务器接收到数据查询请求之后，提取数据查询请求中的目标标签值，然后在位图索引中查询提取得到的每个目标标签值所对应的目标位图。

可选地，若目标标签值有至少两个，则服务器可以根据位图索引查询每个目标标签值所对应的目标位图。比如，数据查询请求中携带的目标标签值包括“A and B”，则服务器可以查询标签值 A 所对应的目标位图 1 以及标签值 B 所对应的目标位图 2。

5 可选地，若预先设置有用于表示标签值所对应的位图是否常驻内存的标签配置信息，则在接收到数据查询请求之后，服务器可以根据目标标签值所对应的标签配置信息确定其对应的位图是否常驻内存，若常驻内存，则确定内存中的该目标标签值所对应的目标位图；而若不常驻内存，则从数据库中加载目标标签值对应的位图至内存中。

步骤 603，确定至少一个目标位图均包含的目标位图位所对应的目标承载体的标识。

10 若目标标签值包括一个，且数据查询请求用于查询具备该目标标签值的数据，则服务器根据目标位图确定所需查询的数据所对应的位图位，并且，由于每个位图位与一个承载体的标识一一对应，因此，可以根据上述对应关系确定查询到的位图位所对应的承载体的标识，查询到的承载体的标识即为目标承载体的标识。

15 比如，假设步骤 802 中确定的目标位图为“01100001”，则服务器可以根据该目标位图确定得到位图位为 2、3 和 8，进而确定得到位图位为 2 所对应的承载体的标识为标识 2、位图位为 3 所对应的承载体的标识为标识 3、位图位为 8 所对应的承载体的标识为标识 8。

若目标标签值包括一个，且数据查询请求用于查询不具备该目标标签值的数据，则服务器可以确定目标位图中用于表示不具备该目标标签值的索引所对应的数据。本与上述确定步骤类似，在此不再赘述。可选地，一种可能实现，服务器还可以对目标位图进行非运算，然后计算为 1 的位图位所对应的承载体的标识。

20 若目标标签值包括至少两个，则若数据查询请求用于查询同时具备至少两个目标标签值的数据，则在步骤 602 中服务器可以确定得到至少两个目标位图；此时在位图为二进制序列时，服务器可以将确定得到的至少两个目标位图进行与运算，确定与运算后取值为 1 的位图位，然后查询确定的位图位所对应的承载体的标识。比如，假设目标标签值有 A 和 B，且数据查询请求用于查询同时具备标签值 A 和标签值 B 的数据，服务器查询得到的标签值 A 所对应的位图为“01100001”，标签值 B 所对应的位图为“01001011”，则服务器进行与运算后得到“01000001”，此时服务器可以确定位图位为 2 和 8 所对应的承载体的标识。

30 若目标标签值包括至少两个，数据查询请求用于查询具备 j 个第一目标标签值或者具备 k 个第二目标标签值的数据，则在位图为二进制序列时，服务器可以将 j 个第一目标标签值所对应的 j 个目标位图进行与运算，将 k 个第二目标标签值所对应的 k 个目标位图进行与运算，然后将两者的与运算结果进行或运算，确定或运算后为“1”的位图位所对应的承载体的标识。其中，j 和 k 为大于等于 1 的整数。并且，在 j 或者 k 为 1 时，上述所说的与运算也可以不执行也即可以认为与运算的结果为其本身。

35 若目标标签值包括至少两个，数据查询请求用于查询具备 x 个第三目标标签值而不具备 y 个第四目标标签值的数据，则服务器将 x 个第三目标标签值所对应的 x 个目标位图进行与运算得到第一结果，将 y 个第四目标标签值所对应的 y 个目标位图进行与运算得到第二结果，然后再计算第一结果和第二结果的非运算，确定非运算后为 1 的位图位所对应的承载体的标识。其中，x 和 y 为大于等于 1 的整数。并且，在 x 或者 y 为 1 时，上述所说的与运算也可以不执行也即可以认为与运算的结果为其本身。

需要说明的是，在位图为数组时，其查询逻辑类似，在此不再赘述。

步骤 604, 查询数据表中目标载体具有的标签值。

数据表中记录已存储的数据的至少一个载体的标识和数据所具备的至少一个标签值的对应关系。可选地, 数据表的结构与上述实施例中所说的结构类似, 在此不再赘述。

在步骤 603 中确定得到目标载体的标识之后, 服务器即可根据数据表中记录的上述  
5 对应关系查询目标载体的标识所对应的至少一个标签值。

步骤 605, 反馈目标载体具有的标签值。

可选地, 服务器将查询到的标签值反馈至终端。

综上所述, 本实施例提供的数据查询方法, 通过在接收到数据查询请求之后, 直接位  
图索引即可确定得到所需查询的数据, 解决了相关技术中数据查询效率较低的问题; 达到  
10 了可以提高数据查询效率的效果。

同时, 在进行多标签查询时, 通过根据内存中的索引信息直接查询, 使得在多标签查  
询时可以达到 ms 级查询。

需要说明的是, 一种可能实现, 若位图索引被划分为多个位图索引分区, 如有 N 个位  
15 图索引分区, N 为大于等于 2 的整数, 则上述实施例中步骤 602 和步骤 603 会适应性的发  
生变化, 因此, 下述实施例将对位图索引分区存储时的数据查询方法做详细介绍。其  
中, 每个位图索引分区中包括至少一个子位图, 且每个位图索引分区中的子位图所对应  
的载体的标识的集合不存在交集, 这与上述数据存储中所涉及的位图索引分区类似, 在此  
不再赘述。

20 请参考图 7, 其示出了本申请一个实施例提供的数据查询方法的方法流程图, 本实施例  
以该数据查询方法用于图 2 所示的服务器中来举例说明, 如图 7 所示, 该数据查询方法包  
括:

步骤 701, 接收数据查询请求, 数据查询请求中携带有至一个目标标签值。

25 步骤 702, 在多个位图索引分区中, 分别确定每个目标标签值对应的目标子位图。

当位图索引在 N 个位图索引分区中分区存储时, 对于每个位图索引分区, 服务器可以  
查询在该位图索引分区中目标标签值所对应的目标子位图。比如, 假设位图索引在 4 个位  
图索引分区中存储, 则服务器在位图索引分区 1 中查询目标标签值所对应的目标子位图 1,  
在位图索引分区 2 中查询目标标签值所对应的目标子位图 2, 在位图索引分区 3 中查询目标  
30 标签值所对应的目标子位图 3, 在位图索引分区 4 中查询目标标签值所对应的目标子位图 4。

可选地, 当数据查询请求中携带的目标标签值有至少两个时, 对于每个位图索引分区,  
服务器可以查询在该位图索引分区中每个目标标签值所对应的目标子位图索引。比如, 假  
设数据查询请求中携带的目标标签值有目标标签值 1、目标标签值 2 和目标标签值 3, 则以  
在位图索引分区 1 中进行查询为例, 服务器可以在位图索引分区 1 中查询目标标签值 1 所  
35 对应的目标子位图 1、目标标签值 2 所对应的目标子位图 2、以及目标标签值 3 所对应的目  
标子位图 3。

步骤 703, 在多个位图索引分区中, 分别确定至少一个目标子位图均包含的目标位图位  
所对应的目标载体的标识。

在确定每个位图索引分区中的至少一个目标子位图之后, 服务器可以在每个位图索引

分区中根据确定的至少一个目标子位图确定目标位图位，并确定目标位图位对应的承载体的标识，在各个位图索引分区中确定的各个承载体的标识即为最终的目标承载体的标识。

其中，对于每个位图索引分区中确定目标位图位以及确定目标承载体的标识的具体确定方式与上述实施例在位图索引中确定目标位图位以及确定目标承载体的标识的确定方式类似，在此不再赘述。

步骤 704，查询数据表中目标载体具有的标签值。

步骤 705，反馈目标载体具有的标签值。

其中，步骤 704 和步骤 705 的具体实现与上述实施例中的步骤 604 和步骤 605 的具体实现类似，在此不再赘述。

综上所述，本实施例提供的的数据查询方法，通过在接收到数据查询请求之后，直接根据位图索引即可确定得到所需查询的数据，解决了相关技术中数据查询效率较低的问题；达到了可以提高数据查询效率的效果。

同时，在进行多标签查询时，通过根据内存中的索引信息直接查询，使得在多标签查询时可以达到 ms 级查询。

图 8 所示为本发明实施例提供的数据存储装置的示意图。数据存储装置 800 可以是计算机设备，该计算机设备可以是上述的服务器(例如图 2 所示的服务器 220)，数据存储装置 800 包括至少一个处理器 801，通信总线 802，存储器 803 以及至少一个通信接口 804。

处理器 801 可以是一个通用中央处理器 (CPU)，微处理器，特定应用集成电路 (application-specific integrated circuit, ASIC)，或一个或多个用于控制本发明方案程序执行的集成电路。

通信总线 802 可包括一通路，在上述组件之间传送信息。所述通信接口 804，使用任何收发器一类的装置，用于与其他设备或通信网络通信，如以太网，无线接入网 (RAN)，无线局域网(Wireless Local Area Networks, WLAN)等。

存储器 803 可以是只读存储器 (read-only memory, ROM) 或可存储静态信息和指令的其他类型的静态存储设备，随机存取存储器 (random access memory, RAM) 或者可存储信息和指令的其他类型的动态存储设备，也可以是电可擦可编程只读存储器 (Electrically Erasable Programmable Read-Only Memory, EEPROM)、只读光盘 (Compact Disc Read-Only Memory, CD-ROM) 或其他光盘存储、光碟存储 (包括压缩光碟、激光碟、光碟、数字通用光碟、蓝光光碟等)、磁盘存储介质或者其他磁存储设备、或者能够用于携带或存储具有指令或数据结构形式的期望的程序代码并能够由计算机存取的任何其他介质，但不限于此。存储器可以是独立存在，通过总线与处理器相连接。存储器也可以和处理器集成在一起。

其中，所述存储器 803 用于存储执行本发明方案的程序代码，并由处理器 801 来控制执行。所述处理器 801 用于执行所述存储器 803 中存储的程序代码。

在具体实现中，作为一种实施例，处理器 801 可以包括一个或多个 CPU，例如图 8 中的 CPU0 和 CPU1。

在具体实现中，作为一种实施例，数据存储装置 800 可以包括多个处理器，例如图 8 中的处理器 801 和处理器 808。这些处理器中的每一个可以是一个单核 (single-CPU) 处理器，也可以是一个多核 (multi-CPU) 处理器。这里的处理器可以指一个或多个设备、电路、

和/或用于处理数据（例如计算机程序指令）的处理核。

在具体实现中，作为一种实施例，数据存储装置 800 还可以包括输出设备 805 和输入设备 806。输出设备 805 和处理器 801 通信，可以以多种方式来显示信息。例如，输出设备 805 可以是液晶显示器（liquid crystal display, LCD）、发光二级管（light emitting diode, LED）显示设备，阴极射线管（cathode ray tube, CRT）显示设备，或投影仪（projector）等。输入设备 806 和处理器 801 通信，可以以多种方式接受用户的输入。例如，输入设备 806 可以是鼠标、键盘、触摸屏设备或传感设备等。

上述的数据存储装置 800 可以是一个通用计算机设备或者是一个专用计算机设备。在具体实现中，数据存储装置 800 可以是台式机、便携式电脑、网络服务器、掌上电脑（Personal Digital Assistant, PDA）、移动手机、平板电脑、无线终端设备、通信设备、嵌入式设备或有图 8 中类似结构的设备。本发明实施例不限定数据存储装置 800 的类型。

数据存储装置的存储器中存储了一个或多个软件模块。数据存储装置可以通过处理器以及存储器中的程序代码来实现软件模块，实现上述实施例所说的数据存储方法。

请参考图 9，其示出了本申请一个实施例提供的数据存储装置的结构示意图，如图 9 所示，该数据存储装置可以包括：获取单元 910 和建立单元 920；

其中，获取单元 910 用于执行上述实施例中的步骤 401；

建立单元 920 用于执行上述实施例中的步骤 402。

可选地，该数据存储装置还可以包括设置单元；

该设置单元，用于执行上述实施例中的步骤 301；

获取单元 910，还用于执行上述实施例中的步骤 302。

可选地，所述标签定义表还包括标签配置信息，所述标签配置信息包括用于表示每个标签值所对应的位图是否常驻内存的信息；该数据存储装置还可以包括加载单元；

该加载单元，用于根据所述标签配置信息，将需要常驻内存的位图加载至内存。

可选地，该装置还包括划分单元；

该划分单元，用于将所述位图索引划分为多个位图索引分区，每个位图索引分区中包括至少一个子位图，每个位图索引分区对应于一个承载体标识的集合，不同位图索引分区所对应的承载体标识的集合不存在交集。

可选地，该装置还包括划分单元和确定单元；

划分单元，用于根据承载体的标识的第一范围将所述数据表划分为 M 个数据分区，每个数据分区包括至少一个子数据表，每个数据分区所对应的承载体标识的集合为所述第一范围，且不同数据分区所对应的承载体标识的集合不存在交集；

确定单元，用于根据所述 M 个数据分区确定 N 个位图索引分区，每个位图索引分区中包括至少一个子位图，每个位图索引分区所对应的承载体的标识的范围大于等于所述第一范围，且不同位图索引分区所对应的承载体的标识的集合不存在交集，M、N 为正整数，N 小于或等于 M，N 大于或等于 2。

可选地，所述每个位图索引分区中的子位图的数量为预先定义的标签定义表中设置的全部标签值的数量。

可选地，所述位图索引中的不同位图中相同的位图位对应于相同的承载体的标识；每

个位图索引分区中的不同子位图的相同的位图位对应于相同的承载体的标识，不同位图索引分区中的不同子位图的相同的位图位对应于不相同的承载体的标识。

可选地，该装置还包括：接收单元、确定单元、查询单元和反馈单元；

接收单元，用于执行上述实施例中的步骤 601 或者步骤 701；

5 确定单元，用于执行上述实施例中的步骤 602 和步骤 703，或者，执行步骤 602 和步骤 703；

查询单元，用于执行上述实施例中的步骤 604 或者步骤 704；

反馈单元，用于执行上述实施例中的步骤 605 或者步骤 705。

10 可选地，获取单元 910，还用于获取新数据，其中，在所述新数据包括的新对应关系中承载体的标识和/或标签值是新的；

该装置还包括更新单元，该更新单元，用于根据所述新对应关系更新所述位图索引。

可选地，更新单元，还用于确定所述新对应关系中的承载体的标识所属的位图索引分区；在确定的位图索引分区中，更新所述新对应关系中的标签值，和/或更新所述新对应关系中的标签值所对应的位图。

15 综上所述，本实施例提供的数据存储装置，通过在获取到至少一条数据记录之后，建立该至少一条数据记录对应的位图索引，且位图索引中包括标签值和其对应的位图的对应关系，位图中包括用于记录载体标识所对应的载体是否具备对应的标签值的至少一个位图位；使得在后续基于标签值查询时，可以根据位图索引直接确定对应的数据，提高了数据查询效率。

20

本申请一个实施例还提供了一种计算机存储介质，该计算机存储介质中存储有指令；数据存储装置（可以是计算机设备，例如服务器）执行该指令，例如数据存储装置中的处理器执行该指令，使得该数据存储装置实现上述实施例所说的数据存储方法。

25 本申请实施例提供一种计算机程序产品，该计算机程序产品包括指令；数据存储装置（可以是计算机设备，例如服务器）执行该指令，使得该数据存储装置执行上述方法实施例的数据存储方法。

30 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，可轻易想到变化或替换，都应涵盖在本申请的保护范围之内。因此，本申请的保护范围应所述以权利要求的保护范围为准。

## 权 利 要 求 书

1、一种数据存储方法，其特征在于，所述方法包括：

5 获取至少一条数据记录，每条数据记录包括一个承载体标识和至少一个标签值，所述至少一条数据记录存储于数据表中，所述数据表用于记录承载体标识与标签值的对应关系；

建立所述至少一条数据记录对应的位图索引，所述位图索引包括至少一个位图，每个位图对应于一个标签值，每个位图包括至少一个位图位，每个位图位用于记录一个承载体标识所对应的承载体是否具备当前位图所对应的标签值。

10 2、根据权利要求1所述的方法，其特征在于，所述获取至少一条数据记录之前，所述方法包括：

预先设置标签定义表，所述预先设置的标签定义表包括预设的多个标签值；

所述获取至少一条数据记录包括：

根据待存储的源数据和所述预先设置的标签定义表，生成所述数据表。

15 3、根据权利要求2所述的方法，其特征在于，所述预先设置的标签定义表还包括标签配置信息，所述标签配置信息包括表示每个标签值所对应的位图是否常驻内存的信息；所述方法还包括：

根据所述标签配置信息，将需要常驻内存的位图加载至内存。

4、根据权利要求1至3任一项所述的方法，其特征在于，所述方法还包括：

20 将所述位图索引划分为多个位图索引分区，每个位图索引分区中包括至少一个子位图，每个位图索引分区对应于一个承载体标识的集合，不同位图索引分区所对应的承载体标识的集合不存在交集。

5、根据权利要求1至3任一项所述的方法，其特征在于，所述方法还包括：

25 根据承载体标识的第一范围将所述数据表划分为M个数据分区，每个数据分区包括至少一个子数据表，每个数据分区所对应的承载体标识的集合为所述第一范围，且不同数据分区所对应的承载体标识的集合不存在交集；

根据所述M个数据分区确定N个位图索引分区，每个位图索引分区中包括至少一个子位图，每个位图索引分区对应的承载体标识的范围大于等于所述第一范围，且不同位图索引分区所对应的承载体标识的集合不存在交集，M、N为正整数，N小于或等于M，N大于或等于2。

30 6、根据权利要求4或5所述的方法，其特征在于，所述每个位图索引分区中的子位图的数量为预先定义的标签定义表中设置的全部标签值的数量。

7、根据权利要求4或5或6所述的方法，其特征在于，所述位图索引中的不同位图中相同的位图位对应于相同的承载体的标识；每个位图索引分区中的不同子位图的相同的位图位对应于相同的承载体的标识，不同位图索引分区中的不同子位图的相同的位图位对应于不相同的承载体的标识。

8、根据权利要求4至7任一项所述的方法，其特征在于，

接收数据查询请求，所述数据查询请求携带有至少一个目标标签值；

在所述多个位图索引分区中，分别确定所述至少一个目标标签值对应的至少一个目标子位图，分别确定所述至少一个目标子位图均包含的目标位图位所对应的目标承载体的标识；

查询所述数据表中所述目标载体具有的标签值；

反馈所述目标载体具有的标签值。

9、根据权利要求 1 至 8 任一项所述的方法，其特征在于，所述方法包括：

获取新数据，其中，在所述新数据包括的新对应关系中承载体的标识和/或标签值是新的；  
5 根据所述新对应关系更新所述位图索引。

10、根据权利要求 9 所述的方法，其特征在于，所述根据所述新对应关系更新位图索引，  
包括：

确定所述新对应关系中的承载体的标识所属的位图索引分区；

10 在确定的位图索引分区中，更新所述新对应关系中的标签值，和/或更新所述新对应关系  
中的标签值所对应的位图。

11、一种数据存储装置，其特征在于，所述装置包括：

获取单元，用于获取至少一条数据记录，每条数据记录包括一个载体标识和至少一个  
标签值，所述至少一条数据记录存储于数据表中，所述数据表用于记录载体标识与标签值  
的对应关系；

15 建立单元，用于建立所述至少一条数据记录对应的位图索引，所述位图索引包括至少一  
个位图，每个位图对应于一个标签值，每个位图包括至少一个位图位，每个位图位用于记录  
一个载体标识所对应的载体是否具备当前位图所对应的标签值。

12、根据权利要求 11 所述的装置，其特征在于，所述装置包括：

20 设置单元，用于在所述获取单元获取所述至少一条数据记录之前，预先设置标签定义表，  
所述预先设置的标签定义表包括预设的多个标签值；

所述获取单元，还用于根据待存储的源数据和所述预先设置的标签定义表，生成所述数  
据表。

25 13、根据权利要求 12 所述的装置，其特征在于，所述预先设置的标签定义表还包括标签  
配置信息，所述标签配置信息包括表示每个标签值所对应的位图是否常驻内存的信息；所述  
装置还包括：

加载单元，用于根据所述标签配置信息，将需要常驻内存的位图加载至内存。

14、根据权利要求 11 至 13 任一项所述的装置，其特征在于，所述装置还包括：

30 划分单元，用于将所述位图索引划分为多个位图索引分区，每个位图索引分区中包括至  
少一个子位图，每个位图索引分区对应于一个载体标识的集合，不同位图索引分区所对应  
的载体标识的集合不存在交集。

15、根据权利要求 11 至 13 任一项所述的装置，其特征在于，所述装置还包括：

划分单元，用于根据载体标识的第一范围将所述数据表划分为 M 个数据分区，每个数  
据分区包括至少一个子数据表，每个数据分区所对应的载体标识的集合为所述第一范围，  
且不同数据分区所对应的载体标识的集合不存在交集；

35 确定单元，用于根据所述 M 个数据分区确定 N 个位图索引分区，每个位图索引分区中包  
括至少一个子位图，每个位图索引分区对应的载体标识的范围大于等于所述第一范围，且  
不同位图索引分区所对应的载体标识的集合不存在交集，M、N 为正整数，N 小于或等于  
M，N 大于或等于 2。

16、根据权利要求 14 或 15 所述的装置，其特征在于，所述每个位图索引分区中的子位

图的数量为预先定义的标签定义表中设置的全部标签值的数量。

17、根据权利要求 14 或 15 或 16 所述的装置，其特征在于，所述位图索引中的不同位图中相同的位图位对应于相同的承载体的标识；每个位图索引分区中的不同子位图的相同的位图位对应于相同的承载体的标识，不同位图索引分区中的不同子位图的相同的位图位对应于不相同的承载体的标识。

18、根据权利要求 14 至 17 任一项所述的装置，其特征在于，所述装置还包括：  
接收单元，用于接收数据查询请求，所述数据查询请求携带有至少一个目标标签值；  
确定单元，用于在所述多个位图索引分区中，分别确定所述至少一个目标标签值对应的至少一个目标子位图，分别确定所述至少一个目标子位图均包含的目标位图位所对应的目标承载体的标识；  
查询单元，用于查询所述数据表中所述目标载体具有的标签值；  
反馈单元，用于反馈所述目标载体具有的标签值。

19、根据权利要求 11 至 18 任一项所述的装置，其特征在于，  
所述获取单元，还用于获取新数据，其中，在所述新数据包括的新对应关系中承载体的标识和/或标签值是新的；

所述装置包括：更新单元；  
所述更新单元，用于根据所述新对应关系更新所述位图索引。

20、根据权利要求 19 所述的装置，其特征在于，所述更新单元，还用于：  
确定所述新对应关系中的承载体的标识所属的位图索引分区；  
在确定的位图索引分区中，更新所述新对应关系中的标签值，和/或更新所述新对应关系中的标签值所对应的位图。

21、一种数据存储装置，其特征在于，所述装置包括：存储器和处理器，所述存储器中存储有指令，所述处理器通过执行所述存储器中存储的指令使得所述数据存储装置实现如权利要求 1 至 10 任一所述的数据存储方法。

22、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中存储有指令，数据存储装置执行所述指令使得所述数据存储装置实现权利要求 1 至 10 任一所述的数据存储方法。

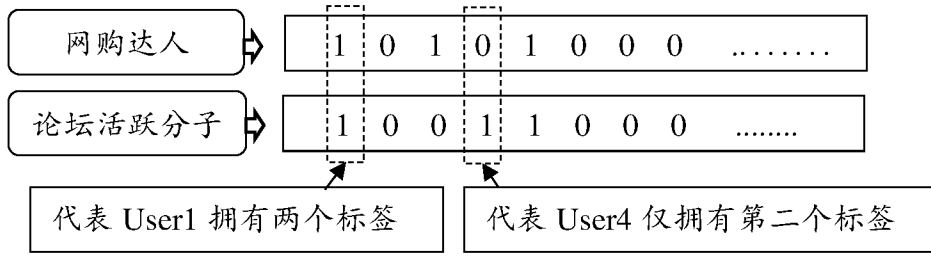


图 1

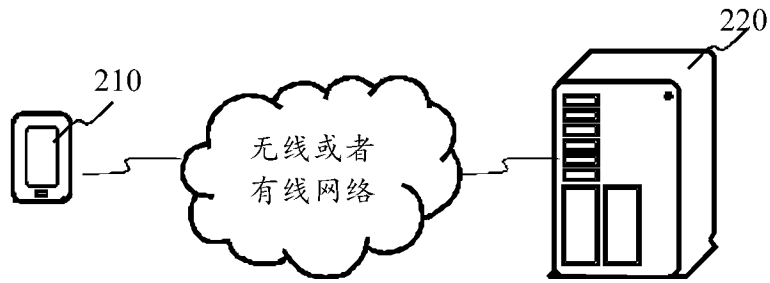


图 2

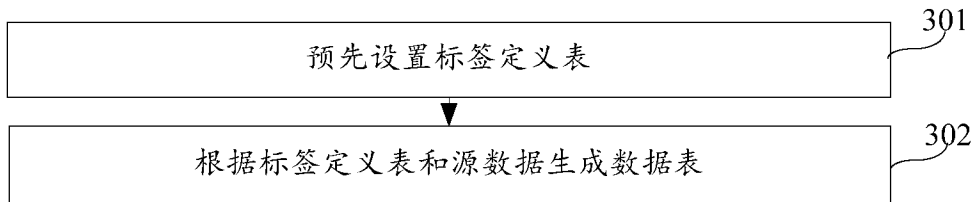


图 3

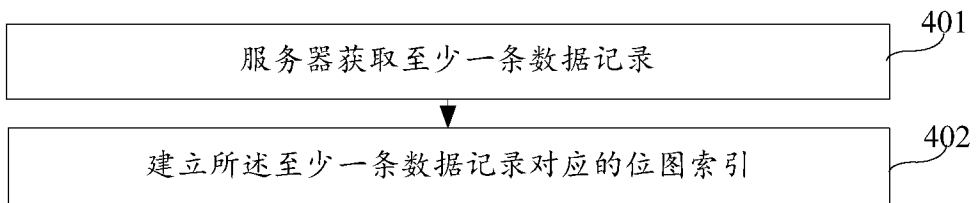


图 4

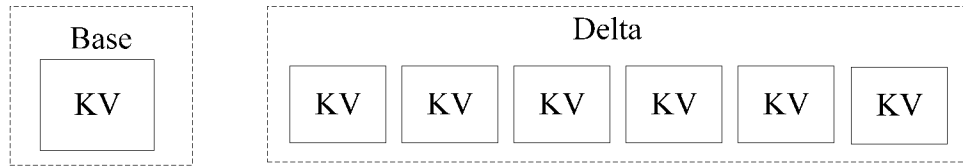


图 5

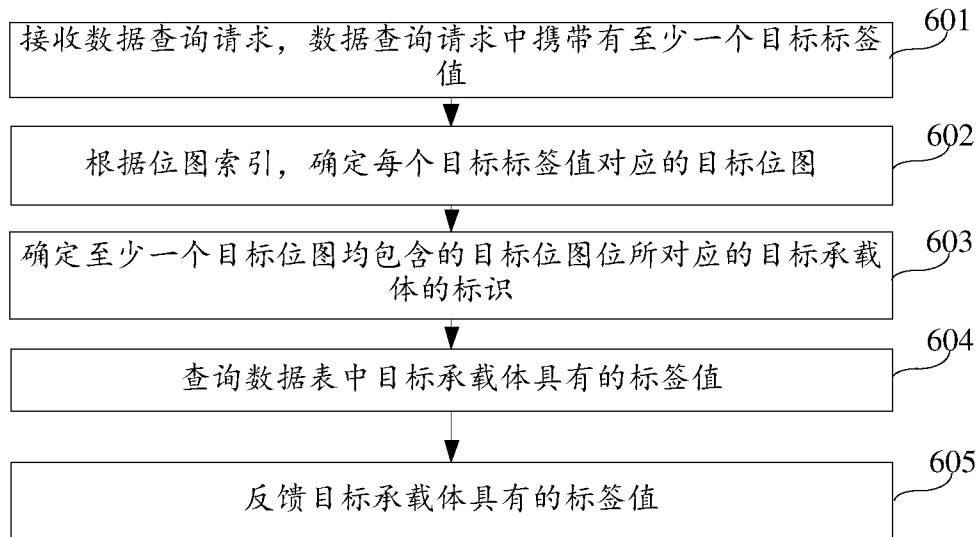


图 6

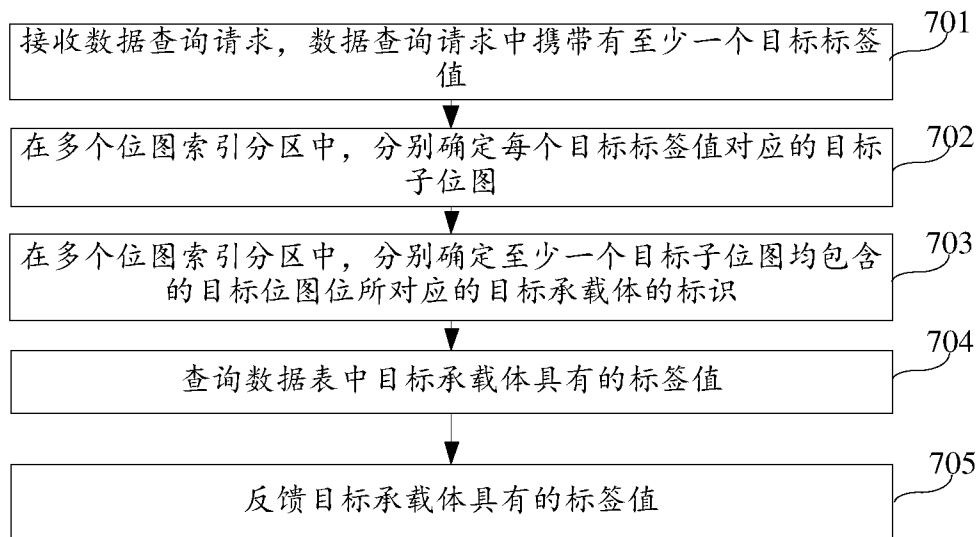


图 7

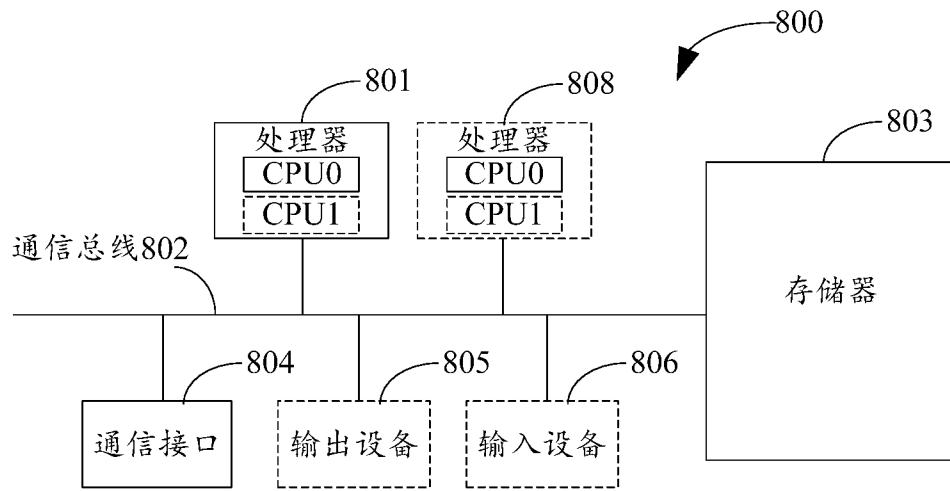


图 8

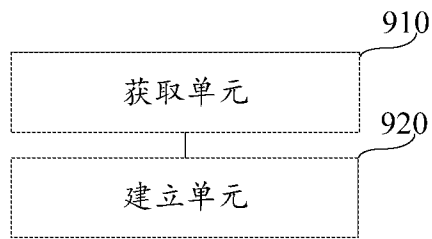


图 9

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/CN2017/095893

## 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)

WPI, EPODOC, CNPAT, CNKI: 标签, 索引, 位图, 数据, 存储, 查询, 标识, 映射, 效率, tag, label, data, stor+, index, query, search+, map+, bitmap, efficien+

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 105630972 A (NETEASE (HANGZHOU) NETWORK CO., LTD.), 01 June 2016 (01.06.2016), description, paragraphs [0040]-[0100], and figure 2	1-3, 9-13, 19-22
A	CN 102254012 A (PEKING UNIVERSITY), 23 November 2011 (23.11.2011), entire document	1-22
A	CN 102110146 A (TSINGHUA UNIVERSITY), 29 June 2011 (29.06.2011), entire document	1-22
A	CN 102779180 A (HUAWEI TECHNOLOGIES CO., LTD.), 14 November 2012 (14.11.2012), entire document	1-22

Further documents are listed in the continuation of Box C.

See patent family annex.

<p>* Special categories of cited documents:</p> <p>“A” document defining the general state of the art which is not considered to be of particular relevance</p> <p>“E” earlier application or patent but published on or after the international filing date</p> <p>“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)</p> <p>“O” document referring to an oral disclosure, use, exhibition or other means</p> <p>“P” document published prior to the international filing date but later than the priority date claimed</p>	<p>“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</p> <p>“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</p> <p>“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</p> <p>“&amp;” document member of the same patent family</p>
---	---

Date of the actual completion of the international search 27 March 2018	Date of mailing of the international search report 07 May 2018
Name and mailing address of the ISA 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  TIAN, Jing  Telephone No. (86-10) 86-(10)-53961334

**INTERNATIONAL SEARCH REPORT**  
Information on patent family members

International application No.  
PCT/CN2017/095893

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 105630972 A	01 June 2016	None	
CN 102254012 A	23 November 2011	None	
CN 102110146 A	29 June 2011	None	
CN 102779180 A	14 November 2012	None	

国际检索报告

国际申请号

PCT/CN2017/095893

<p><b>A. 主题的分类</b></p> <p>G06F 17/30(2006.01)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>B. 检索领域</b></p> <p>检索的最低限度文献(标明分类系统和分类号)</p> <p>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))</p> <p>WPI, EPODOC, CNPAT, CNKI: 标签, 索引, 位图, 数据, 存储, 查询, 标识, 映射, 效率, tag, label, data, stor+, index, query, search+, map+, bitmap, efficien+</p>																											
<p><b>C. 相关文件</b></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>X</td> <td>CN 105630972 A (网易杭州网络有限公司) 2016年 6月 1日 (2016 - 06 - 01) 说明书第[0040]-[0100]段, 图2</td> <td>1-3, 9-13, 19-22</td> </tr> <tr> <td>A</td> <td>CN 102254012 A (北京大学) 2011年 11月 23日 (2011 - 11 - 23) 全文</td> <td>1-22</td> </tr> <tr> <td>A</td> <td>CN 102110146 A (清华大学) 2011年 6月 29日 (2011 - 06 - 29) 全文</td> <td>1-22</td> </tr> <tr> <td>A</td> <td>CN 102779180 A (华为技术有限公司) 2012年 11月 14日 (2012 - 11 - 14) 全文</td> <td>1-22</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          “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          “&amp;” 同族专利的文件</p> <table border="1"> <tr> <td>国际检索实际完成的日期</td> <td>国际检索报告邮寄日期</td> </tr> <tr> <td>2018年 3月 27日</td> <td>2018年 5月 7日</td> </tr> <tr> <td>ISA/CN的名称和邮寄地址</td> <td>受权官员</td> </tr> <tr> <td>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</td> <td>田晶</td> </tr> <tr> <td>传真号 (86-10)62019451</td> <td>电话号码 86-(10)-53961334</td> </tr> </table>			类型*	引用文件, 必要时, 指明相关段落	相关的权利要求	X	CN 105630972 A (网易杭州网络有限公司) 2016年 6月 1日 (2016 - 06 - 01) 说明书第[0040]-[0100]段, 图2	1-3, 9-13, 19-22	A	CN 102254012 A (北京大学) 2011年 11月 23日 (2011 - 11 - 23) 全文	1-22	A	CN 102110146 A (清华大学) 2011年 6月 29日 (2011 - 06 - 29) 全文	1-22	A	CN 102779180 A (华为技术有限公司) 2012年 11月 14日 (2012 - 11 - 14) 全文	1-22	国际检索实际完成的日期	国际检索报告邮寄日期	2018年 3月 27日	2018年 5月 7日	ISA/CN的名称和邮寄地址	受权官员	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	田晶	传真号 (86-10)62019451	电话号码 86-(10)-53961334
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求																									
X	CN 105630972 A (网易杭州网络有限公司) 2016年 6月 1日 (2016 - 06 - 01) 说明书第[0040]-[0100]段, 图2	1-3, 9-13, 19-22																									
A	CN 102254012 A (北京大学) 2011年 11月 23日 (2011 - 11 - 23) 全文	1-22																									
A	CN 102110146 A (清华大学) 2011年 6月 29日 (2011 - 06 - 29) 全文	1-22																									
A	CN 102779180 A (华为技术有限公司) 2012年 11月 14日 (2012 - 11 - 14) 全文	1-22																									
国际检索实际完成的日期	国际检索报告邮寄日期																										
2018年 3月 27日	2018年 5月 7日																										
ISA/CN的名称和邮寄地址	受权官员																										
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	田晶																										
传真号 (86-10)62019451	电话号码 86-(10)-53961334																										

国际检索报告  
关于同族专利的信息

国际申请号

PCT/CN2017/095893

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	105630972	A	2016年 6月 1日	无	
CN	102254012	A	2011年 11月 23日	无	
CN	102110146	A	2011年 6月 29日	无	
CN	102779180	A	2012年 11月 14日	无	