



(12)发明专利申请

(10)申请公布号 CN 111309704 A  
(43)申请公布日 2020.06.19

(21)申请号 201811517469.7

(22)申请日 2018.12.12

(71)申请人 北京沃东天骏信息技术有限公司  
地址 100176 北京市大兴区北京经济技术  
开发区科创十一街18号院2号楼4层  
A402室

(72)发明人 高飞

(74)专利代理机构 北京成创同维知识产权代理  
有限公司 11449  
代理人 蔡纯 张靖琳

(51)Int.Cl.  
G06F 16/21(2019.01)  
G06F 16/25(2019.01)

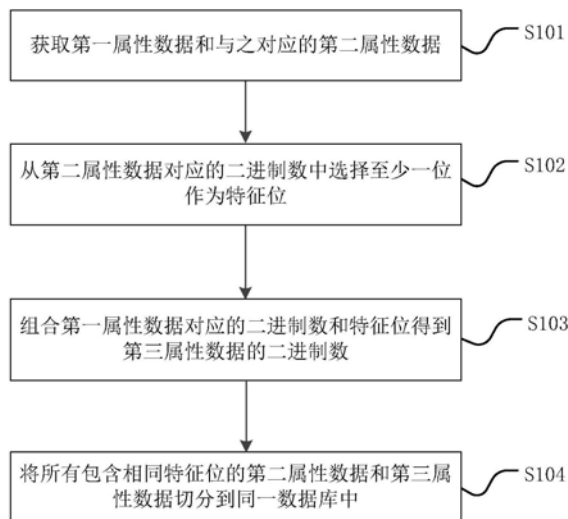
权利要求书2页 说明书9页 附图7页

(54)发明名称

数据库操作方法和数据库操作系统

(57)摘要

本发明实施例提供一种数据库操作方法和数据库操作系统,该数据库操作方法包括:获取第一属性数据和与之对应的第二属性数据;从所述第二属性数据对应的二进制数中选择至少一位作为特征位;组合所述第一属性数据对应的二进制数和所述特征位得到第三属性数据的二进制数;以及将所有包含相同特征位的所述第二属性数据和所述第三属性数据切分到同一数据库中。通过从第二属性数据中提取特征位并将其添加至第一属性数据对应的二进制数中组成第三属性数据,使得第二属性数据和第三属性数据均含有特征位,再按照特征位切分数据库,从而采用第二属性数据和第三属性数据查询都可以直接定位到数据库。



1. 一种数据库操作方法,其特征在于,包括:  
获取第一属性数据和与之对应的第二属性数据;  
从所述第二属性数据对应的二进制数中选择至少一位作为特征位;  
组合所述第一属性数据对应的二进制数和所述特征位得到第三属性数据的二进制数;  
以及  
将所有包含相同特征位的所述第二属性数据和所述第三属性数据切分到同一数据库中。
2. 根据权利要求1所述的数据库操作方法,其特征在于,从所述第二属性数据对应的二进制数的末尾取所述特征位。
3. 根据权利要求2所述的数据库操作方法,其特征在于,还包括:根据所述第二属性数据除以需要切分的数据库的个数取余的结果切分数据库。
4. 根据权利要求1所述的数据库操作方法,其特征在于,将所述特征位添加至所述第一属性数据对应的二进制数的末尾,组成第三属性数据对应的二进制数。
5. 根据权利要求4所述的数据库操作方法,其特征在于,还包括:将带有特征位的所述第三属性数据除以所述数据库的个数取余的结果相同的所有所述第三属性数据切分到同一个数据库中。
6. 根据权利要求1所述的数据库操作方法,其特征在于,当所述数据库的个数为N时,所述特征位的位数为(N-1)对应的最简单的二进制数的位数,N为大于等于1的整数。
7. 根据权利要求1所述的数据库操作方法,其特征在于,所述第一属性数据为订单数据,所述第二属性数据为买家数据。
8. 根据权利要求1所述的数据库操作方法,其特征在于,采用分布式服务全局ID算法生成与所述第一属性数据对应的二进制数。
9. 根据权利要求1所述的数据库操作方法,其特征在于,采用水平切分方式进行所述数据库的切分。
10. 一种数据库操作系统,其特征在于,包括:  
数据获取单元,用于获取第一属性数据和与之对应的第二属性数据;  
特征位选择单元,用于从所述第二属性数据对应的二进制数中选择至少一位作为特征位;  
数据组建单元,用于组合所述第一属性数据对应的二进制数和所述特征位得到第三属性数据的二进制数;以及  
合并单元,用于将所有包含相同特征位的所述第二属性数据和所述第三属性数据切分到同一数据库中。
11. 根据权利要求10所述的数据库操作系统,其特征在于,所述特征位选择单元从所述第二属性数据对应的二进制数的末尾取所述特征位。
12. 根据权利要求11所述的数据库操作系统,其特征在于,还包括:数据库切分单元,用于根据所述第二属性数据除以需要切分的数据库的个数取余的结果切分数据库。
13. 根据权利要求10所述的数据库操作系统,其特征在于,所述数据组建单元将所述特征位添加至所述第一属性数据对应的二进制数的末尾,组成第三属性数据对应的二进制数。

14. 根据权利要求13所述的数据库操作系统,其特征在于,还包括:归类单元,用于将带有特征位的所述第三属性数据除以所述数据库的个数取余的结果相同的所有所述第三属性数据切分到同一个数据库中。

15. 根据权利要求10所述的数据库操作系统,其特征在于,当所述数据库的个数为N时,所述特征位的位数为(N-1)对应的最简单的二进制数的位数,N为大于等于1的整数。

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计算机指令,所述计算机指令被执行时实现如权利要求1至9中任一项所述的数据库操作方法。

17. 一种数据库操作装置,其特征在于,包括:

存储器,用于存储计算机指令;

处理器,耦合到所述存储器,所述处理器被配置为基于所述存储器存储的计算机指令执行实现如权利要求1至9中任一项所述的数据库操作方法。

## 数据库操作方法和数据库操作系统

### 技术领域

[0001] 本发明涉及计算机技术领域,具体涉及一种数据库操作方法和数据库操作系统。

### 背景技术

[0002] 随着互联网的飞速发展,应用数据量及访问量快速增长,单台数据库服务器的资源通常难以支撑大量的数据量和数据库操作请求。为了解决该问题,需要对数据库进行分库分表,一种是按业务分类进行垂直切分;另一种是按一定的规则(如数值范围、数值哈希)进行水平切分,即把一个数据库水平切分成多个部分放到不同的数据库服务器上。然而在水平分库分表之后,大表中的数据分散存储在各个数据库中。在进行查询时往往需要通过范围法或者哈希法找到对应的数据库进行查询,但这只能满足按照关键字来进行查询,而不能按照其他属性进行直接查询。

[0003] 按照目前的数据库切分方法,在根据其他属性进行数据查询时会用到例如索引表法、缓存映射法和路由表法等方法。索引表是用来维护其他属性与关键字的对应关系,当访问数据表时,先通过索引表找到属性对应的关键字,再通过关键字按照范围法或哈希法找到对应的数据库进行查询;缓存映射法是将映射的结果存储在缓存中,如果数据量过大,需要根据属性自动进行Cache的水平切分,查询时要先进行Cache查询;路由表法是单独维护一张路由表,查询请求先通过属性查找路由表,找到该条数据所在的数据库,再进行查询。这些方法都需要多经过一次查询步骤,操作复杂,步骤冗余,性能下降。

[0004] 因此,发明人认为,上述的数据库切分方法在按照其他属性进行数据查询时有很大的局限性,不能直接定位到需要的数据表中,而是需要耗费多余的步骤和时间,查询结果不直观且效率低下。

### 发明内容

[0005] 有鉴于此,本发明实施例提供一种数据库操作方法和数据库操作系统,从第二属性数据中提取特征位,将特征位添加至与第二属性数据具有关联关系的第一属性数据中,组成第三属性数据,并基于特征位水平切分数据库,同时满足按照第二属性数据和第三属性数据进行数据查询,提高了查询效率。

[0006] 根据本发明第一方面,提供一种数据库操作方法,包括:获取第一属性数据和与之对应的第二属性数据;从所述第二属性数据对应的二进制数中选择至少一位作为特征位;组合所述第一属性数据对应的二进制数和所述特征位得到第三属性数据的二进制数;以及将所有包含相同特征位的所述第二属性数据和所述第三属性数据切分到同一数据库中。

[0007] 优选地,从所述第二属性数据对应的二进制数的末尾取所述特征位。

[0008] 优选地,所述数据库操作方法还包括:根据所述第二属性数据除以需要切分的数据库的个数取余的结果切分数据库。

[0009] 优选地,将所述特征位添加至所述第一属性数据对应的二进制数的末尾,组成第三属性数据对应的二进制数。

[0010] 优选地,所述数据库操作方法还包括:将带有特征位的所述第三属性数据除以所述数据库的个数取余的结果相同的所有所述第三属性数据切分到同一个数据库中。

[0011] 优选地,当所述数据库的个数为N时,所述特征位的位数为(N-1)对应的最简单的二进制数的位数,N为大于等于1的整数。

[0012] 优选地,所述第一属性数据为订单数据,所述第二属性数据为买家数据。

[0013] 优选地,采用分布式服务全局ID算法生成与所述第一属性数据对应的二进制数。

[0014] 优选地,采用水平切分方式进行所述数据库的切分。

[0015] 根据本发明第二方面,提供一种数据库操作系统,包括:数据获取单元,用于获取第一属性数据和与之对应的第二属性数据;特征位选择单元,用于从所述第二属性数据对应的二进制数中选择至少一位作为特征位;数据组建单元,用于组合所述第一属性数据对应的二进制数和所述特征位得到第三属性数据的二进制数;以及合并单元,用于将所有包含相同特征位的所述第二属性数据和所述第三属性数据切分到同一数据库中。

[0016] 优选地,所述特征位选择单元从所述第二属性数据对应的二进制数的末尾取所述特征位。

[0017] 优选地,所述数据库操作系统还包括:数据库切分单元,用于根据所述第二属性数据除以需要切分的数据库的个数取余的结果切分数据库。

[0018] 优选地,所述数据组建单元将所述特征位添加至所述第一属性数据对应的二进制数的末尾,组成第三属性数据对应的二进制数。

[0019] 优选地,所述数据库操作系统还包括:归类单元,用于将带有特征位的所述第三属性数据除以所述数据库的个数取余的结果相同的所有所述第三属性数据切分到同一个数据库中。

[0020] 优选地,当所述数据库的个数为N时,所述特征位的位数为(N-1)对应的最简单的二进制数的位数,N为大于等于1的整数。

[0021] 根据本发明第三方面,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计算机指令,所述计算机指令被执行时,实现如上述所述的数据库操作方法。

[0022] 根据本发明第四方面,提供一种数据库操作装置,包括:存储器,用于存储计算机指令;处理器,耦合到所述存储器,所述处理器被配置为基于所述存储器存储的计算机指令执行实现如上述所述的数据库操作方法。

[0023] 本发明的实施例具有以下优点或有益效果:获取具有相互关系的第一属性数据和第二属性数据,从第二属性数据中选取特征位并将其添加至第一属性数据中,组成第三属性数据,然后将所有包含相同特征位的第二属性数据和第三属性数据切分到同一数据库中。通过基于特征位切分数据库的方法,使得具有相互关系的第一属性数据和带有特征位的第三属性数据能切分至同一个数据库中,因此不但能满足按照第三属性数据来进行的查询,而且能够满足按照第二属性数据来进行快速查询,极大地提高了数据库数据的查询效率,同时提升了服务器的稳定性和负载能力。

[0024] 本发明的另一优选实施例具有以下优点或有益效果:选择第二属性数据对应的二进制数的末尾几位作为特征位,并根据第二属性数据除以数据库的个数取余的结果来进行数据库的切分;再将特征位添加至第一属性数据对应的二进制数末尾,组成第三属性数据的二进制数,将带有特征位的第三属性数据进行除以数据库个数取余操作,将具有相同结

果的第三属性数据归类到同一个数据库中。通过采用二级制数取余等操作进行数据库切分,便于计算机处理,而且因为第三属性数据和第二属性数据末尾均带有特征位,取余后的结果相同,可以快速地将具有相互关联关系的第三属性数据和第二属性数据切分到同一个数据库中,也便于后续的查询操作。

[0025] 本发明的另一优选实施例具有以下优点或有益效果:当数据库的个数为N时,特征位的位数为(N-1)对应的最简单的二进制数的位数(N为大于等于1的整数),建立起了特征位的个数与数据库的个数之间的相互关系,引出取余划分数据库的操作,保证能切分成足够数量的数据库存储第二属性数据,避免造成冗余计算步骤,节约计算资源。

### 附图说明

[0026] 通过参照以下附图对本发明实施例的描述,本发明的上述以及其它目的、特征和优点将更为清楚,在附图中:

[0027] 图1示出了本发明实施例中的数据库操作方法的流程图;

[0028] 图2示出了本发明实施例中汇总的数据库操作方法的流程图;

[0029] 图3示出了本发明实施例中的数据组建过程的示意图;

[0030] 图4示出了本发明实施例中数据库查询的操作示意图;

[0031] 图5示出了本发明实施例中数据库操作系统的结构图;

[0032] 图6示出了本发明实施例中汇总的数据库操作系统的结构图;

[0033] 图7示出了根据本发明实施例的数据库操作装置的结构图。

### 具体实施方式

[0034] 以下基于实施例对本发明进行描述,但是本发明并不仅仅限于这些实施例。在下文对本发明的细节描述中,详尽描述了一些特定的细节部分。对本领域技术人员来说没有这些细节部分的描述也可以完全理解本发明。为了避免混淆本发明的实质,公知的方法、过程、流程没有详细叙述。另外附图不一定是按比例绘制的。

[0035] 图1示出了本发明实施例中的数据库操作方法的流程图,具体步骤包括S101-S103。

[0036] 在步骤S101中,获取第一属性数据和与之对应的第二属性数据。

[0037] 在步骤S102中,从第二属性数据对应的二进制数中选择至少一位作为特征位。

[0038] 在步骤S103中,组合第一属性数据对应的二进制数和特征位得到第三属性数据的二进制数。

[0039] 在步骤S104中,将所有包含相同特征位的第二属性数据和第三属性数据切分到同一数据库中。

[0040] 如背景技术中所述,现有的数据库在进行水平分库后,大表中的数据分散存储在各个数据库中,在进行数据查询时往往只能通过关键字直接找到对应的数据库进行查询,而不能直接通过其他的属性进行直接查询,在使用其他属性查询时,首先要通过不同的算法查找到其他属性对应的关键字,然后再通过关键字查询,查询效率低下,性能下降。

[0041] 因而在本实施例中,通过获取具有相互关系的第一属性数据和第二属性数据,从第二属性数据中选取特征位并将其添加至第一属性数据中,组成第三属性数据的二进制

数,第一属性数据和第三属性数据带有相同的属性,然后基于特征位切分数据库,将具有第一属性(例如是关键字)的第三属性数据和具有第二属性(例如是其他属性)的第二属性数据均切分至同一数据库中,从而不但能满足按照第一属性来进行直接查询,而且能够满足按照第二属性进行直接查询,大大提高了数据的查询效率。

[0042] 首先,获取待分库的大量的第一属性数据和第二属性数据,第一属性数据和第二属性数据之间具有关联关系。第一属性数据带有的属性可以理解为现有技术中提到的关键字的特征,第二属性数据带有的属性可以理解为现有技术中提到的其他属性特征。现有技术的分库方法,查询数据时只能按照关键字直接查询,而本实施例的分库方法,在数据查询时,通过第一属性和第二属性均可以直接定位到相应的数据库。

[0043] 一般情况下,切分数据库时,会按照将第一属性数据作为唯一字段的方法将数据进行分库,为保证数据表中需要查询的带有其它属性的其他字段能快速定位到目标数据库,那么需要将查询字段作为分库的数据基因融入到唯一字段上。第一属性数据和第二属性数据之间具有关联性,如果使两种数据均带有相同的数据基因,再按照数据基因分库,那么就可以很容易地将两种数据切分到同一个数据库中,进行数据查询时,不论通过哪一种属性都可以直接定位到。所以在获取数据之后,要进行数据基因的确定,数据在计算机中存储使用的是二进制数,所以从第二属性数据对应的二进制数中选择至少一位作为特征位,将特征位作为数据基因。

[0044] 然后获取第一属性数据对应的二进制数,在第一属性数据对应的二进制数中加入提取的特征位,组成第三属性数据对应的二进制数,使第三属性数据和第二属性数据带有相同的数据基因,而第一属性数据和第三属性数据表示的属性相同(例如关键字)。

[0045] 最后根据数据基因分库,将所有包含相同特征位的第三属性数据和第二属性数据切分到同一数据库中,从而保证用户通过第三属性数据和第二属性数据查询时均能快速定位到数据库,提高查询速度。

[0046] 本实施例的数据库操作方法,首先获取具有相互关系的第一属性数据和第二属性数据,从第二属性数据中选取特征位并将其添加至第一属性数据中,组成第三属性数据,然后将所有包含相同特征位的第二属性数据和第三属性数据切分到同一数据库中。通过基于特征位切分数据库的方法,使得具有相互关系的第三属性数据和第二属性数据能切分至同一个数据库中,因此不但能满足按照第一属性(关键字)来进行的查询,而且能够满足按照第二属性(其他属性)来进行快速查询,极大地提高了数据库数据的查询效率,同时提升了服务器的稳定性和负载能力。

[0047] 图2示出了本发明实施例中汇总的数据库操作方法的流程图,具体包括以下步骤。

[0048] 在步骤S201中,获取第一属性数据和与之对应的第二属性数据。

[0049] 在步骤S202中,从第二属性数据对应的二进制数中选择至少一位作为特征位。

[0050] 在步骤S203中,根据第二属性数据除以需要切分的数据库的个数取余的结果切分数据库。

[0051] 在步骤S204中,组合第一属性数据对应的二进制数和特征位得到第三属性数据的二进制数。

[0052] 在步骤S205中,将带有特征位的第三属性数据除以数据库的个数取余的结果相同的所有第三属性数据切分到同一个数据库中。

[0053] 在步骤S206中,将所有包含相同特征位的第二属性数据和第三属性数据切分到同一数据库中。

[0054] 本实施例是比前述实施例更加完善的数据库操作方法。其中,步骤S201-S202、步骤S204和步骤S206与图1的步骤S101-S104相同,这里就不再赘述。

[0055] 在步骤S202中,从第二属性数据对应的二进制数中获取到了作为数据基因的特征位,然而特征位的位数和位置的选取,对数据库的切分有很大的影响。

[0056] 在一个实施例中,从第二属性数据对应的二进制数的末尾取特征位。二进制数只有0和1两种表示,末尾(即个位)的最后一位为0时,对应的十进制数为0,末尾最后一位为1时,对应的十进制数为1,除此之外,其他位上的数表示的十进制数都是2的倍数。而在实际操作中,进行数据分类时,经常会按照一个十进制数除以另一个数进行取余操作分类,当除数为 $2^x$ ( $x$ 大于等于1)时,当被除数为奇数,即对应的二进制数的个位为1时,余数为奇数;而当被除数为偶数,即对应的二进制数的个位数为0时,余数为偶数。所以数字对应的二进制数的末尾一位对于取余的结果非常重要。所以本实施例中,取第二属性数据对应的二进制数的末尾至少一位作为特征位,而采取取余方式进行数据库分库。

[0057] 在步骤S203中,根据第二属性数据除以需要切分的数据库的个数取余的结果切分数据库。

[0058] 在本实施例中,采用第二属性数据的十进制数除以大于1的整数取余的结果进行数据库的切分,除数为需要切分的数据库的个数,取余的结果取决于特征位的位数,而特征位的位数是由数据库的个数来决定的。

[0059] 在一个可选的实施例中,根据需要切分的数据库的个数来决定特征位的位数,当数据库的个数为 $N$ 时,特征位的位数为 $(N-1)$ 对应的最简单的二进制数的位数, $N$ 为大于等于1的整数。

[0060] 例如,数据库的个数为2时, $(2-1)$ 对应的最简单的二进制数为1,所以特征位的位数为1;另一方面,利用第二属性数据除以2取余的结果进行划分数据库,因此需要取余的结果为0和1,对应的二进制数为0和1,所以对应的特征位的位数为1。在一个实施例中,还可以按照取余结果对数据库命名,例如,第二属性数据为3和8,分别除以2取余后的结果为1和0,则数据库对应可以命名为“数据库1”和“数据库0”例如,当数据库的个数为5时, $(5-1)$ 对应的最简单的二进制数为100,所以特征位的位数为3,利用第二属性数据除以5取余的结果进行划分数据库,需要第二属性数据除以5后取余的结果为0、1、2、3、4,则对应的特征位可以有5种表示:000、001、010、011、100,另外,可以将数据库命名为“数据库0”、“数据库1”、“数据库2”、“数据库3”和“数据库4”。例如,当第二属性数据对应的十进制数分别为10、21、12、33和34时,分别对应的二进制数的最后三位分别为000、001、010、011和100,除以5后取余的结果为0、1、2、3、4,则根据取余结果可以将5个第二属性数据分别划分到5个数据库中。

[0061] 数据库的个数也可以称之为模,其本质是第二属性数据的最后几个位(例如2位)决定这行数据落在哪个库上,这两个位就是分库数据基因,也就是特征位。

[0062] 在一个实施例中,可以根据第二属性数据的个数来设置数据库的个数,例如一个编号的第二属性数据切分到一个数据库中,那么如果有32个第二属性数据,则对应建立32个数据库。而一个第二属性数据可能对应多个第一属性数据,添加了特征位后的第三属性数据和第二属性数据具有相同的数据基因,应该切分至同一个数据库中,所以一个数据库



中例如可以有一个第二属性数据和对应的多个第三属性数据。

[0063] 因为特征位的位数是由数据库的个数决定的,所以当数据库的个数为32时,(32-1)对应的最简单的二进制数表示为“11111”,选取的特征位的位数就为5,对应有31个特征位的表示,包含相同特征位的第三属性数据和第二属性数据均应划分到同一数据库中。

[0064] 在另一个实施例中,从第二属性数据对应的二进制数的末尾取特征位,然后将特征位添加至第一属性数据对应的二进制数的末尾,组成第三属性数据对应的二进制数。上述过程中,根据第二属性数据和特征位来切分数据库,是通过取余方式进行的,主要是依赖于特征位,所以在第一属性数据进行数据库归类时,也可以采用取余方式,所以也要将特征位设置在第一属性数据对应的二进制数的末尾,从而对应的十进制数与第二属性数据具有同样的特性,在取余时,也可以将带有相同特征位的第三属性数据和第二属性数据切分至同一数据库中。

[0065] 在步骤S205中,将带有特征位的第三属性数据除以数据库的个数取余的结果相同的所有第三属性数据切分到同一个数据库中。上述步骤中描述到,一个第二属性数据可能对应多个第一属性数据,将第二属性数据中提取的数据基因添加至与第二属性数据具有关联关系的每一个第一属性数据中,组成多个与第二属性数据具有关联关系的第三属性数据,然后进行同样的取余操作,由于含有相同的特征位,所以相同特征位对应的取余结果相同,就可以把所有含有相同特征位的第三属性数据切分至同一个数据库中。

[0066] 本实施例中采用特征位进行数据库分库,巧妙地利用取余的方式,使得所有带有相同的特征位的第三属性数据和第二属性数据均能划分至同一个数据库中,便于后续的数据查询,计算简便,分库结果简洁明了,节约了计算资源,提高了服务的稳定性。

[0067] 在一个实施例中,采用水平切分方式进行数据库的切分。水平切分分为库内分表和分库分表,是根据表内数据内在的逻辑关系,将同一个表按不同的条件分散到多个数据库或多个表中,每个表中只包含一部分数据,从而使得单个表的数据量变小,达到分布式的效果。本发明实施例的操作方法在水平分库的场景下能够提高按照其它属性查询数据的性能,基于数据基因水平切分数据库的数据库操作方法,能够提升系统的稳定性和负载能力。

[0068] 在另一个可选的实施例中,第一属性数据为订单数据,第二属性数据为买家数据,第三属性数据为添加特征位后的订单数据,即查询时需要用的订单数据。以下基于图3详细说明买家数据和订单数据的数据库分库过程。

[0069] 图3示出了本发明实施例中的数据组建过程的示意图。结合图2,从第二属性数据的末尾取几位特征位,添加至第一属性数据的末尾组成第三属性数据,然后就可以进行数据库的切分,当第一属性数据为订单数据,第二属性数据为买家数据时,对应的数据组建过程即为图3所示。

[0070] 如图3所示,301表示第一属性数据,302表示第二属性数据,303表示特征位,304表示第三属性数据。第一属性数据为订单数据,第二属性数据为买家数据,第三属性数据为加入特征为以后的订单数据。将第三属性数据对应的订单数据简称为oid,代表唯一字段,将第二属性数据对应的买家数据简称为buyer\_id,代表其它字段,按照数据基因水平分库的原理,需要把其它字段buyer\_id的数据基因,融入到301中。

[0071] 步骤401表示取分库基因得到过程,例如,buyer\_uid=1,二进制数为0000 0000 0000 0001,从buyer\_uid中取特征位,然后通过buyer\_uid分库,假设分为4个库,采用

buyer\_uid%4的方式来进行数据库路由,数据库的个数即为模,图中模为4,其本质是buyer\_uid的最后2个位(bit)决定这行数据落在哪个库上,这2个位(bit),就是分库数据基因。所以,买家数据302buyer\_uid的最后两位为01,则分库基因是01,分库基因就是特征位303。

[0072] 步骤402和步骤403是表示将特征位303融入第一属性数据301中生成第三属性数据304的过程,例如,将特征位303表示的基因“01”添加到第一属性数据301对应的二进制数的末尾,生成第三属性数据304对应的二进制数。

[0073] 在生成订单数据oid的二进制数据时,先使用一种分布式服务全局ID算法生成与第一属性数据301对应的二进制数,为oid的前62位(bit)。在不分表的情况下,数据的唯一ID可以通过数据库自增ID来生成,不需要业务中进行实现,但如果由于数据量不断的增大,会对数据进行分库,分表,这样原来的数据库自增ID就用不上了。所以在这种情况下,需要采用分布式服务全局ID算法生成一个特有的服务全局ID,即分布式服务全局ID,例如:twitter开源的snowflake算法。

[0074] 然后将分库基因,即特征位303加入到第一属性数据301对应的62位分布式服务全局ID的末尾,形成最后2个位(bit),拼装成最终64位(bit)的新的订单数据oid,新数据即第三属性数据304对应的64位二进制数的末尾两位为01。

[0075] 由于买家数据buyer\_uid=1的用户可能购买多种商品,则一个买家数据可以对应多个订单数据,一个buyer\_uid对应的所有的oid的末端均带有分库数据基因,让同一个buyer\_uid下的所有订单数据都含有相同的数据基因01,分库时就可以落在同一个数据库中。

[0076] 通过这种方法,可以保证同一个用户下的所有订单数据oid,都落在同一个数据库上,oid的最后2位都相同,于是通过buyer\_uid%4能够定位到数据库,通过oid%4也能定位到数据库,两种属性都可以直接查找到相应的数据库,提升数据查找效率。

[0077] 在一个实施例中,数据库操作方法还包括:数据库分库后,分别采用不同的属性进行数据查询。具体的数据查询过程参见图4。

[0078] 图4示出了本发明实施例中数据库查询的操作示意图。如图所示,4个大圆柱DB0、DB1、DB2和DB3代表4个不同的数据库,订单数据oid和买家数据buyer\_uid组成的总订单数据表示为Order,oid和buyer\_uid可以分别作为两个数据列的名称。

[0079] 在本实施例的数据库操作方法中,数据的存储和查询例如采用SQL编程语言,如图4所示,步骤411表示插入数据的过程,执行“inset into Order(oid,buyer\_id,...) values(9,1,...)”之后,订单数据与买家数据组成一个大的总订单数据Order。步骤510表示分库操作,是将总订单数据Order按照oid取余水平切分的方式切分到四个库DB0、DB1、DB2、DB3中,实线箭头所示的过程是将数据进行数据库分库的过程,即根据oid除以4取余的方式切分数据库,并将订单数据存储到库DB1、DB2、DB3、DB4之中的一个。然后将buyer\_uid也对应存储至相应的数据库中,划分结束后,每一个数据库中,都含有两个订单数据oid和一个买家数据buyer\_uid。

[0080] 此时的数据库满足按照两个查询需求(oid和buyer\_id)进行数据查询的要求。查询时,也可以采用SQL语言,例如,步骤412和步骤413,发送指令:

[0081] select\*from Order where oid=9

[0082] `select*from Order where buyer_id=1`

[0083] 步骤412是根据订单数据oid定位数据库,步骤413是根据买家数据buyer\_uid定位数据库。计算机执行查询指令,虚线所示的过程是数据的查询过程,根据定位,两种查询方法都可以直接定位到同一个数据库中。

[0084] 所以通过这种方法可以保证同一个用户下的所有订单数据oid,都落在同一个数据库上,oid的最后2位都相同,于是通过buyer\_uid%4能够定位到数据库,通过oid%4也能定位到数据库。

[0085] 应该指出的是,本发明不仅限于上述实施例提供的规则,其他规则,也可以实践本发明实施例提供的数据库操作方法。

[0086] 图5示出了本发明实施例中的数据库操作系统的结构图。

[0087] 该数据库操作系统500包括数据获取单元501、特征位选择单元502、数据组建单元503和合并单元504。

[0088] 数据获取单元501用于获取第一属性数据和与之对应的第二属性数据;特征位选择单元502用于从第二属性数据对应的二进制数中选择至少一位作为特征位;数据组建单元503用于组合第一属性数据对应的二进制数和特征位得到第三属性数据的二进制数;合并单元504用于将所有包含相同特征位的第二属性数据和第三属性数据切分到同一数据库中。

[0089] 在一个实施例中,特征位选择单元502从所述第二属性数据对应的二进制数的末尾取所述特征位。

[0090] 在本实施例中,数据库操作系统500首先获取具有相互关系的第一属性数据和第二属性数据,从第二属性数据中选取特征位并将其添加至第一属性数据中,组成第三属性数据,第三属性数据是由第一属性数据变形得来,与第一属性数据包含相同的数据属性,然后将所有包含相同特征位的第三属性数据和第二属性数据切分到同一数据库中。通过基于特征位切分数据库,使得具有相互关系的第三属性数据和第二属性数据能切分至同一个数据库中,因此不但能满足按照第一属性来进行的查询,而且能够满足按照第二属性来进行快速查询,极大地提高了数据库数据的查询效率,同时提升了服务器和数据库操作系统的稳定性和负载能力。

[0091] 图6示出了汇总的数据库操作系统600的结构图。图6所示的实施例在图5的实施例的基础上增加了数据库切分单元601和归类单元602。

[0092] 特征位选择单元502从第二属性数据对应的二进制数的末尾取特征位,而数据库切分单元601用于根据第二属性数据除以需要切分的数据库的个数取余的结果切分数据库。

[0093] 在一个实施例中,数据组建单元503将特征位添加至第一属性数据对应的二进制数的末尾,组成第三属性数据的二进制数。则归类单元602用于将带有特征位的第三属性数据除以数据库的个数取余的结果相同的所有第三属性数据切分到同一个数据库中。在一个实施例中,当数据库的个数为N时,特征位的位数为(N-1)对应的最简单的二进制数的位数,N为大于等于1的整数。

[0094] 应该理解,本发明实施例的系统和方法是对应的,因此,在系统的描述中以相对简略的方式进行。

[0095] 图7示出了本发明实施例的数据库操作装置的结构图。图7示出的设备仅仅是一个示例,不对本发明实施例的功能和使用范围构成任何限制。

[0096] 参考图7,该数据库操作装置700包括通过总线连接的处理器701、存储器702和输入输出设备703。存储器702包括只读存储器(ROM)和随机访问存储器(RAM),存储器702内存储有执行系统功能所需的各种计算机指令和数据,处理器701从存储器702中读取各种计算机指令以执行各种适当的动作和处理。输入输出设备包括键盘、鼠标等的输入部分;包括诸如阴极射线管(CRT)、液晶显示器(LCD)等以及扬声器等的输出部分;包括硬盘等的存储部分;以及包括诸如LAN卡、调制解调器等的网络接口卡的通信部分。存储器702还存储有以下的计算机指令以完成本发明实施例的数据库操作方法规定的操作:获取第一属性数据和与之对应的第二属性数据;从第二属性数据对应的二进制数中选择至少一位作为特征位;组合第一属性数据对应的二进制数和特征位得到第三属性数据的二进制数;以及将所有包含相同特征位的第二属性数据和第三属性数据切分到同一数据库中。

[0097] 相应地,本发明实施例提供一种计算机可读存储介质,该计算机可读存储介质存储有计算机指令,所述计算机指令被执行时实现上述数据库操作方法所规定的操作。

[0098] 附图中的流程图、框图图示了本发明实施例的系统、方法、装置的可能的体系框架、功能和操作,流程图和框图上的方框可以代表一个模块、程序段或仅仅是一段代码,所述模块、程序段和代码都是用来实现规定逻辑功能的可执行指令。也应当注意,所述实现规定逻辑功能的可执行指令可以重新组合,从而生成新的模块和程序段。因此附图的方框以及方框顺序只是用来更好的图示实施例的过程和步骤,而不应以此作为对发明本身的限制。

[0099] 系统的各个模块或单元可以通过硬件、固件或软件实现。软件例如包括采用JAVA、C/C++/C#、SQL等各种编程语言形成的编码程序。虽然在方法以及方法图例中给出本发明实施例的步骤以及步骤的顺序,但是所述步骤实现规定的逻辑功能的可执行指令可以重新组合,从而生成新的步骤。所述步骤的顺序也不应该仅仅局限于所述方法以及方法图例中的步骤顺序,可以根据功能的需要随时进行调整。例如将其中的某些步骤并行或按照相反顺序执行。

[0100] 根据本发明的系统和方法可以部署在单个或多个服务器上。例如,可以将不同的模块分别部署在不同的服务器上,形成专用服务器。或者,可以在多个服务器上分布式部署相同的功能单元、模块或系统,以减轻负载压力。所述服务器包括但不限于在同一个局域网以及通过Internet连接的多个PC机、PC服务器、刀片服务器、超级计算机等。

[0101] 以上所述仅为本发明的优选实施例,并不用于限制本发明,对于本领域技术人员而言,本发明可以有各种改动和变化。凡在本发明的精神和原理之内所作的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。

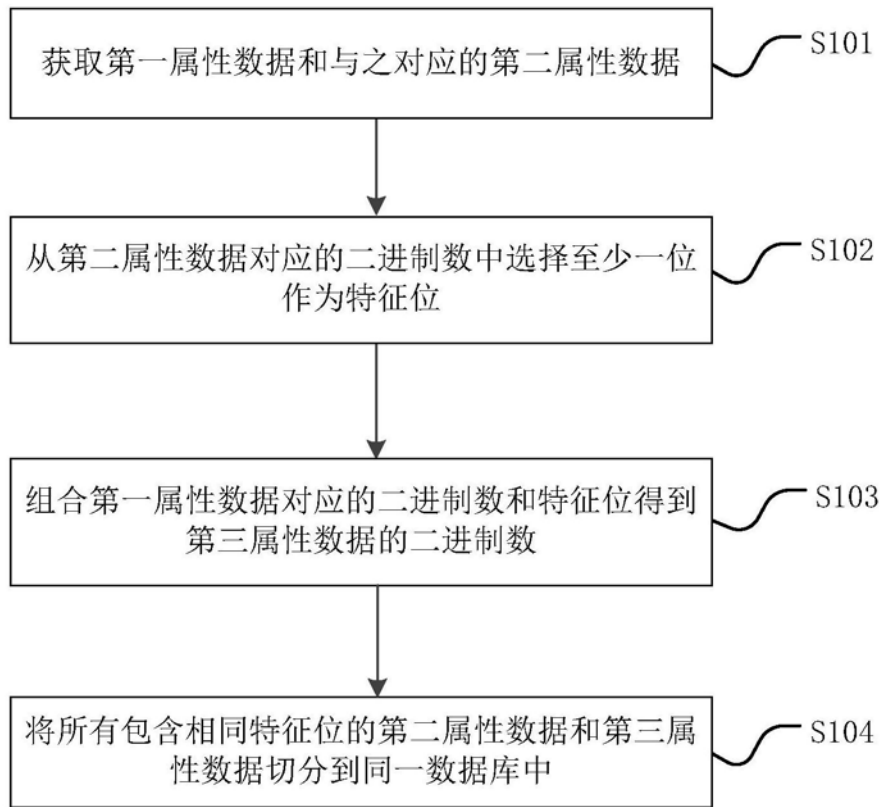


图1

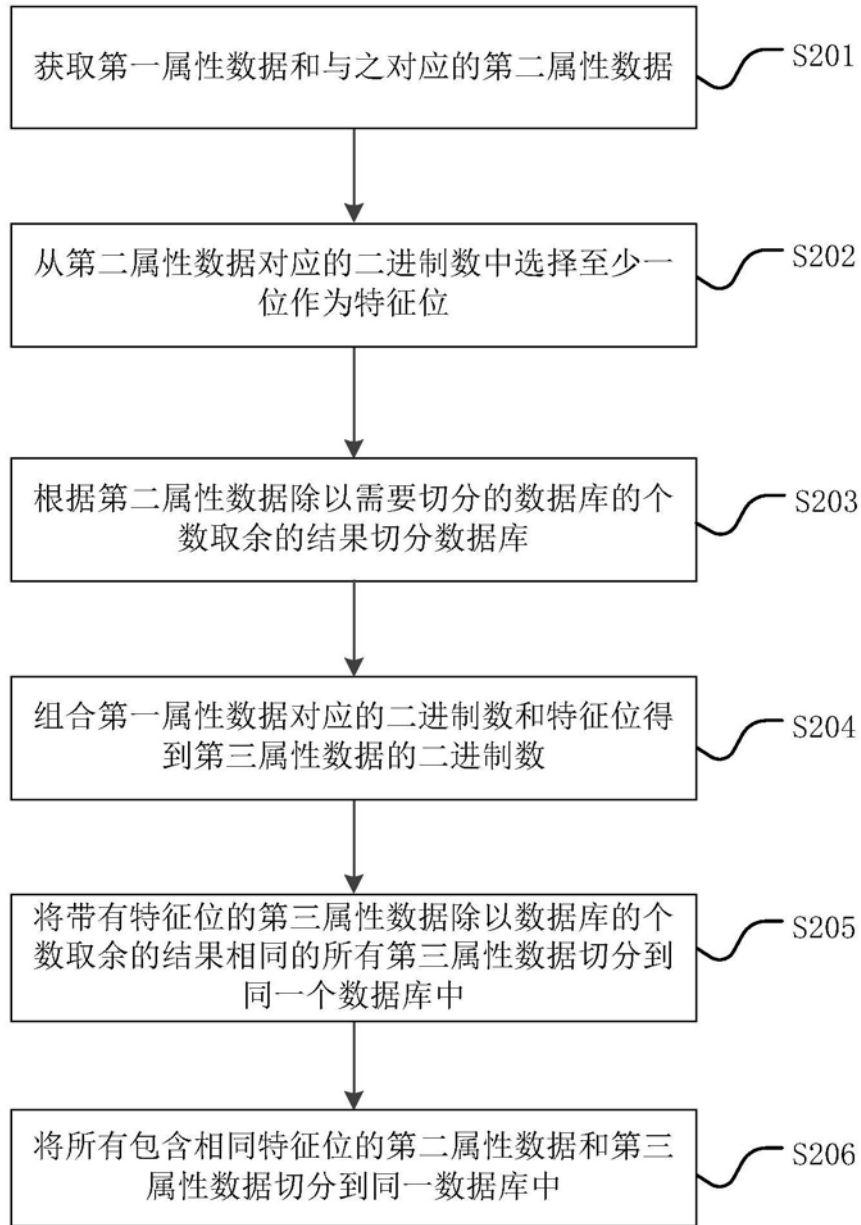


图2

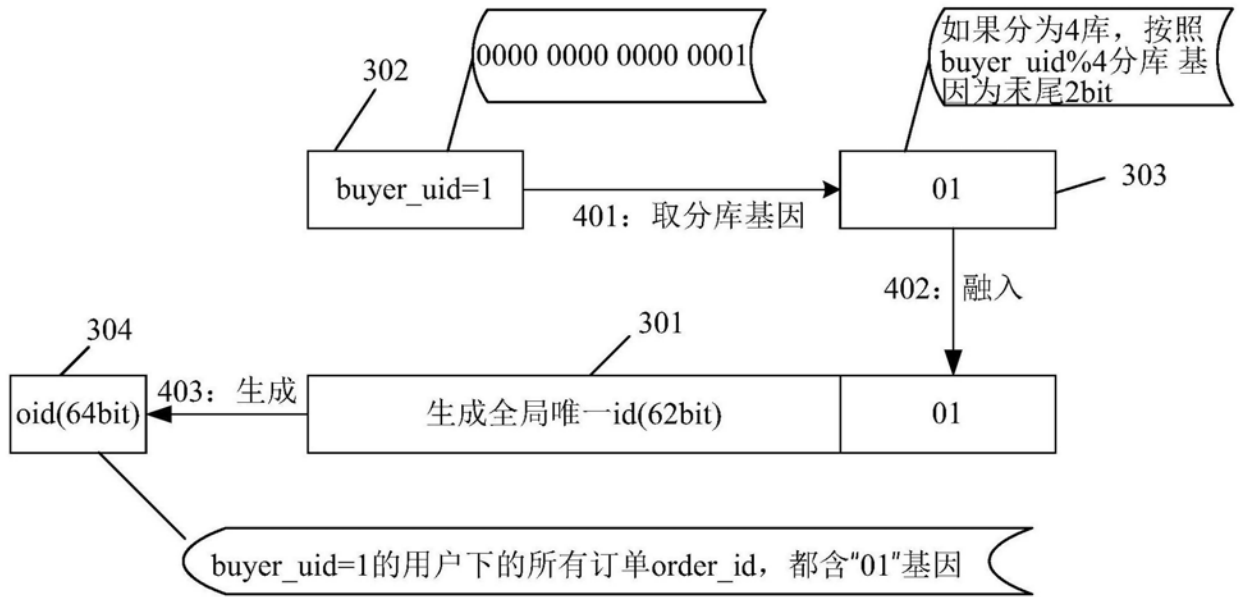


图3

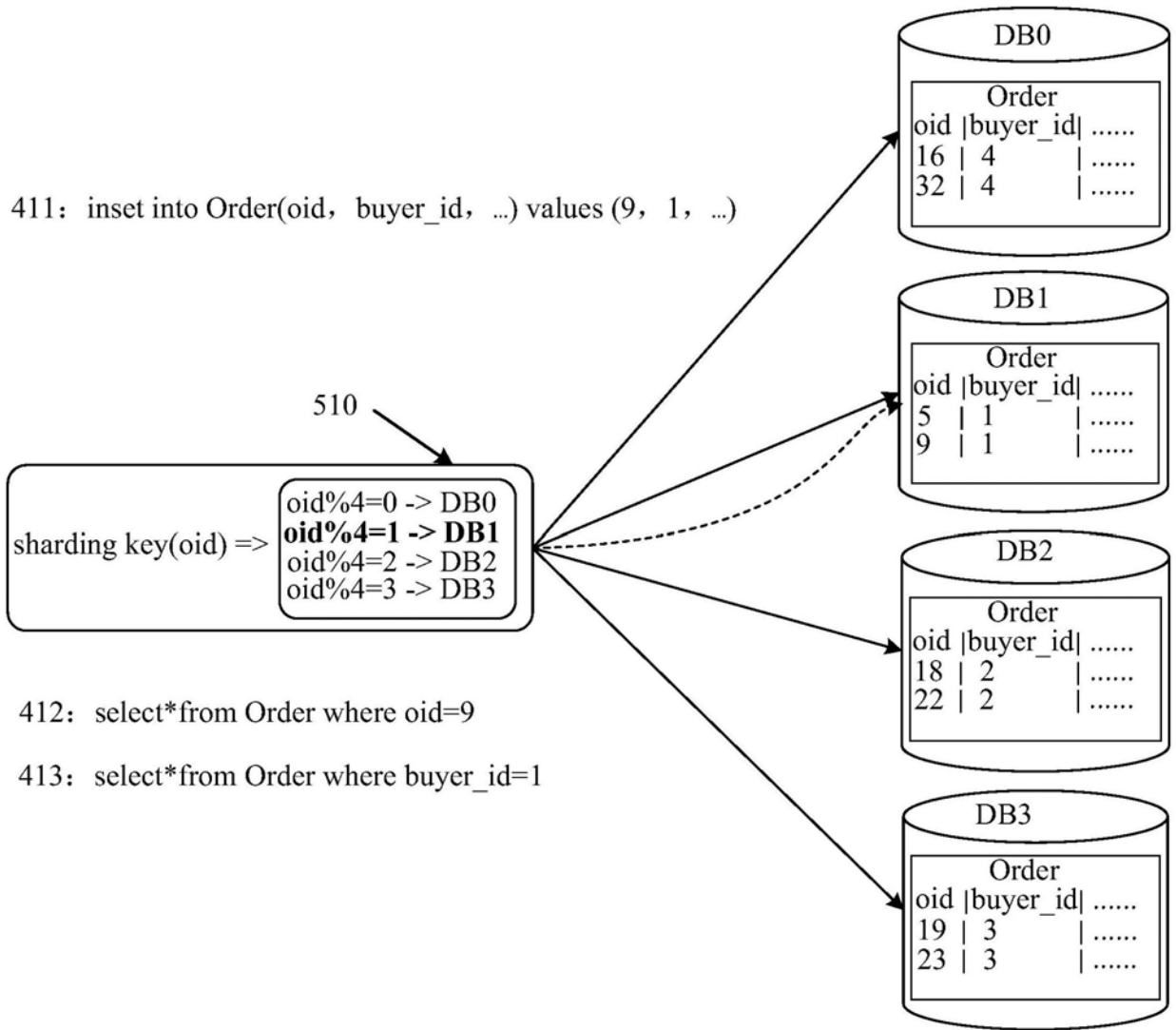


图4



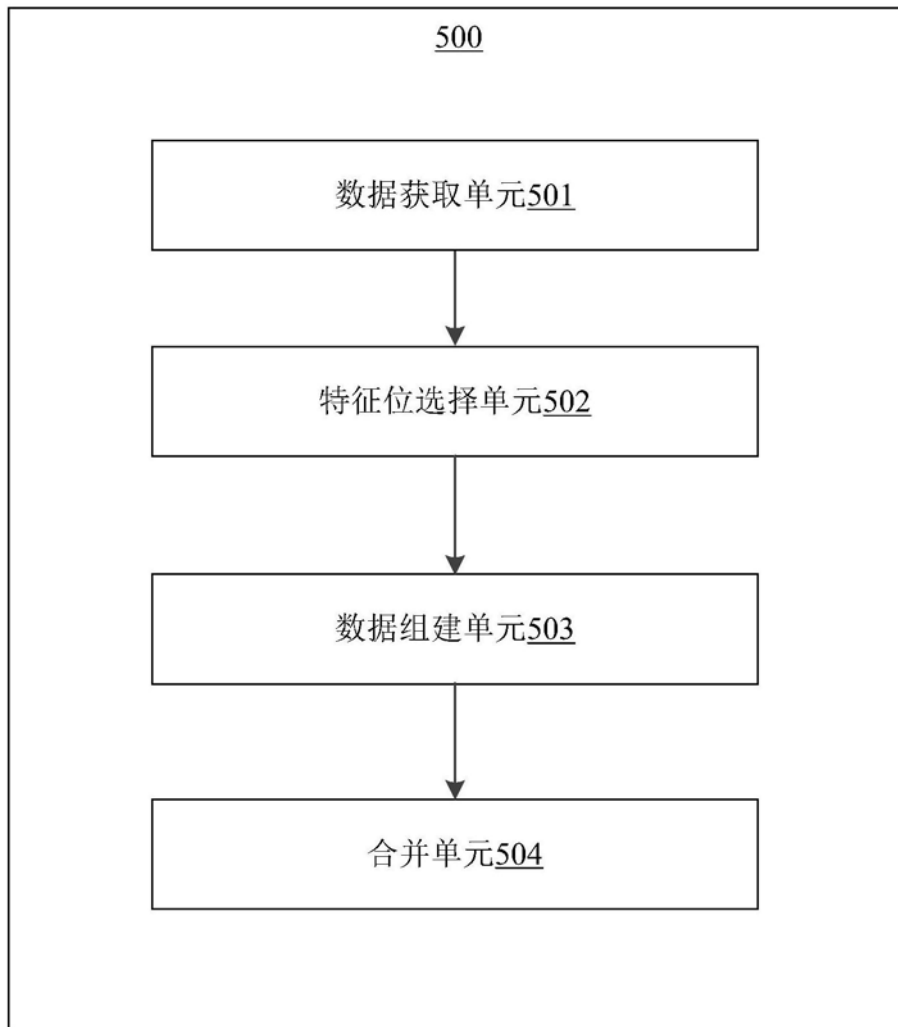


图5

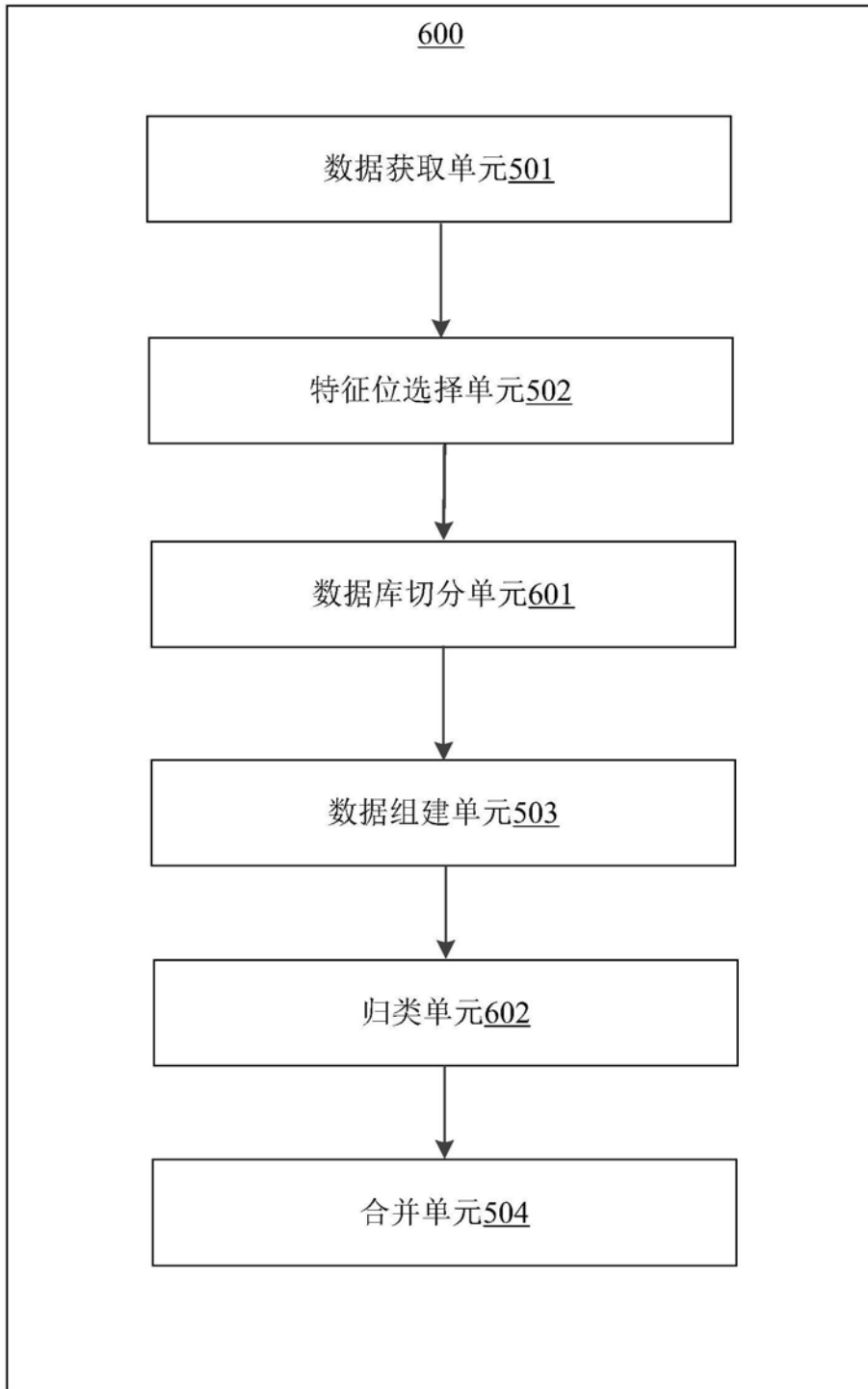


图6

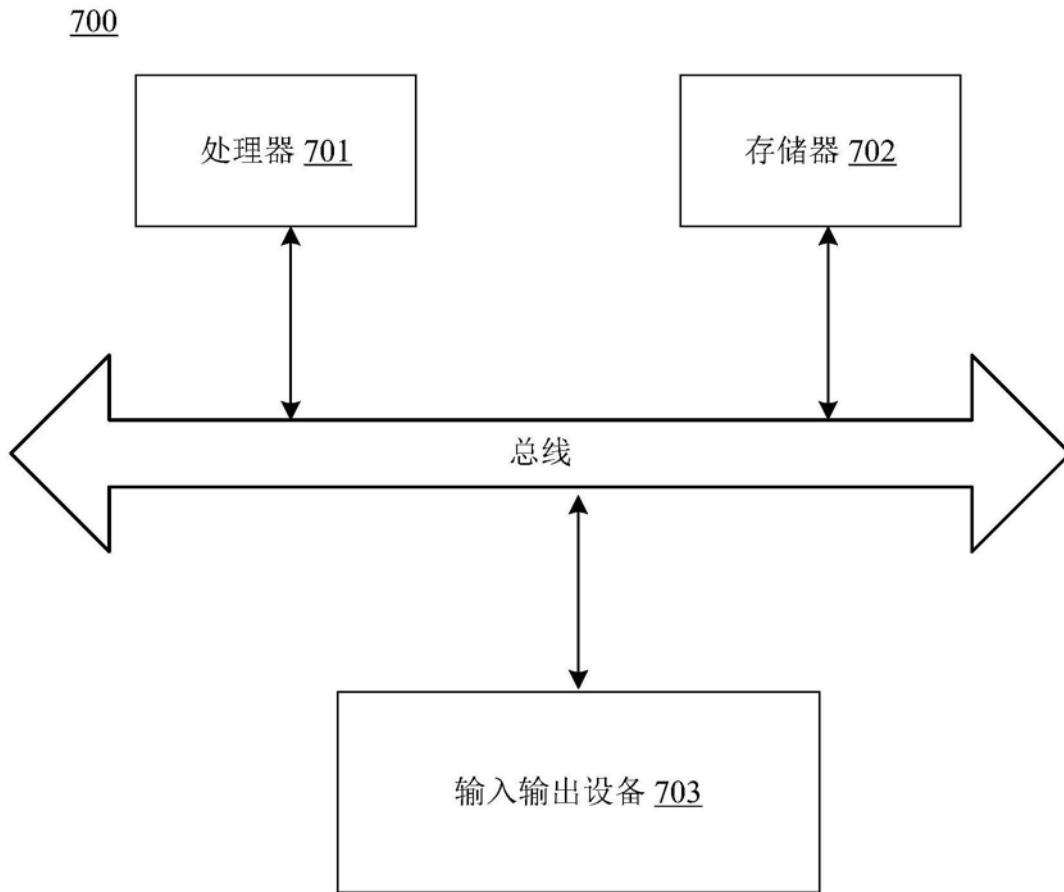


图7