

(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
2020 年 6 月 25 日 (25.06.2020)



(10) 国际公布号
WO 2020/124491 A1

(51) 国际专利分类号:
G06F 16/00 (2019.01)

(21) 国际申请号: PCT/CN2018/122380

(22) 国际申请日: 2018 年 12 月 20 日 (20.12.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201811559134.1 2018年12月19日 (19.12.2018) CN

(71) 申请人: 深圳市优必选科技有限公司(**UBTECH ROBOTICS CORP**) [CN/CN]; 中国广东省深圳市南山区学苑大道1001号南山智园C1栋16、22楼, Guangdong 518000 (CN)。

(72) 发明人: 熊友军(**XIONG, Youjun**); 中国广东省深圳市南山区学苑大道1001号南山智园C1栋16、22

楼, Guangdong 518000 (CN)。方曦(**FANG, Xi**); 中国广东省深圳市南山区学苑大道1001号南山智园C1栋16、22楼, Guangdong 518000 (CN)。

(74) 代理人: 深圳中细软知识产权代理有限公司(**SHENZHEN CIPRUN INTELLECTUAL PROPERTY AGENCY CO., LTD.**); 中国广东省深圳市南山区粤海街道粤兴四道1号中山大学深圳产学研大楼11楼, 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,

(54) **Title:** METHOD AND DEVICE FOR SEGMENTING DATA, COMPUTER DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 数据的切分方法、装置、计算机设备及存储介质

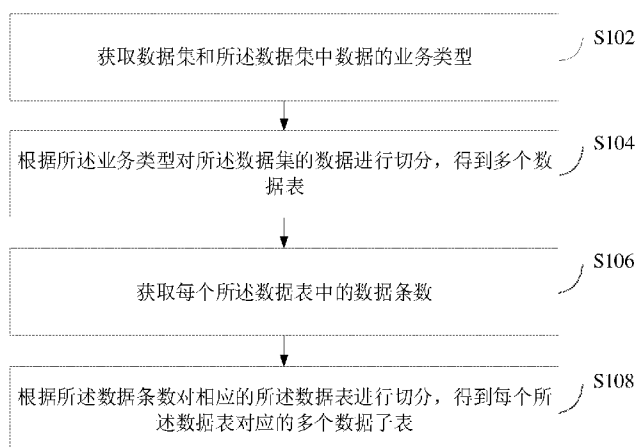


图 1

S102 Acquire a dataset and service types of data in the dataset
S104 Segment the data of the dataset on the basis of the service types to produce multiple data tables
S106 Acquire the number of data entries in each of the data tables
S108 Segment the data tables on the basis of the corresponding number of data entries to produce multiple data child tables corresponding to each of the data tables

(57) **Abstract:** Disclosed are a method and device for segmenting data, a computer device, and a storage medium, comprising: acquiring a dataset and service types of data in the dataset; segmenting the data of the dataset on the basis of the service types to produce multiple data tables; acquiring the number of data entries in each of the data tables; segmenting the data tables on the basis of the corresponding number of data entries to produce multiple data child tables corresponding to each of the data tables. The segmentation of data per the method increases the efficiency of data query when data query is required.



WO 2020/124491 A1

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。

- (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: 本发明实施例公开了一种数据的切分方法、装置、计算机设备及存储介质, 包括: 获取数据集和
所述数据集中数据的业务类型; 根据所述业务类型对所述数据集的数据进行切分, 得到多个数据表; 获取每
个所述数据表中的数据条数; 根据所述数据条数对相应的所述数据表进行切分, 得到每个所述数据表对应的
多个数据子表。通过上述方式对数据进行切分, 能够在需要进行数据查询时, 提高数据查询效率。

数据的切分方法、装置、计算机设备及存储介质

技术领域

[0001] 本发明涉及数据库技术领域，尤其涉及一种数据的切分方法、装置、计算机设备及存储介质。

背景技术

[0002] 随着云时代的来临，大数据也吸引了相关领域人员越来越多的关注。大数据是大量数据的集合，通过大数据，可以对用户行为、产品价格趋势等进行预测。

发明概述

技术问题

[0003] 随着大数据技术的不断发展，服务器存储的数据量越来越大，在对千万乃至上亿级以上的数据量进行查询的时候，由于服务器中存储的数据量巨大，导致服务器数据库由于查询对象过多导致无法及时响应。

问题的解决方案

技术解决方案

[0004] 基于此，有必要针对上述问题，提出一种提高查询效率的数据的切分方法、装置、计算机设备及存储介质。

[0005] 一种数据的切分方法，所述方法包括：

[0006] 获取数据集和所述数据集中数据的业务类型；

[0007] 根据所述业务类型对所述数据集的数据进行切分，得到多个数据表；

[0008] 获取每个所述数据表中的数据条数；

[0009] 根据所述数据条数对相应的所述数据表进行切分，得到每个所述数据表对应的多个数据子表。

[0010] 在其中一个实施例中，在所述根据所述业务类型对所述数据集的数据进行切分，得到多个数据表之后，还包括：根据所述数据表的个数，确定数据库的个数；将所述多个数据表存放于确定的多个数据库中。

[0011] 在其中一个实施例中，所述根据所述数据条数对相应的所述数据表进行切分，

得到每个所述数据表对应的多个数据子表，包括：获取每个所述数据表中的每条数据对应的占用空间；根据所述数据表中的数据条数和所述数据表中的每条数据对应的占用空间确定所述数据表对应的数据子表的个数；根据所述数据子表的个数对所述数据表进行切分，得到每个所述数据表对应的多个数据子表。

[0012] 在其中一个实施例中，在所述根据所述数据子表的个数对所述数据表进行切分，得到每个所述数据表对应的多个数据子表之后，还包括：根据所述数据子表的个数，确定数据库的个数；将所述多个数据子表存放于确定的多个数据库中。

[0013] 在其中一个实施例中，所述根据所述数据条数对相应的所述数据表进行切分，得到每个所述数据表对应的多个数据子表，包括：获取每个所述数据表中每条数据对应的字段数字标识；对所述字段数字标识进行求模运算，得到求模结果；根据所述求模结果和所述数据条数对相应的数据表进行切分，得到每个所述数据表对应的多个数据子表。

[0014] 在其中一个实施例中，在所述根据所述数据条数对相应的所述数据表进行切分，得到每个所述数据表对应的多个数据子表之后，还包括：将所述业务类型和所述数据表的对应关系、所述数据表与相应的所述数据子表的对应关系以及所述数据子表中的每条数据对应的数据标识发送到数据表管理中心，以使所述数据表管理中心根据接收到的数据处理请求中的业务类型和数据标识从相应的数据子表中查找与所述数据标识对应的数据。

[0015] 一种数据的切分装置，所述装置包括：

[0016] 第一获取模块，用于获取数据集和所述数据集中数据的业务类型；

[0017] 第一切分模块，用于根据所述业务类型对所述数据集的数据进行切分，得到多个数据表；

[0018] 第二获取模块，用于获取每个所述数据表中的数据条数；

[0019] 第二切分模块，用于根据所述数据条数对相应的所述数据表进行切分，得到每个所述数据表对应的多个数据子表。

[0020] 在其中一个实施例中，所述数据的切分装置，还包括：第一个数确定模块，用于根据所述数据表的个数，确定数据库的个数；第一数据库存放模块，用于将

所述多个数据表存放于确定的多个数据库中。

[0021] 在其中一个实施例中，所述第二切分模块，包括：占用空间获取模块，用于获取每个所述数据表中的每条数据对应的占用空间；占用空间确定模块，用于根据所述数据表中的数据条数和所述数据表中的每条数据对应的占用空间确定所述数据表对应的数据子表的个数；占用空间切分模块，用于根据所述数据子表的个数对所述数据表进行切分，得到每个所述数据表对应的多个数据子表。

[0022] 在其中一个实施例中，所述数据的切分装置，还包括：第二个数确定模块，用于根据所述数据子表的个数，确定数据库的个数；第二数据库存放模块，用于将所述多个数据子表存放于确定的多个数据库中。

[0023] 在其中一个实施例中，所述第二切分模块，还包括：字段数字标识获取模块，用于获取每个所述数据表中每条数据对应的字段数字标识；求模模块，用于对所述字段数字标识进行求模运算，得到求模结果；求模切分模块，用于根据所述求模结果和所述数据条数对相应的数据表进行切分，得到每个所述数据表对应的多个数据子表。

[0024] 一种计算机设备，包括存储器和处理器，所述存储器存储有计算机程序，所述计算机程序被所述处理器执行时，使得所述处理器执行以下步骤：

[0025] 获取数据集和所述数据集中数据的业务类型；

[0026] 根据所述业务类型对所述数据集的数据进行切分，得到多个数据表；

[0027] 获取每个所述数据表中的数据条数；

[0028] 根据所述数据条数对相应的所述数据表进行切分，得到每个所述数据表对应的多个数据子表。

[0029] 一种计算机可读存储介质，存储有计算机程序，所述计算机程序被处理器执行时，使得所述处理器执行以下步骤：

[0030] 获取数据集和所述数据集中数据的业务类型；

[0031] 根据所述业务类型对所述数据集的数据进行切分，得到多个数据表；

[0032] 获取每个所述数据表中的数据条数；

[0033] 根据所述数据条数对相应的所述数据表进行切分，得到每个所述数据表对应的多个数据子表。

发明的有益效果

有益效果

[0034] 实施本发明实施例，将具有如下效果：

[0035] 本发明提出了一种数据的切分方法、装置、设备及存储介质，首先获取数据集和所述数据集中数据的业务类型；并且根据所述业务类型对所述数据集的数据进行切分，得到多个数据表；然后获取每个所述数据表中的数据条数；最后根据所述数据条数对相应的所述数据表进行切分，得到每个所述数据表对应的多个数据子表。可见，通过上述方式，由于获取到了多个数据子表对原始的数据集进行存放，而不是将原始的数据集存放于一个数据表中，所以在后续要查数据的时候，可以直接在某一个数据子表中查询，而不是在一个大的存储原始数据集的数据表中进行查询，由于查询量减少，所以能够提高查询效率，提高响应效率，并且，上述数据的切分方式将数据首先按业务切分，然后再对经过业务切分的表进行再次切分，使得最终得到的数据子表相对只有一次切分得到的数据子表更小，进一步的提高了查询效率和响应效率。

对附图的简要说明

附图说明

[0036] 为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本发明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

[0037] 其中：

[0038] 图1为一个实施例中数据的切分方法的实现流程示意图；

[0039] 图2为一个实施例中一条数据的示意图；

[0040] 图3为一个实施例中一条数据的示意图；

[0041] 图4为一个实施例中一条数据的示意图；

[0042] 图5为一个实施例中数据的切分方法的实现流程示意图；

[0043] 图6为一个实施例中数据的切分方法对应的切分数据的过程示意图；

[0044] 图7为一个实施例中数据的切分方法的实现流程示意图；

[0045] 图8为一个实施例中数据表管理中心根据业务类型和数据标识确定数据子表的示意图；

[0046] 图9为一个实施例中数据的切分装置的结构框图；

[0047] 图10为一个实施例中计算机设备的结构框图。

发明实施例

本发明的实施方式

[0048] 下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

[0049] 如图1所示，在一个实施例中，提供了一种数据的切分方法。该方法可以应用于服务器，也可以应用于终端。所述服务器为高性能计算机或由多个高性能计算机组成的高性能计算机集群，所述终端，可以是台式终端，也可以是移动终端，例如，台式终端为台式电脑，移动终端为笔记本电脑。

[0050] 如图1所示，本发明实施例所述的数据的切分方法具体包括如下步骤：

[0051] 步骤102，获取数据集和所述数据集中数据的业务类型。

[0052] 所述数据集，为多个数据的集合。

[0053] 所述数据的业务类型，指示不同类型的数据，例如，对某一数据集，其中的数据的业务类型可能有订单业务类型和用户业务类型。其中，订单业务类型的数据表可能包含的数据有用户字段类型的数据（用户信息数据）、订单金额字段类型的数据（订单金额数据）、下单时间字段类型的数据（下单时间数据）、优惠券字段类型的数据（优惠券数据）和购买日期类型的数据（购买日期数据）；用户业务类型的数据表可能包含的数据有用户字段类型的数据（用户信息数据）、用户性别字段类型的数据（用户性别数据）、用户年龄字段类型的数据（用户年龄数据）和用户送货地址字段类型的数据（用户送货地址数据）。

[0054] 步骤S104，根据所述业务类型对所述数据集的数据进行切分，得到多个数据表。

[0055] 数据集的第一次切分主要是根据数据的业务类型对数据进行的切分，一个业务

类型对应一个数据表，例如，订单类型对应订单数据表，用户类型对应用户数据表，地址类型对应地址数据表，具体的，将属于订单业务类型的数据存入订单数据表，即将用户信息数据、订单金额数据、下单时间数据、优惠券数据和购买日期数据存入订单数据表；将属于用户业务类型的数据存入用户数据表，即将用户信息数据、用户性别数据、用户年龄数据和用户送货地址数据存入用户数据表。

[0056] 作为本发明的一种实施例，在步骤104所述根据所述业务类型对所述数据集的数据进行切分，得到多个数据表之后，还包括：根据所述数据表的个数，确定数据库的个数；将所述多个数据表存放于确定的多个数据库中。

[0057] 例如，若经过切分得到3个数据表，则将3个数据表分别放入3个数据库中；或者，经过切分得到3个数据表，将其中两个数据表放入数据库1，将第三个数据表放入数据库2。

[0058] 步骤S106，获取每个所述数据表中的数据条数。

[0059] 所述数据条数，指代数据表中的一行数据或者一列数据，具体的，如图2所示的订单数据表，每一行的数据代表一条数据，或者，如图3所示的订单数据表，每一列的数据代表一条数据。

[0060] 需要说明的是，数据表中包括多个字段的数据，可以将其中的某一个字段设置为主字段，通过设置主字段，能够将数据表中的数据较好的划分为多条数据。例如如图2或图3，字段为：用户字段、订单金额字段、下单时间字段、优惠券字段和商品类型字段，为了更好的划分数据，将主字段设置为用户，如图2所示，通过将用户字段设置为主字段，能够将订单数据表中的数据划分为3条数据。当然，也可以将其他字段设置为主字段，例如图4所示，将下单时间设置为主字段，当然，具体要将什么字段设置为主字段，可以根据实际需求进行确定，比如订单表，显然将用户字段设置为主字段更好，通过将用户字段设置为主字段，能够更好的对单个用户的数据进行分析，当然，如果将下单时间设置为主字段，能够对各个时间段的购买情况进行更好的分析。

[0061] 步骤108，根据所述数据条数对相应的所述数据表进行切分，得到每个所述数据表对应的多个数据子表。

[0062] 在这里，根据所述数据条数可以采用多种方式对相应的数据表进行切分，得到每个所述数据表对应的多个数据子表。

[0063] 可选的，根据预置的切分比例对数据表进行切分，具体的，所述根据所述数据条数对相应的所述数据表进行切分，得到每个所述数据表对应的多个数据子表，包括：获取预置的切分比例；根据预置的切分比例和所述数据条数对相应的所述数据表进行切分，得到每个所述数据表对应的多个数据子表。

[0064] 若预置的切分比例为将每个数据表均匀的切分为3个数据子表，则此时可以根据数据表中的数据条数，例如60条数据，将数据表均匀的切分成3个数据子表，每个数据子表中有20条数据；若预置的切分比例为将每个数据表按1/6，1/3和1/2进行切分，则此时，可以根据数据表中的数据条数，例如60条数据，将数据表切分为包含10条数据的数据子表1，包含20条数据的数据子表2和包含30条数据的数据子表3。

[0065] 可选的，根据占用空间对数据表进行切分，具体的，所述根据所述数据条数对相应的所述数据表进行切分，得到每个所述数据表对应的多个数据子表，包括：获取每个所述数据表中的每条数据对应的占用空间；根据所述数据表中的数据条数和所述数据表中的每条数据对应的占用空间确定所述数据表对应的数据子表的个数；根据所述数据子表的个数对所述数据表进行切分，得到每个所述数据表对应的多个数据子表。

[0066] 如图2所示，可以采用如下方法计算每条数据的占用空间，假设一条数据中的一个数据占用空间是1kb，那么由于一条数据包含有4个数据，所以一条数据的占用空间是4kb。

[0067] 所述根据所述数据表中的数据条数和所述数据表中的每条数据对应的占用空间确定所述数据表对应的数据子表的个数，具体包括：一、根据数据条数和每条数据对应的占用空间确定数据表的占用空间，根据数据表的占用空间确定数据子表的个数；二、根据每条数据的占用空间，确定每个数据子表存放的数据条数，根据每个数据子表存放的数据条数和所述数据表的数据条数，确定数据子表的个数。

[0068] 对于第一种情况，例如，每条数据的占用空间为0.4M，数据条数为1000条，那

么总共的占用空间是400M，由于该空间较大，可以用两个大的数据子表进行存放，即数据子表的个数为2，每个数据子表存200M数据，也可以用更多的存储量小的数据子表进行存放，例如，用10个数据子表存放400M数据，每个数据子表中存40M数据，即100条数据。对于第二种情况，例如，每条数据的占用空间为0.4M，预设的每个数据子表的存放数据量是100M（当然预置的每个数据子表的存储空间也可以不太相同，这里只是为了方便说明），于是，可以确定每个数据子表存放250条数据，假设数据表有1000条数据，那么需要4个数据子表。

[0069] 在确定了数据子表的个数之后，可以将数据表中的数据均匀的存入每个数据子表，也可以根据每个数据子表需要存放的数据的具体大小，将数据表中的数据非均匀的存入数据子表。

[0070] 可选的，根据字段对数据表进行切分，具体的，所述根据所述数据条数对相应的所述数据表进行切分，得到每个所述数据表对应的多个数据子表，包括：获取每个所述数据表中每条数据对应的字段数字标识；对所述字段数字标识进行求模运算，得到求模结果；根据所述求模结果和所述数据条数对相应的数据表进行切分，得到每个所述数据表对应的多个数据子表。

[0071] 所述字段数字标识，用于唯一标识一个字段的数字。例如，用户字段，为每个用户分配一个字段数字标识，用户1的字段数字标识为1256，用户2的字段数字标识为1058，用户3的字段数字标识为2002。

[0072] 然后对字段数字标识进行求模运算，得到求模结果。例如，用户1的字段数字标识1255的求模结果为1，用户2的字段数字标识1058的求模结果是3，用户3的字段数字标识2002的求模结果为2。

[0073] 最后根据求模结果和数据条数对相应的数据表进行切分。例如，将求模结果为1的数据放入数据子表1，将求模结果为2的数据放入数据子表2，将求模结果为3的数据放入数据子表3；进一步的，若某一求模结果的数据较多，例如，求模结果为1的数据较多，那么可以用两个数据子表来存放求模结果为1的数据，以使最终得到的数据子表中存放的数据量都差不多。

[0074] 进一步的，如图5所示，在步骤108所述根据所述数据子表的个数对所述数据表进行切分，得到每个所述数据表对应的多个数据子表之后，还包括：

[0075] 步骤109, 根据所述数据子表的个数, 确定数据库的个数。

[0076] 步骤110, 将所述多个数据子表存放于确定的多个数据库中。

[0077] 例如, 假设数据表1有6个数据子表, 数据表2有2个数据子表, 数据表3有15个数据子表; 则为数据表1设置2个数据库, 每个数据库放3个数据子表; 数据表2设置一个数据库, 用于存放两个数据子表; 数据表3由于数据子表个数较多, 于是可以设置多个数据库, 将不同的数据子表放入多个数据库中, 例如, 设置5个数据库, 每个数据库存放3个数据子表。当然, 在确定每个数据库存放的数据子表的数据量的时候, 不一定设置为每个数据库存放的数据子表的数量相同。例如, 数据表3的5个数据库, 第一个数据库存2个数据子表, 第二个数据库存放4个数据子表, 第三个数据库存放3个数据子表, 第四个数据库存放2个数据子表, 第五个数据库存放4个数据子表。

[0078] 为了更清楚的说明本发明实施例中数据的切分, 如图6所示, 首先按业务类型将数据切分, 得到多个数据表, 然后再对每个数据表进行切分, 得到多个数据子表。

[0079] 上述数据的切分方法, 首先获取数据集和所述数据集中数据的业务类型; 并且根据所述业务类型对所述数据集的数据进行切分, 得到多个数据表; 然后获取每个所述数据表中的数据条数; 最后根据所述数据条数对相应的所述数据表进行切分, 得到每个所述数据表对应的多个数据子表。可见, 通过上述方式, 由于获取到了多个数据子表对原始的数据集进行存放, 而不是将原始的数据集存放于一个数据表中, 所以在后续要查数据的时候, 可以直接在某一个数据子表中查询, 而不是在一个大的存储原始数据集的数据表中进行查询, 由于查询量减少, 所以能够提高查询效率, 提高响应效率, 并且, 上述数据的切分方式将数据首先按业务切分, 然后再对经过业务切分的表进行再次切分, 使得最终得到的数据子表相对只有一次切分得到的数据子表更小, 进一步的提高了查询效率和响应效率。

[0080] 如图7所示, 本发明实施例提供一种数据的切分方法, 具体包括如下步骤:

[0081] 步骤702, 获取数据集和所述数据集中数据的业务类型。

[0082] 步骤704, 根据所述业务类型对所述数据集的数据进行切分, 得到多个数据表

[0083] 步骤706，获取每个所述数据表中的数据条数。

[0084] 步骤708，根据所述数据条数对相应的所述数据表进行切分，得到每个所述数据表对应的多个数据子表。

[0085] 步骤710，将所述业务类型和所述数据表的对应关系、所述数据表与相应的所述数据子表的对应关系以及所述数据子表中的每条数据对应的数据标识发送到数据表管理中心，以使所述数据表管理中心根据接收到的数据处理请求中的业务类型和数据标识从相应的数据子表中查找与所述数据标识对应的数据。

[0086] 所述数据标识，为序号，用于指示数据子表中的第几条数据，例如，用数据标识0001标识该数据子表中的第一条数据，用0002标识数据子表中的第二条数据；所述数据标识，为字段标识，用于标识数据子表中某一字段的数据，例如，某一数据子表的主字段为用户，那么用数据标识0910标识用户1，0920标识用户2。

[0087] 所述对应关系，指示只要知道业务类型就知道该业务类型的数据存放在哪个数据表，或者只要知道数据表就能知道该数据表存放的是哪个业务类型的数据，同时，根据对应关系还知道表存放的位置（所述位置反映表所在的数据库和服务器，即不同的表不仅可以存放于不同的数据库，甚至可以存放在不同的服务器上面），例如表1所示。

[0088] 表1

□

业务类型 1	数据表 1	数据子表 1-1	多个数据标识	位置 1
		数据子表 1-2	多个数据标识	位置 2
业务类型 2	数据表 2	数据子表 2-1	多个数据标识	位置 3
		数据子表 2-2	多个数据标识	位置 4
		数据子表 2-3	多个数据标识	位置 5
业务类型 3	数据表 3	数据子表 3-1	多个数据标识	位置 6
		数据子表 3-2	多个数据标识	位置 7

[0089] 如图8所示，数据表管理中心接收到终端发送的数据处理请求，然后从该数据

处理请求中获取到业务类型和数据标识，首先根据获取到的业务类型，知道该业务类型的数据存放于数据表3中，进一步的，根据数据标识，知道该数据标识对应的数据存放于数据子表2中，于是从数据子表2中查找与所述数据标识对应的数据，这样，能够大大的提高数据的查询效率，因为其只用在数据子表2中进行查找既可。

[0090] 如图9所示，本发明实施例提供一种数据的切分装置900，该装置900包括：

[0091] 第一获取模块902，用于获取数据集和所述数据集中数据的业务类型；

[0092] 第一切分模块904，用于根据所述业务类型对所述数据集的数据进行切分，得到多个数据表；

[0093] 第二获取模块906，用于获取每个所述数据表中的数据条数；

[0094] 第二切分模块908，用于根据所述数据条数对相应的所述数据表进行切分，得到每个所述数据表对应的多个数据子表。

[0095] 上述数据的切分装置，首先获取数据集和所述数据集中数据的业务类型；并且根据所述业务类型对所述数据集的数据进行切分，得到多个数据表；然后获取每个所述数据表中的数据条数；最后根据所述数据条数对相应的所述数据表进行切分，得到每个所述数据表对应的多个数据子表。可见，通过上述方式，由于获取到了多个数据子表对原始的数据集进行存放，而不是将原始的数据集存放于一个数据表中，所以在后续要查数据的时候，可以直接在某一个数据子表中查询，而不是在一个大的存储原始数据集的数据表中进行查询，由于查询量减少，所以能够提高查询效率，提高响应效率，并且，上述数据的切分方式将数据首先按业务切分，然后再对经过业务切分的表进行再次切分，使得最终得到的数据子表相对只有一次切分得到的数据子表更小，进一步的提高了查询效率和响应效率。

[0096] 在其中一个实施例中，所述数据的切分装置900，还包括：

[0097] 第一个数确定模块，用于根据所述数据表的个数，确定数据库的个数；

[0098] 第一数据库存放模块，用于将所述多个数据表存放于确定的多个数据库中。

[0099] 在其中一个实施例中，所述第二切分模块908，包括：

[0100] 占用空间获取模块，用于获取每个所述数据表中的每条数据对应的占用空间；

- [0101] 占用空间确定模块，用于根据所述数据表中的数据条数和所述数据表中的每条数据对应的占用空间确定所述数据表对应的数据子表的个数；
- [0102] 占用空间切分模块，用于根据所述数据子表的个数对所述数据表进行切分，得到每个所述数据表对应的多个数据子表。
- [0103] 在其中一个实施例中，所述数据的切分装置900，还包括：
- [0104] 第二个数确定模块，用于根据所述数据子表的个数，确定数据库的个数；
- [0105] 第二数据库存放模块，用于将所述多个数据子表存放于确定的多个数据库中。
- [0106] 在其中一个实施例中，所述第二切分模块908，还包括：
- [0107] 字段数字标识获取模块，用于获取每个所述数据表中每条数据对应的字段数字标识；
- [0108] 求模模块，用于对所述字段数字标识进行求模运算，得到求模结果；
- [0109] 求模切分模块，用于根据所述求模结果和所述数据条数对相应的数据表进行切分，得到每个所述数据表对应的多个数据子表。
- [0110] 图10示出了一个实施例中计算机设备的内部结构图。该计算机设备具体可以是服务器。如图10所示，该计算机设备包括通过系统总线连接的处理器、存储器和网络接口。其中，存储器包括非易失性存储介质和内存储器。该计算机设备的非易失性存储介质存储有操作系统，还可存储有计算机程序，该计算机程序被处理器执行时，可使得处理器实现数据的切分方法。该内存储器中也可储存有计算机程序，该计算机程序被处理器执行时，可使得处理器执行数据的切分方法。网络接口用于与外部进行通信。本领域技术人员可以理解，图10中示出的结构，仅仅是与本申请方案相关的部分结构的框图，并不构成对本申请方案所应用于其上的计算机设备的限定，具体的计算机设备可以包括比图中所示更多或更少的部件，或者组合某些部件，或者具有不同的部件布置。
- [0111] 在一个实施例中，本申请提供的数据的切分方法可以实现为一种计算机程序的形式，计算机程序可在如图10所示的计算机设备上运行。计算机设备的存储器中可存储组成数据的切分装置的各个程序模板。比如，第一获取模块902、第一切分模块904、第二获取模块906以及第二切分模块908。
- [0112] 具体的，本发明实施例所述的计算机设备，包括存储器和处理器，所述存储器

存储有计算机程序，所述计算机程序被所述处理器执行时，使得所述处理器执行如下步骤：

- [0113] 获取数据集和所述数据集中数据的业务类型；
- [0114] 根据所述业务类型对所述数据集的数据进行切分，得到多个数据表；
- [0115] 获取每个所述数据表中的数据条数；
- [0116] 根据所述数据条数对相应的所述数据表进行切分，得到每个所述数据表对应的多个数据子表。
- [0117] 上述计算机设备，首先获取数据集和所述数据集中数据的业务类型；并且根据所述业务类型对所述数据集的数据进行切分，得到多个数据表；然后获取每个所述数据表中的数据条数；最后根据所述数据条数对相应的所述数据表进行切分，得到每个所述数据表对应的多个数据子表。可见，通过上述方式，由于获取到了多个数据子表对原始的数据集进行存放，而不是将原始的数据集存放于一个数据表中，所以在后续要查数据的时候，可以直接在某一个数据子表中查询，而不是在一个大的存储原始数据集的数据表中进行查询，由于查询量减少，所以能够提高查询效率，提高响应效率，并且，上述数据的切分方式将数据首先按业务切分，然后再对经过业务切分的表进行再次切分，使得最终得到的数据子表相对只有一次切分得到的数据子表更小，进一步的提高了查询效率和响应效率。
- [0118] 在其中一个实施例中，上述计算机程序被所述处理器执行时，还用于执行以下步骤：
 - [0119] 根据所述数据表的个数，确定数据库的个数；
 - [0120] 将所述多个数据表存放于确定的多个数据库中。
- [0121] 在其中一个实施例中，所述根据所述数据条数对相应的所述数据表进行切分，得到每个所述数据表对应的多个数据子表，包括：
 - [0122] 获取每个所述数据表中的每条数据对应的占用空间；
 - [0123] 根据所述数据表中的数据条数和所述数据表中的每条数据对应的占用空间确定所述数据表对应的数据子表的个数；
 - [0124] 根据所述数据子表的个数对所述数据表进行切分，得到每个所述数据表对应的

多个数据子表。

[0125] 在其中一个实施例中，上述计算机程序被所述处理器执行时，还用于执行以下步骤：

[0126] 根据所述数据子表的个数，确定数据库的个数；

[0127] 将所述多个数据子表存放于确定的多个数据库中。

[0128] 在其中一个实施例中，所述根据所述数据条数对相应的所述数据表进行切分，得到每个所述数据表对应的多个数据子表，包括：

[0129] 获取每个所述数据表中每条数据对应的字段数字标识；

[0130] 对所述字段数字标识进行求模运算，得到求模结果；

[0131] 根据所述求模结果和所述数据条数对相应的数据表进行切分，得到每个所述数据表对应的多个数据子表。

[0132] 在其中一个实施例中，上述计算机程序被所述处理器执行时，还用于执行以下步骤：

[0133] 将所述业务类型和所述数据表的对应关系、所述数据表与相应的所述数据子表的对应关系以及所述数据子表中的每条数据对应的数据标识发送到数据表管理中心，以使所述数据表管理中心根据接收到的数据处理请求中的业务类型和数据标识从相应的数据子表中查找与所述数据标识对应的数据。

[0134] 一种计算机可读存储介质，存储有计算机程序，所述计算机程序被处理器执行时，使得所述处理器执行如下步骤：

[0135] 获取数据集和所述数据集中数据的业务类型；

[0136] 根据所述业务类型对所述数据集的数据进行切分，得到多个数据表；

[0137] 获取每个所述数据表中的数据条数；

[0138] 根据所述数据条数对相应的所述数据表进行切分，得到每个所述数据表对应的多个数据子表。

[0139] 上述计算机可读存储介质，首先获取数据集和所述数据集中数据的业务类型；并且根据所述业务类型对所述数据集的数据进行切分，得到多个数据表；然后获取每个所述数据表中的数据条数；最后根据所述数据条数对相应的所述数据表进行切分，得到每个所述数据表对应的多个数据子表。可见，通过上述方式

，由于获取到了多个数据子表对原始的数据集进行存放，而不是将原始的数据集存放于一个数据表中，所以在后续要查数据的时候，可以直接在某一个数据子表中查询，而不是在一个大的存储原始数据集的数据表中进行查询，由于查询量减少，所以能够提高查询效率，提高响应效率，并且，上述数据的切分方式将数据首先按业务切分，然后再对经过业务切分的表进行再次切分，使得最终得到的数据子表相对只有一次切分得到的数据子表更小，进一步的提高了查询效率和响应效率。

[0140] 在其中一个实施例中，上述计算机程序被所述处理器执行时，还用于执行以下步骤：

[0141] 根据所述数据表的个数，确定数据库的个数；

[0142] 将所述多个数据表存放于确定的多个数据库中。

[0143] 在其中一个实施例中，所述根据所述数据条数对相应的所述数据表进行切分，得到每个所述数据表对应的多个数据子表，包括：

[0144] 获取每个所述数据表中的每条数据对应的占用空间；

[0145] 根据所述数据表中的数据条数和所述数据表中的每条数据对应的占用空间确定所述数据表对应的数据子表的个数；

[0146] 根据所述数据子表的个数对所述数据表进行切分，得到每个所述数据表对应的多个数据子表。

[0147] 在其中一个实施例中，上述计算机程序被所述处理器执行时，还用于执行以下步骤：

[0148] 根据所述数据子表的个数，确定数据库的个数；

[0149] 将所述多个数据子表存放于确定的多个数据库中。

[0150] 在其中一个实施例中，所述根据所述数据条数对相应的所述数据表进行切分，得到每个所述数据表对应的多个数据子表，包括：

[0151] 获取每个所述数据表中每条数据对应的字段数字标识；

[0152] 对所述字段数字标识进行求模运算，得到求模结果；

[0153] 根据所述求模结果和所述数据条数对相应的数据表进行切分，得到每个所述数据表对应的多个数据子表。

[0154] 在其中一个实施例中，上述计算机程序被所述处理器执行时，还用于执行以下步骤：

[0155] 将所述业务类型和所述数据表的对应关系、所述数据表与相应的所述数据子表的对应关系以及所述数据子表中的每条数据对应的数据标识发送到数据表管理中心，以使所述数据表管理中心根据接收到的数据处理请求中的业务类型和数据标识从相应的数据子表中查找与所述数据标识对应的数据。

[0156] 需要说明的是，上述数据的切分方法实施例、数据的切分装置实施例、计算机设备实施例和计算机可读存储介质实施例属于同一发明构思，方法实施例、装置实施例、计算机设备实施例和计算机可读存储介质实施例中的内容可相互适用。

[0157] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，是可以通过计算机程序来指令相关的硬件来完成，所述的程序可存储于一非易失性计算机可读取存储介质中，该程序在执行时，可包括如上述各方法的实施例的流程。其中，本申请所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用，均可包括非易失性和/或易失性存储器。非易失性存储器可包括只读存储器(ROM)、可编程ROM(PROM)、电可编程ROM(EPROM)、电可擦除可编程ROM(EEPROM)或闪存。易失性存储器可包括随机存取存储器(RAM)或者外部高速缓冲存储器。作为说明而非局限，RAM以多种形式可得，诸如静态RAM(SRAM)、动态RAM(DRAM)、同步DRAM(SDRAM)、双数据率SDRAM(DDRSDRAM)、增强型SDRAM(ESDRAM)、同步链路(Synchlink) DRAM(SLDRAM)、存储器总线(Rambus)直接RAM(RDRAM)、直接存储器总线动态RAM(DRDRAM)、以及存储器总线动态RAM(RDRAM)等。

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

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

进，这些都属于本申请的保护范围。因此，本申请专利的保护范围应以所附权利要求为准。

权利要求书

- [权利要求 1] 一种数据的切分方法，其特征在于，所述方法包括：
获取数据集和所述数据集中数据的业务类型；
根据所述业务类型对所述数据集的数据进行切分，得到多个数据表；
获取每个所述数据表中的数据条数；
根据所述数据条数对相应的所述数据表进行切分，得到每个所述数据表对应的多个数据子表。
- [权利要求 2] 如权利要求1所述的方法，其特征在于，在所述根据所述业务类型对所述数据集的数据进行切分，得到多个数据表之后，还包括：
根据所述数据表的个数，确定数据库的个数；
将所述多个数据表存放于确定的多个数据库中。
- [权利要求 3] 如权利要求1所述的方法，其特征在于，所述根据所述数据条数对相应的所述数据表进行切分，得到每个所述数据表对应的多个数据子表，包括：
获取每个所述数据表中的每条数据对应的占用空间；
根据所述数据表中的数据条数和所述数据表中的每条数据对应的占用空间确定所述数据表对应的数据子表的个数；
根据所述数据子表的个数对所述数据表进行切分，得到每个所述数据表对应的多个数据子表。
- [权利要求 4] 如权利要求3所述的方法，其特征在于，在所述根据所述数据子表的个数对所述数据表进行切分，得到每个所述数据表对应的多个数据子表之后，还包括：
根据所述数据子表的个数，确定数据库的个数；
将所述多个数据子表存放于确定的多个数据库中。
- [权利要求 5] 如权利要求1所述的方法，其特征在于，所述根据所述数据条数对相应的所述数据表进行切分，得到每个所述数据表对应的多个数据子表，包括：
获取每个所述数据表中每条数据对应的字段数字标识；

对所述字段数字标识进行求模运算，得到求模结果；

根据所述求模结果和所述数据条数对相应的数据表进行切分，得到每个所述数据表对应的多个数据子表。

[权利要求 6] 如权利要求1至5任一项所述的方法，其特征在于，在所述根据所述数据条数对相应的所述数据表进行切分，得到每个所述数据表对应的多个数据子表之后，还包括：

将所述业务类型和所述数据表的对应关系、所述数据表与相应的所述数据子表的对应关系以及所述数据子表中的每条数据对应的数据标识发送到数据表管理中心，以使所述数据表管理中心根据接收到的数据处理请求中的业务类型和数据标识从相应的数据子表中查找与所述数据标识对应的数据。

[权利要求 7] 一种数据的切分装置，其特征在于，所述装置包括：

第一获取模块，用于获取数据集和所述数据集中数据的业务类型；

第一切分模块，用于根据所述业务类型对所述数据集的数据进行切分，得到多个数据表；

第二获取模块，用于获取每个所述数据表中的数据条数；

第二切分模块，用于根据所述数据条数对相应的所述数据表进行切分，得到每个所述数据表对应的多个数据子表。

[权利要求 8] 如权利要求7所述的装置，其特征在于，还包括：

第一个数确定模块，用于根据所述数据表的个数，确定数据库的个数；

第一数据库存放模块，用于将所述多个数据表存放于确定的多个数据库中。

[权利要求 9] 一种计算机设备，包括存储器和处理器，所述存储器存储有计算机程序，所述计算机程序被所述处理器执行时，使得所述处理器执行如权利要求1至6中任一项所述方法的步骤。

[权利要求 10] 在一种计算机可读存储介质，存储有计算机程序，所述计算机程序被处理器执行时，使得所述处理器执行如权利要求1至6中任一项所述方

法的步骤。

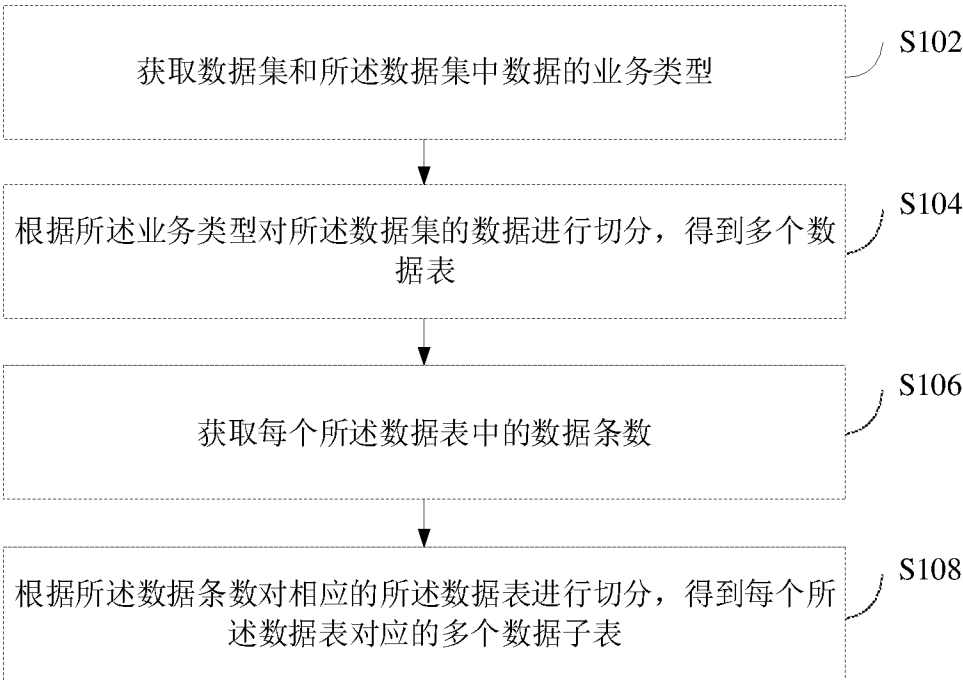


图 1

用户	订单金额	下单时间	优惠券	商品类型	一条数据
用户1	100	2019.10.10	满20减5	包包	一条数据
用户2	200	2019.10.11	满20减5	衣服	
用户3	50	2019.10.11	满20减10	裤子	一条数据

图 2

	一条数据			一条数据	
用户	用户1	用户2	用户3	用户4	
订单金额	100	100	100	100	
下单时间	2019.10.10	2019.10.11	2019.10.11	2019.10.12	
优惠券	满20减5	满20减5	满20减10	满20减10	
商品类型	包包	衣服	裤子	鞋子	

图 3

	一条数据			一条数据	
下单时间	2019.10.10	2019.10.11	2019.10.11	2019.10.12	
用户	用户1	用户2	用户3	用户4	
优惠券	满20减5	满20减5	满20减10	满20减10	
商品类型	包包	衣服	裤子	鞋子	
订单金额	100	100	50	50	

图 4



图 5

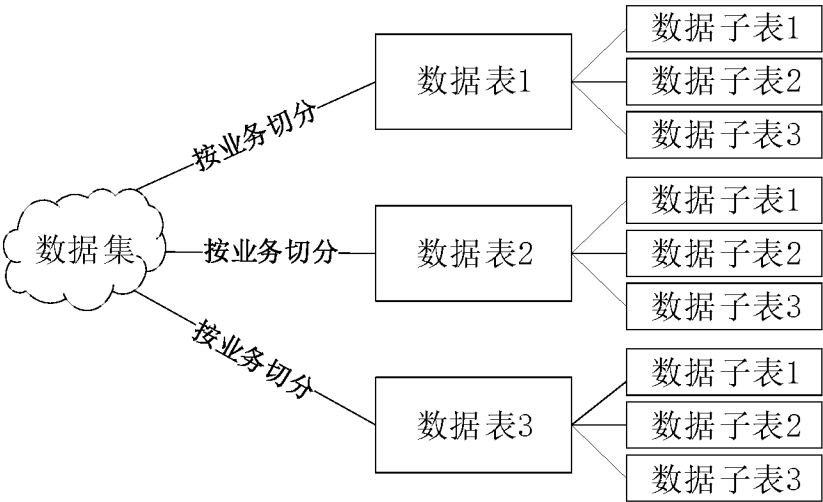


图 6

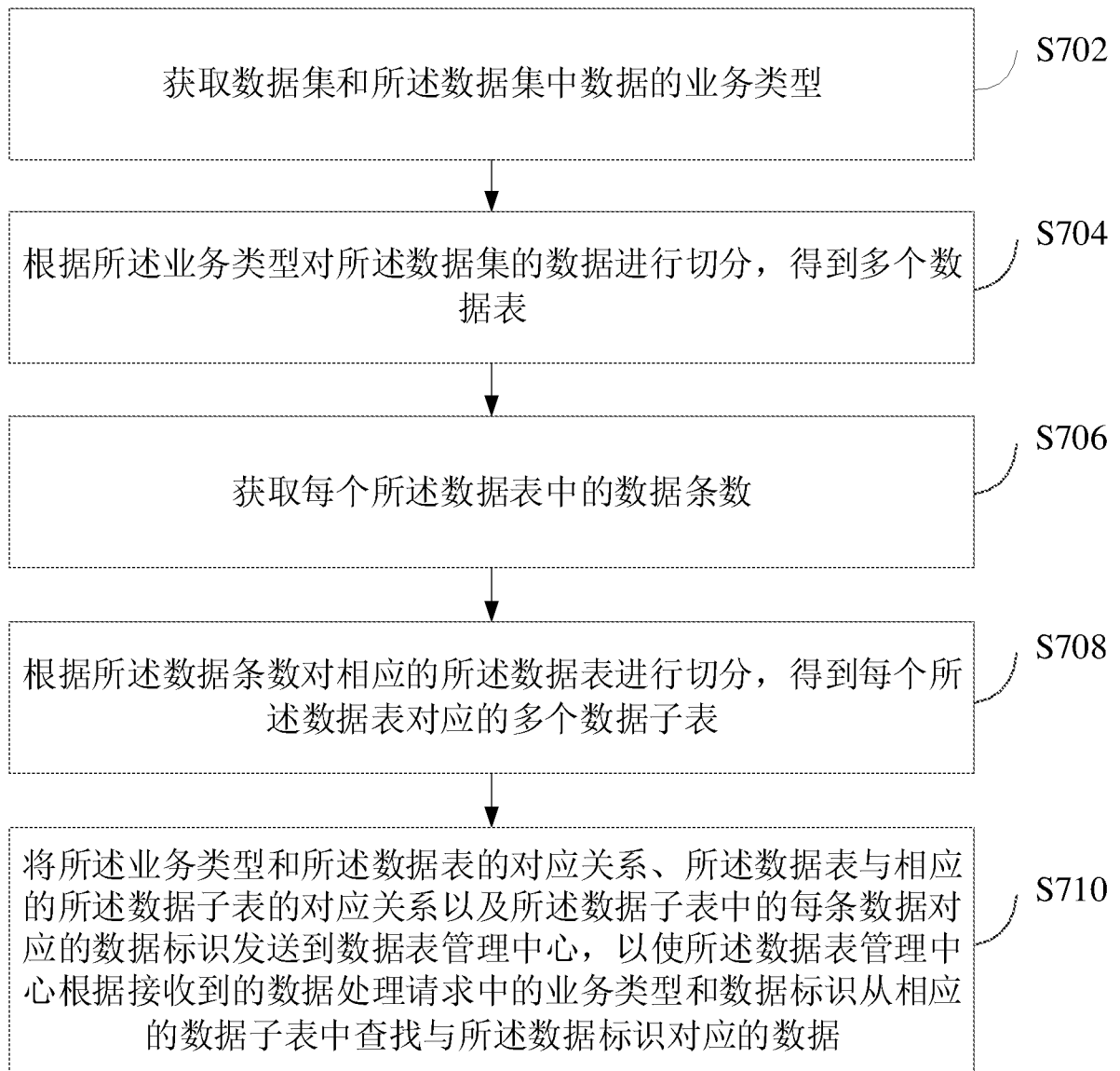


图 7

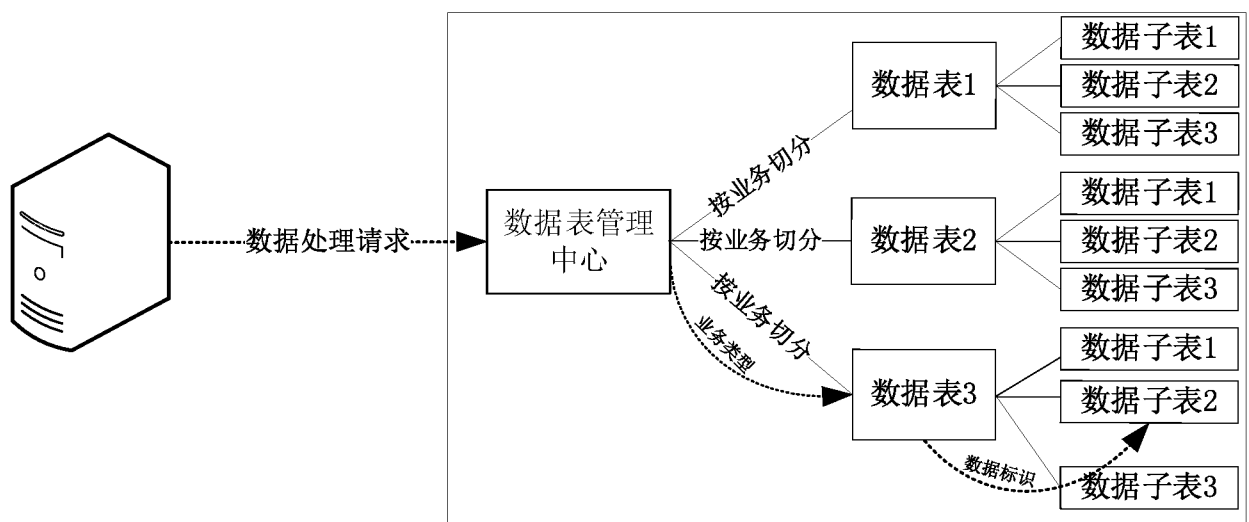


图 8

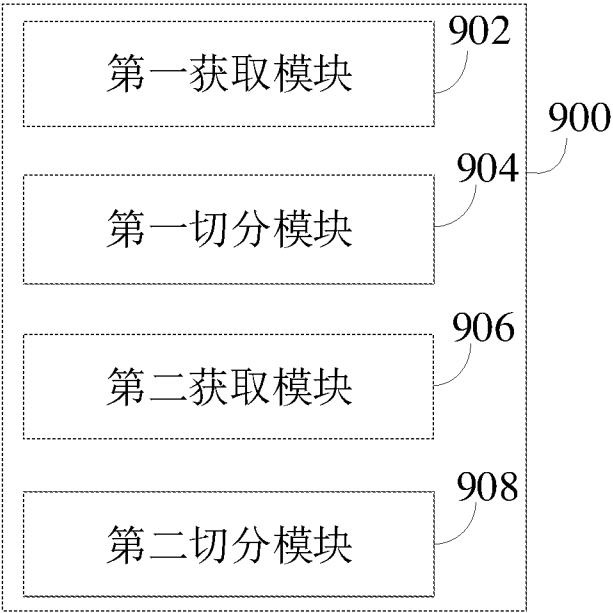


图 9

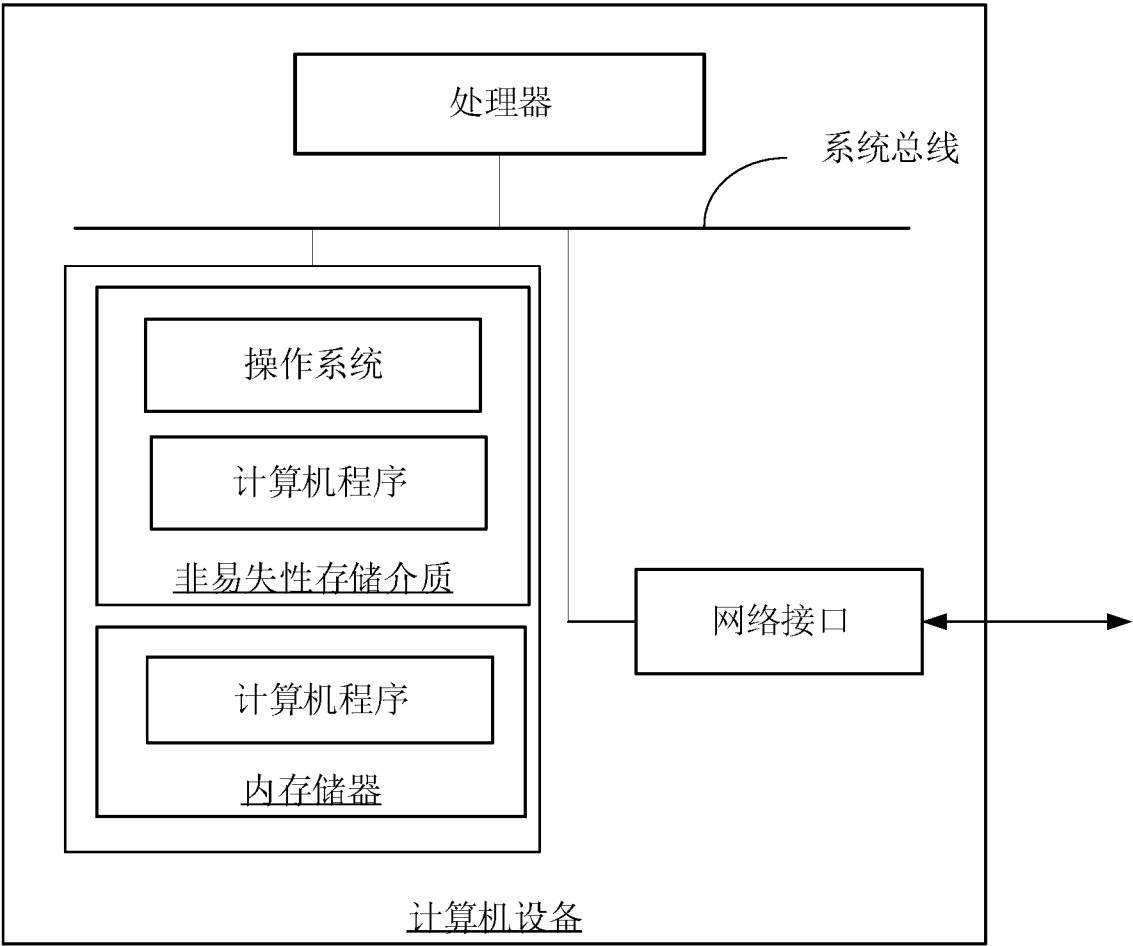


图 10

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/122380

A. CLASSIFICATION OF SUBJECT MATTER

G06F 16/00(2019.01)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)

CNABS; CNTXT; VEN; WOTXT; EPTXT; USTXT; CNKI: 数据, 切分, 划分, 数据表, 表, 子表, 记录数, 行, 查询, 查找, 效率, data, divide, datasheet, table, subtable, record numbers, row, query, efficiency

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 103605755 A (HUAZHONG UNIVERSITY OF SCIENCE AND TECHNOLOGY) 26 February 2014 (2014-02-26) description, paragraphs 29-33	1-10
Y	CN 106294740 A (BEIJING CHUANGRUI CULTURE & MEDIA CO., LTD.) 04 January 2017 (2017-01-04) description, paragraphs 28-40 and 70	1-10
A	US 9600553 B1 (SYMANTEC CORPORATION) 21 March 2017 (2017-03-21) entire document	1-10

☐ 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

19 August 2019

Date of mailing of the international search report

18 September 2019

Name and mailing address of the ISA/CN

National Intellectual Property Administration, PRC (ISA/
CN)
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/CN2018/122380

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)		Publication date (day/month/year)	
CN	103605755	A	26 February 2014	CN	103605755	B	29 July 2015
CN	106294740	A	04 January 2017	None			
US	9600553	B1	21 March 2017	None			

国际检索报告

国际申请号

PCT/CN2018/122380

A. 主题的分类

G06F 16/00 (2019.01) 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))

CNABS; CNTXT; VEN; WOTXT; EPTXT; USTXT; CNKI: 数据, 切分, 划分, 数据表, 表, 子表, 记录数, 行, 查询, 查找, 效率, data, divide, datasheet, table, subtable, record numbers, row, query, efficiency

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
Y	CN 103605755 A (华中科技大学) 2014年 2月 26日 (2014 - 02 - 26) 说明书第29-33段	1-10
Y	CN 106294740 A (北京创锐文化传媒有限公司) 2017年 1月 4日 (2017 - 01 - 04) 说明书第28-40、70段	1-10
A	US 9600553 B1 (SYMANTEC CORPORATION) 2017年 3月 21日 (2017 - 03 - 21) 全文	1-10

☐ 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

2019年 8月 19日

国际检索报告邮寄日期

2019年 9月 18日

ISA/CN的名称和邮寄地址

中国国家知识产权局 (ISA/CN)

中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

受权官员

郑舒玲

电话号码 86-(20)-28950720

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/122380

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	103605755	A	2014年 2月 26日	CN 103605755 B	2015年 7月 29日
CN	106294740	A	2017年 1月 4日	无	
US	9600553	B1	2017年 3月 21日	无	