



(12) 发明专利申请

(10) 申请公布号 CN 117669709 A

(43) 申请公布日 2024. 03. 08

(21) 申请号 202211097889.0

(22) 申请日 2022.09.08

(71) 申请人 华为技术有限公司

地址 518129 广东省深圳市龙岗区坂田华为总部办公楼

(72) 发明人 吕政 冯霖

(74) 专利代理机构 北京同达信恒知识产权代理有限公司 11291

专利代理师 田小倩

(51) Int. Cl.

G06N 5/01 (2023.01)

G06F 7/08 (2006.01)

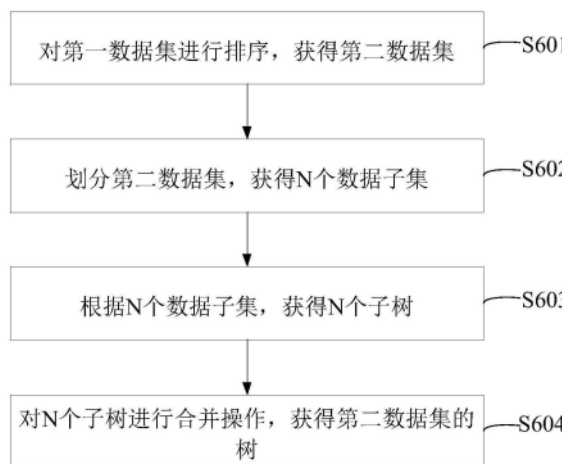
权利要求书4页 说明书22页 附图10页

(54) 发明名称

一种数据处理方法及装置

(57) 摘要

本申请提供一种数据处理方法及装置,涉及计算机技术领域。在所述方法中,加速装置对第一数据集进行排序,获得第二数据集,其中,所述第一数据集和所述第二数据集均包括多个数据,划分所述第二数据集,获得多个数据子集,并对多个数据子集分别进行处理,获得多个子树,其中一个数据子集对应一个子树,并根据这多个子树获得第二数据集的树,由于加速装置创建树的过程中,无需处理器或加速装置生成指令以及译码指令等,简化了创建树的流程,提高了创建树的效率,也就提高了数据处理的效率,并且,加速装置可以并行获得多个子树,能够提高创建树的效率,也就能提高加速装置处理数据的效率。



1. 一种数据处理方法,其特征在于,应用于加速装置,所述方法包括:

对第一数据集进行排序,获得第二数据集,其中,所述第一数据集和所述第二数据集均包括多个数据;

划分所述第二数据集,获得N个数据子集,所述N个数据子集中的每个数据子集包括所述第二数据集中连续排列的至少两个数据,N为大于或等于2的整数;

根据所述N个数据子集中的第i个数据子集,确定第i个子树,所述i取遍1到N中的任意一个正整数,共获得N个子树,其中,所述第i个子树用于查找所述第i个数据子集包括的数据;

对所述N个子树进行合并操作,获得所述第二数据集的树,所述树用于查找所述第二数据集包括的数据。

2. 根据权利要求1所述的方法,其特征在于,对第一数据集进行排序,获得第二数据集,包括:

确定第一数据组,所述第一数据组为对所述第一数据集中的部分数据进行排序后的结果;

将所述第一数据组中的每个第一位置上的数据,与第二数据组中与所述每个第一位置对应的位置上的目标数据进行比较,确定目标位置,其中,所述目标位置是指在所述第一数据组中用于插入所述目标数据的位置,所述第二数据组与所述第一数据组包括的数据的个数相同,且所述第二数据组包括的任一数据均为所述目标数据,所述目标数据为所述第一数据集中除了所述部分数据之外的数据;

将所述目标数据插入到所述目标位置中,获得所述第二数据集。

3. 根据权利要求1或2所述的方法,其特征在于,所述方法还包括:

确定第三数据组,其中,所述第三数据组包括多个目标键,所述目标键为目标分组的键,所述第三数据组与所述第二数据集包括的数据的个数相同;

将所述第二数据集中的每个第二位置上的数据,与所述第三数据组中的与所述每个第二位置对应的位置上的数据进行比较,确定所述第二数据集中与目标键匹配的至少一个数据;

将所述至少一个数据确定为所述目标分组;

将所述目标分组的信息和所述目标分组写入外存中,其中,所述目标分组的信息包括所述目标键。

4. 根据权利要求3所述的方法,其特征在于,所述目标分组的信息还包括所述至少一个数据包括的数据的个数、所述至少一个数据的最大值、所述至少一个数据的最小值以及所述至少一个数据的求和结果中的一种或多种。

5. 根据权利要求1-4任一项所述的方法,其特征在于,所述方法还包括:

确定目标执行计划,所述目标执行计划用于指示对所述第一数据集执行的操作。

6. 根据权利要求1-5任一项所述的方法,其特征在于,所述方法还包括:

从处理器接收第一请求,其中,所述第一请求用于请求对所述第一数据集进行处理;根据所述第一请求获取所述第一数据集。

7. 根据权利要求6所述的方法,其特征在于,所述方法还包括:

所述加速装置和所述处理器均设置在计算设备中,所述加速装置通过快捷外围部件互

连标准PCIe与所述处理器连接。

8. 一种数据处理方法,其特征在于,应用于加速装置中,所述方法包括:

对第一数据集进行排序,获得第二数据集,其中,所述第一数据集和所述第二数据集均包括多个数据;

确定第三数据组,其中,所述第三数据组包括多个第一键,所述第一键为目标分组的键,所述第三数据组与所述第二数据集包括的数据的个数相同;

将所述第二数据集中的每个第二位置上的数据,与所述第三数据组中的与所述每个第二位置对应的位置上的数据进行比较,确定所述第二数据集中与目标键匹配的至少一个数据;

将所述至少一个数据确定为所述目标分组;

将所述目标分组的信息和所述目标分组写入外存中,其中,所述目标分组的信息包括所述目标键。

9. 根据权利要求8所述的方法,其特征在于,所述目标分组的信息还包括所述至少一个数据的数据个数、所述至少一个数据的最大值、所述至少一个数据的最小值以及所述至少一个数据的求和结果中的一种或多种。

10. 一种数据处理方法,其特征在于,应用于加速装置中,所述方法包括:

获取第一数据集,所述第一数据集包括多个数据;

确定第一数组,所述第一数组为所述第一数据集中的部分数据进行排序的结果;

将所述第一数组中的每个第一位置上的数据,与第二数据组中与所述每个第一位置对应的位置上的目标数据进行比较,确定目标位置,其中,所述目标位置是指在所述第一数据组中用于插入所述目标数据的位置,所述第二数据组与所述第一数据组包括的数据的个数相同,且所述第二数据组包括的任一数据均为所述目标数据,所述目标数据为所述第一数据集中除了所述部分数据之外的数据;

将所述目标数据插入到所述目标位置中,获得所述第二数据集。

11. 一种加速装置,其特征在于,包括:

排序模块,用于对第一数据集进行排序,获得第二数据集,其中,所述第一数据集和所述第二数据集均包括多个数据;

树创建模块,用于划分所述第二数据集,获得N个数据子集,所述N个数据子集中的每个数据子集包括所述第二数据集中连续排列的至少两个数据,N为大于或等于2的整数,根据所述N个数据子集中的第i个数据子集,确定第i个子树,所述i取遍1到N中的任意一个正整数,共获得N个子树,其中,所述第i个子树用于查找所述第i个数据子集包括的数据,以及对所述N个子树进行合并操作,获得所述第二数据集的树,所述树用于查找所述第二数据集包括的数据。

12. 根据权利要求11所述的装置,其特征在于,所述排序模块具体用于:

确定第一数据组,所述第一数据组为对所述第一数据集中的部分数据进行排序后的结果;

将所述第一数据组中的每个第一位置上的数据,与第二数据组中与所述每个第一位置对应的位置上的目标数据进行比较,确定目标位置,其中,所述目标位置是指在所述第一数据组中用于插入所述目标数据的位置,所述第二数据组与所述第一数据组包括的数据的个

数相同,且所述第二数据组包括的任一数据均为所述目标数据,所述目标数据为所述第一数据集中除了所述部分数据之外的数据;

将所述目标数据插入到所述目标位置中,获得所述第二数据集。

13. 根据权利要求11或12所述的装置,其特征在于,所述装置还包括分组模块,用于:

确定第三数据组,其中,所述第三数据组包括多个目标键,所述目标键为目标分组的键,所述第三数据组与所述第二数据集包括的数据的个数相同;

将所述第二数据集中的每个第二位置上的数据,与所述第三数据组中的与所述每个第二位置对应的位置上的数据进行比较,确定所述第二数据集中与目标键匹配的至少一个数据;

将所述至少一个数据确定为所述目标分组;

将所述目标分组的信息和所述目标分组写入外存中,其中,所述目标分组的信息包括所述目标键。

14. 根据权利要求13所述的装置,其特征在于,所述目标分组的信息还包括所述至少一个数据包括的数据的个数、所述至少一个数据的最大值、所述至少一个数据的最小值以及所述至少一个数据的求和结果中的一种或多种。

15. 根据权利要求11-14任一项所述的装置,其特征在于,所述装置还包括执行计划确定模块,用于确定目标执行计划,所述目标执行计划用于指示对所述第一数据集执行的操作。

16. 根据权利要求15所述的装置,其特征在于,所述装置还包括数据获取模块,用于从处理器接收第一请求,其中,所述第一请求用于请求对所述第一数据集进行处理;根据所述第一请求获取所述第一数据集。

17. 根据权利要求16所述的装置,其特征在于,所述装置和所述处理器均设置在第一设备中,所述装置通过快捷外围部件互连标准PCIe与所述处理器连接。

18. 一种加速装置,其特征在于,包括:

排序模块,用于对第一数据集进行排序,获得第二数据集,其中,所述第一数据集和所述第二数据集均包括多个数据;

分组模块,用于确定第三数据组,其中,所述第三数据组包括多个第一键,所述第一键为目标分组的键,所述第三数据组与所述第二数据集包括的数据的个数相同,将第三数据组中的每个第二位置上的第一键,与所述第二数据集中的与所述每个第二位置对应位置上的数据进行比较,确定所述第二数据集中与目标键匹配的至少一个数据,以及将所述至少一个数据确定为所述目标分组;将所述目标分组的信息和所述目标分组写入硬盘中,其中,所述目标分组的信息包括所述目标键。

19. 一种加速装置,其特征在于,包括:

数据获取模块,获取第一数据集,所述第一数据集包括多个数据;

排序模块,确定第一数组,所述第一数组为所述第一数据集中的部分数据进行排序的结果,将所述第一数组中的每个第一位置上的数据,与第二数据组中与所述每个第一位置对应的位置上的目标数据进行比较,确定目标位置,其中,所述目标位置是指在所述第一数据组中用于插入所述目标数据的位置,所述第二数据组与所述第一数据组包括的数据的个数相同,且所述第二数据组包括的任一数据均为所述目标数据,所述目标数据为所述第

一数据集中除了所述部分数据之外的数据,以及将所述目标数据插入到所述目标位置中,获得所述第二数据集。

20. 一种加速装置,其特征在于,包括处理器和供电电路,所述供电电路用于为所述处理器供电,所述处理器用于执行如权利要求1-7、8-9或10任一项所述的方法。

21. 一种计算设备,其特征在于,包括如权利要求20所述的加速装置。

22. 一种计算设备,其特征在于,所述计算设备包括加速装置和处理器;

所述处理器,用于向所述加速装置发送第一请求,其中,所述第一请求用于请求所述加速装置对第一数据集进行处理;

所述加速装置,用于执行如权利要求1-7、8-9或10任一项所述的方法,以对所述第一数据集进行处理。

23. 一种包含指令的计算机程序产品,其特征在于,当所述指令被计算设备集群运行时,使得所述计算设备集群执行如权利要求1-7、8-9或10任一项所述的方法。

24. 一种计算机可读存储介质,其特征在于,所述存储介质中存储有计算机程序或指令,当所述计算机程序或指令被通信装置执行时,实现如权利要求1-7、8-9或10任一项所述的方法。

## 一种数据处理方法及装置

### 技术领域

[0001] 本申请涉及计算机技术领域,尤其涉及一种数据处理方法及装置。

### 背景技术

[0002] 数据能够为企业提供确定决策的依据。企业的计算设备采集的数据一般是零散的,通常需要计算设备对数据进行处理,获得数据处理结果,如此,企业的工作人员可以根据数据处理结果,确定相应的决策。

[0003] 目前,数据处理一般是由计算设备中的中央处理器(central processing unit, CPU)执行的。CPU在对每个数据进行处理过程中,需要单独译码获得数据对应的指令,进而执行指令,CPU处理单个数据的过程较为繁琐,CPU处理单个数据的效率较低。另外,一旦需要处理的数据较多,这会进一步导致CPU处理数据的效率较低。

### 发明内容

[0004] 本申请提供一种数据处理方法及装置,用于提高数据处理效率。

[0005] 第一方面,本申请实施例提供一种数据处理方法,所述方法可以由加速装置执行,加速装置例如为片上系统(system on chip, SoC)或数据处理单元(data processing unit, DPU)等,所述方法包括:对第一数据集进行排序,获得第二数据集,其中,所述第一数据集和所述第二数据集均包括多个数据;划分所述第二数据集,获得N个数据子集,所述N个数据子集中的每个数据子集包括所述第二数据集中连续排列的至少两个数据,N为大于或等于2的整数;根据所述N个数据子集中的第i个数据子集,确定第i个子树,所述i取遍1到N中的任意一个正整数,共获得N个子树,其中,所述第i个子树用于查找所述第i个数据子集包括的数据;对所述N个子树进行合并操作,获得所述第二数据集的树,所述树用于查找所述第二数据集包括的数据。

[0006] 在本申请实施例中,加速装置可以对排序后的数据集进行划分,获得多个数据子集,并分别根据多个数据子集,确定多个子树,进而根据这多个子树确定树,加速装置在确定树的过程中,无需处理器或加速装置生成指令,以及译码指令等过程,减少了数据处理过程,有利于提高数据处理的效率。并且,加速装置可以并行根据多个数据子集,确定多个子树,从而提高加速装置多个子树的效率,也就提高了确定数据集的树的效率,由于创建树的过程属于数据处理过程,因此提高了确定树的效率,也就提高了处理数据的效率。可选的,加速装置可以是计算设备中的独立于处理器的器件,由专门的加速装置对数据进行处理,可减少计算设备中的处理器的负载。

[0007] 在一种可能的实施方式中,加速装置对第一数据集进行排序的方式具体包括:确定第一数据组,所述第一数据组为对所述第一数据集中的部分数据进行排序后的结果;将所述第一数据组中的每个第一位置上的数据,与第二数据组中与所述每个第一位置对应的位置上的目标数据进行比较,确定目标位置,其中,所述目标位置是指在所述第一数据组中用于插入所述目标数据的位置,所述第二数据组与所述第一数据组包括的数据的个数相

同,且所述第二数据组包括的任一数据均为所述目标数据,所述目标数据为所述第一数据集中除了所述部分数据之外的数据;将所述目标数据插入到所述目标位置中,获得所述第二数据集。

[0008] 在上述实施方式中,加速装置可以将已排序的数据(即第一数据组)中的各个数据并行与待排序的数据(目标数据)进行比较,而不是将一个数据与多个数据进行逐一比较,即相当于并行将目标数据与多个数据进行比较,提高加速装置对数据集进行排序的效率,对数据集进行排序的过程也属于数据处理过程,因此也就能够提高数据处理的效率。并且,在加速装置进行排序的过程中,无需加速装置或处理器生成和译码指令等过程,也有利于提高数据处理的效率。

[0009] 在一种可能的实施方式中,加速装置还可以对第二数据集进行分组,一种加速装置对第二数据集进行分组的过程包括:确定第三数据组,其中,所述第三数据组包括多个目标键,所述目标键为目标分组的键,所述第三数据组与所述第二数据集包括的数据的个数相同;将所述第二数据集中的每个第二位置上的数据,与所述第三数据组中的与所述每个第二位置对应的位置上的数据进行比较,确定所述第二数据集中与目标键匹配的至少一个数据;将所述至少一个数据确定为所述目标分组;将所述目标分组的信息和所述目标分组写入外存中,其中,所述目标分组的信息包括所述目标键。

[0010] 在上述实施方式中,加速装置可以将待分组的数据(即第二数据集)均与目标分组的分组键进行比较,从而可以一次性确定与目标键匹配的至少一个数据,也就确定了目标分组包括这至少一个数据,提高了加速装置对数据集进行分组的效率,对数据集进行分组的过程也属于数据处理过程,因此也就能够提高数据处理的效率。

[0011] 在一种可能的实施方式中,所述目标分组的信息还可以包括所述至少一个数据的数据个数、所述至少一个数据的最大值、所述至少一个数据的最小值以及所述至少一个数据的求和结果中的一种或多种。

[0012] 在上述实施方式中,加速装置在确定目标分组之后,还可以确定目标分组的信息,以便于为用户提供更多的目标分组中的数据统计信息。

[0013] 在一种可能的实施方式中,所述方法还包括:确定目标执行计划,所述目标执行计划用于指示对所述第一数据集执行的操作。

[0014] 在上述实施方式中,加速装置可以确定目标执行计划,目标执行计划包括对第一数据集执行的操作,例如上述中的排序操作、分组操作或创建树操作等一种或多种。如此,便于加速装置后续按照目标执行计划,对第一数据集执行相应的操作。

[0015] 在一种可能的实施方式中,所述方法还包括:从处理器接收第一请求,其中,所述第一请求用于请求对所述第一数据集进行处理;根据所述第一请求获取所述第一数据集。

[0016] 在上述实施方式中,加速装置可以从处理器接收第一请求,根据第一请求获取第一数据集,提供了一种加速装置获取第一数据集的方式。并且,无需处理器对第一数据集进行处理,有利于减少处理器的处理量。

[0017] 在一种可能的实施方式中,所述加速装置和所述处理器可以均设置在计算设备中,所述加速装置可以通过PCIe与所述处理器连接。

[0018] 在上述实施方式中,加速装置与处理器可通过PCIe连接,无需单独设计加速装置与处理器之间的连接方式,有利于降低计算设备的成本。另外,加速装置可以替代处理器对

数据执行排序操作、分组操作和创建树操作等一种或多种,有利于减少处理器的处理量。

[0019] 第二方面,本申请实施例提供一种数据处理方法,该方法可以由加速装置执行,所述方法包括:对第一数据集进行排序,获得第二数据集,其中,所述第一数据集和所述第二数据集均包括多个数据;确定第三数据组,其中,所述第三数据组包括多个第一键,所述第一键为目标分组的键,所述第三数据组与所述第二数据集包括的数据的个数相同;将所述第二数据集中的每个第二位置上的数据,与所述第三数据组中的与所述每个第二位置对应的位置上的数据进行比较,确定所述第二数据集中与目标键匹配的至少一个数据;将所述至少一个数据确定为所述目标分组;将所述目标分组的信息和所述目标分组写入外存中,其中,所述目标分组的信息包括所述目标键。

[0020] 在一种可能的实施方式中,所述目标分组的信息可以还包括所述至少一个数据的数据个数、所述至少一个数据的最大值、所述至少一个数据的最小值以及所述至少一个数据的求和结果中的一种或多种。

[0021] 在一种可能的实施方式中,对第一数据集进行排序,获得第二数据集,包括:确定第一数据组,所述第一数据组为对所述第一数据集中的部分数据进行排序后的结果;将所述第一数据组中的每个第一位置上的数据,与第二数据组中与所述每个第一位置对应的位置上的目标数据进行比较,确定目标位置,其中,所述目标位置是指在所述第一数据组中用于插入所述目标数据的位置,所述第二数据组与所述第一数据组包括的数据的个数相同,且所述第二数据组包括的任一数据均为所述目标数据,所述目标数据为所述第一数据集中除了所述部分数据之外的数据;将所述目标数据插入到所述目标位置中,获得所述第二数据集。

[0022] 在一种可能的实施方式中,所述方法包括:划分所述第二数据集,获得N个数据子集,所述N个数据子集中的每个数据子集包括所述第二数据集中连续排列的至少两个数据,N为大于或等于2的整数;根据所述N个数据子集中的第i个数据子集,确定第i个子树,所述i取遍1到N中的任意一个正整数,共获得N个子树,其中,所述第i个子树用于查找所述第i个数据子集包括的数据;对所述N个子树进行合并操作,获得所述第二数据集的树,所述树用于查找所述第二数据集包括的数据。

[0023] 在一种可能的实施方式中,所述方法还包括:确定目标执行计划,所述目标执行计划用于指示对所述第一数据集执行的操作。

[0024] 在一种可能的实施方式中,所述方法还包括:从处理器接收第一请求,其中,所述第一请求用于请求对所述第一数据集进行处理;根据所述第一请求获取所述第一数据集。

[0025] 在一种可能的实施方式中,所述方法还包括:所述加速装置和所述处理器均设置在计算设备中,所述加速装置通过PCIe与所述处理器连接。

[0026] 第三方面,本申请实施例提供一种数据处理方法,所述方法可以由加速装置执行,所述方法包括:获取第一数据集,所述第一数据集包括多个数据;确定第一数组,所述第一数组为所述第一数据集中的部分数据进行排序的结果;将所述第一数组中的每个第一位置上的数据,与第二数据组中与所述每个第一位置对应的位置上的目标数据进行比较,确定目标位置,其中,所述目标位置是指在所述第一数据组中用于插入所述目标数据的位置,所述第二数据组与所述第一数据组包括的数据的个数相同,且所述第二数据组包括的任一数据均为所述目标数据,所述目标数据为所述第一数据集中除了所述部分数据之外的数

据;将所述目标数据插入到所述目标位置中,获得所述第二数据集。

[0027] 在一种可能的实施方式中,所述方法包括:确定第三数据组,其中,所述第三数据组包括多个第一键,所述第一键为目标分组的键,所述第三数据组与所述第二数据集包括的数据的个数相同;将所述第二数据集中的每个第二位置上的数据,与所述第三数据组中的与所述每个第二位置对应的位置上的数据进行比较,确定所述第二数据集中与目标键匹配的至少一个数据;将所述至少一个数据确定为所述目标分组;将所述目标分组的信息和所述目标分组写入外存中,其中,所述目标分组的信息包括所述目标键。

[0028] 在一种可能的实施方式中,所述目标分组的信息还包括所述至少一个数据包括的数据的个数、所述至少一个数据的最大值、所述至少一个数据的最小值以及所述至少一个数据的求和结果中的一种或多种。

[0029] 在一种可能的实施方式中,所述方法包括:划分所述第二数据集,获得N个数据子集,所述N个数据子集中的每个数据子集包括所述第二数据集中连续排列的至少两个数据,N为大于或等于2的整数;根据所述N个数据子集中的第i个数据子集,确定第i个子树,所述i取遍1到N中的任意一个正整数,共获得N个子树,其中,所述第i个子树用于查找所述第i个数据子集包括的数据;对所述N个子树进行合并操作,获得所述第二数据集的树,所述树用于查找所述第二数据集包括的数据。

[0030] 在一种可能的实施方式中,所述方法还包括:确定目标执行计划,所述目标执行计划用于指示对所述第一数据集执行的操作。

[0031] 在一种可能的实施方式中,所述方法还包括:从处理器接收第一请求,其中,所述第一请求用于请求对所述第一数据集进行处理;根据所述第一请求获取所述第一数据集。

[0032] 在一种可能的实施方式中,所述方法还包括:所述加速装置和所述处理器均设置在计算设备中,所述加速装置通过PCIe与所述处理器连接。

[0033] 第四方面,本申请实施例提供一种加速装置,所述装置包括:排序模块,用于对第一数据集进行排序,获得第二数据集,其中,所述第一数据集和所述第二数据集均包括多个数据;树创建模块,用于划分所述第二数据集,获得N个数据子集,所述N个数据子集中的每个数据子集包括所述第二数据集中连续排列的至少两个数据,N为大于或等于2的整数,根据所述N个数据子集中的第i个数据子集,确定第i个子树,所述i取遍1到N中的任意一个正整数,共获得N个子树,其中,所述第i个子树用于查找所述第i个数据子集包括的数据,以及对所述N个子树进行合并操作,获得所述第二数据集的树,所述树用于查找所述第二数据集包括的数据。

[0034] 在一种可能的实施方式中,所述排序模块具体用于:确定第一数据组,所述第一数据组为对所述第一数据集中的部分数据进行排序后的结果;将所述第一数据组中的每个第一位置上的数据,与第二数据组中与所述每个第一位置对应的位置上的目标数据进行比较,确定目标位置,其中,所述目标位置是指在所述第一数据组中用于插入所述目标数据的位置,所述第二数据组与所述第一数据组包括的数据的个数相同,且所述第二数据组包括的任一数据均为所述目标数据,所述目标数据为所述第一数据集中除了所述部分数据之外的数据;将所述目标数据插入到所述目标位置中,获得所述第二数据集。

[0035] 在一种可能的实施方式中,所述装置还包括分组模块,所述分组模块具体用于:确定第三数据组,其中,所述第三数据组包括多个目标键,所述目标键为目标分组的键,所述

第三数据组与所述第二数据集包括的数据的个数相同;将所述第二数据集中的每个第二位置上的数据,与所述第三数据组中的与所述每个第二位置对应的位置上的数据进行比较,确定所述第二数据集中与目标键匹配的至少一个数据;将所述至少一个数据确定为所述目标分组;将所述目标分组的信息和所述目标分组写入外存中,其中,所述目标分组的信息包括所述目标键。

[0036] 在一种可能的实施方式中,所述目标分组的信息还包括所述至少一个数据包括的数据的个数、所述至少一个数据的最大值、所述至少一个数据的最小值以及所述至少一个数据的求和结果中的一种或多种。

[0037] 在一种可能的实施方式中,所述装置还包括执行计划确定模块,所述执行计划确定模块用于:确定目标执行计划,所述目标执行计划用于指示对所述第一数据集执行的操作。

[0038] 在一种可能的实施方式中,所述装置包括数据获取模块,所述数据获取模块用于:从处理器接收第一请求,其中,所述第一请求用于请求对所述第一数据集进行处理;根据所述第一请求获取所述第一数据集。

[0039] 第五方面,本申请实施例提供一种加速装置,所述装置包括:排序模块,用于对所述第一数据集进行排序,获得第二数据集,其中,所述第一数据集和所述第二数据集均包括多个数据;确定第三数据组,其中,所述第三数据组包括多个第一键,所述第一键为目标分组的键,所述第三数据组与所述第二数据集包括的数据的个数相同;分组模块,用于将所述第二数据集中的每个第二位置上的数据,与所述第三数据组中的与所述每个第二位置对应的位置上的数据进行比较,确定所述第二数据集中与目标键匹配的至少一个数据,将所述至少一个数据确定为所述目标分组,以及将所述目标分组的信息和所述目标分组写入外存中,其中,所述目标分组的信息包括所述目标键。

[0040] 在一种可能的实施方式中,所述目标分组的信息还包括所述至少一个数据包括的数据的个数、所述至少一个数据的最大值、所述至少一个数据的最小值以及所述至少一个数据的求和结果中的一种或多种。

[0041] 在一种可能的实施方式中,所述排序模块具体用于:确定第一数据组,所述第一数据组为对所述第一数据集中的部分数据进行排序后的结果;将所述第一数据组中的每个第一位置上的数据,与第二数据组中与所述每个第一位置对应的位置上的目标数据进行比较,确定目标位置,其中,所述目标位置是指在所述第一数据组中用于插入所述目标数据的位置,所述第二数据组与所述第一数据组包括的数据的个数相同,且所述第二数据组包括的任一数据均为所述目标数据,所述目标数据为所述第一数据集中除了所述部分数据之外的数据;将所述目标数据插入到所述目标位置中,获得所述第二数据集。

[0042] 在一种可能的实施方式中,所述装置还包括树创建模块,所述树创建模块用于划分所述第二数据集,获得N个数据子集,所述N个数据子集中的每个数据子集包括所述第二数据集中连续排列的至少两个数据,N为大于或等于2的整数;根据所述N个数据子集中的第i个数据子集,确定第i个子树,所述i取遍1到N中的任意一个正整数,共获得N个子树,其中,所述第i个子树用于查找所述第i个数据子集包括的数据;对所述N个子树进行合并操作,获得所述第二数据集的树,所述树用于查找所述第二数据集包括的数据。

[0043] 在一种可能的实施方式中,所述装置还包括执行计划确定模块,所述执行计划确

定模块用于：确定目标执行计划，所述目标执行计划用于指示对所述第一数据集执行的操作。

[0044] 在一种可能的实施方式中，所述装置还包括数据获取模块，所述数据获取模块用于：从处理器接收第一请求，其中，所述第一请求用于请求对所述第一数据集进行处理；根据所述第一请求获取所述第一数据集。

[0045] 在一种可能的实施方式中，所述装置和所述处理器均设置在计算设备中，所述装置通过快捷外围部件互连标准PCIe与所述处理器连接。

[0046] 第六方面，本申请实施例提供一种加速装置，所述装置包括：数据获取模块，用于获取第一数据集，所述第一数据集包括多个数据；排序模块，用于确定第一数组，所述第一数组为所述第一数据集中的部分数据进行排序的结果；将所述第一数组中的每个第一位置上的数据，与第二数据组中与所述每个第一位置对应的位置上的目标数据进行比较，确定目标位置，其中，所述目标位置是指在所述第一数据组中用于插入所述目标数据的位置，所述第二数据组与所述第一数据组包括的数据的个数相同，且所述第二数据组包括的任一数据均为所述目标数据，所述目标数据为所述第一数据集中除了所述部分数据之外的数据；将所述目标数据插入到所述目标位置中，获得所述第二数据集。

[0047] 在一种可能的实施方式中，所述装置还包括分组模块，用于：确定第三数据组，其中，所述第三数据组包括多个第一键，所述第一键为目标分组的键，所述第三数据组与所述第二数据集包括的数据的个数相同；将所述第二数据集中的每个第二位置上的数据，与所述第三数据组中的与所述每个第二位置对应的位置上的数据进行比较，确定所述第二数据集中与目标键匹配的至少一个数据；将所述至少一个数据确定为所述目标分组；将所述目标分组的信息和所述目标分组写入外存中，其中，所述目标分组的信息包括所述目标键。

[0048] 在一种可能的实施方式中，所述目标分组的信息还包括所述至少一个数据包括的数据的个数、所述至少一个数据的最大值、所述至少一个数据的最小值以及所述至少一个数据的求和结果中的一种或多种。

[0049] 在一种可能的实施方式中，所述装置还包括树创建模块，用于：划分所述第二数据集，获得N个数据子集，所述N个数据子集中的每个数据子集包括所述第二数据集中连续排列的至少两个数据，N为大于或等于2的整数；根据所述N个数据子集中的第i个数据子集，确定第i个子树，所述i取遍1到N中的任意一个正整数，共获得N个子树，其中，所述第i个子树用于查找所述第i个数据子集包括的数据；对所述N个子树进行合并操作，获得所述第二数据集的树，所述树用于查找所述第二数据集包括的数据。

[0050] 在一种可能的实施方式中，所述装置还包括执行计划确定模块，用于：确定目标执行计划，所述目标执行计划用于指示对所述第一数据集执行的操作。

[0051] 在一种可能的实施方式中，所述装置还包括数据获取模块，用于：从处理器接收第一请求，其中，所述第一请求用于请求对所述第一数据集进行处理；根据所述第一请求获取所述第一数据集。

[0052] 在一种可能的实施方式中，所述装置和所述处理器均设置在计算设备中，所述装置通过快捷外围部件互连标准PCIe与所述处理器连接。

[0053] 第七方面，本申请实施例提供一种加速装置，包括：处理器和供电电路；所述供电电路为所述处理器供电，所述处理器用于执行第一方面至第三方面中任一的数据处理方

法。

[0054] 在一种可能的实施方式中,所述加速装置还包括其他部件,例如,天线,输入输出模块,接口等等。这些部件可以是硬件,软件,或者软件和硬件的结合。

[0055] 第八方面,本申请实施例提供一种计算设备,所述计算设备包括第七方面的加速装置。

[0056] 第九方面,本申请实施例提供一种计算设备,所述计算设备包括加速装置和处理器;所述处理器用于向所述加速装置发送第一请求,其中,所述第一请求用于请求所述加速装置对第一数据集进行处理;所述加速装置用于执行第一方面至第三方面中任一的数据处理方法,以对所述第一数据集进行处理。

[0057] 第十方面,本申请实施例提供一种芯片系统,该芯片系统包括:处理器和接口。其中,该处理器用于从该接口调用并运行指令,当该处理器执行该指令时,实现上述第一方面至第三方面中任一的数据处理方法。

[0058] 第十一方面,提供一种计算机可读存储介质,该计算机可读存储介质用于存储计算机程序或指令,当其被运行时,实现上述第一方面至第三方面中任一的数据处理方法。

[0059] 第十二方面,提供一种包含指令的计算机程序产品,当其在计算机上运行时,实现上述第一方面至第三方面中任一的数据处理方法。

[0060] 关于第二方面至第十二方面的有益效果,可参照第一方面论述的有益效果,此处不再列举。

## 附图说明

[0061] 图1为本申请实施例适用的一种场景示意图;

[0062] 图2为本申请实施例适用的另一种场景示意图;

[0063] 图3为本申请实施例提供了一种数据处理系统的结构示意图;

[0064] 图4为本申请实施例提供的另一种数据处理系统的结构示意图;

[0065] 图5为本申请实施例提供的又一种数据处理系统的结构示意图;

[0066] 图6为本申请实施例提供了一种数据处理方法的流程示意图一;

[0067] 图7为本申请实施例提供了一种对第一数据集执行转换操作的原理示意图;

[0068] 图8为本申请实施例提供了一种对第一数据集进行排序的过程示意图;

[0069] 图9为本申请实施例提供了一种创建树的过程示意图;

[0070] 图10为本申请实施例提供了一种创建树的原理示意图;

[0071] 图11为本申请实施例提供了一种对数据进行分组的过程示意图;

[0072] 图12为本申请实施例提供了一种页的结构示意图;

[0073] 图13至图16为本申请实施例提供的几种数据处理方法的流程示意图;

[0074] 图17至图18为本申请实施例提供的两种加速装置的结构示意图。

## 具体实施方式

[0075] 为了使本申请实施例的目的、技术方案和优点更加清楚,下面将结合附图对本申请实施例作进一步地详细描述。

[0076] 以下,对本申请实施例中的部分用语进行解释说明,以便于本领域技术人员理解。

[0077] 1、终端设备,是一种具有无线收发功能的设备,可以是固定设备,移动设备、手持设备、穿戴设备、车载设备,或内置于上述设备中的无线装置(例如,通信模块或芯片系统等)。所述终端设备用于连接人,物,机器等,可广泛用于各种场景,例如包括但不限于以下场景:蜂窝通信、设备到设备通信(device-to-device,D2D)、车到一切(vehicle to everything,V2X)、机器到机器/机器类通信(machine-to-machine/machine-type communications,M2M/MTC)、物联网(internet of things,IoT)、虚拟现实(virtual reality,VR)、增强现实(augmented reality,AR)、工业控制(industrial control)、无人驾驶(self driving)、远程医疗(remote medical)、智能电网(smart grid)、智能家具、智能办公、智能穿戴、智能交通,智慧城市(smart city)、无人机、机器人等场景的终端设备。所述终端设备有时可称为用户设备(user equipment,UE)、终端、接入站、UE站、远方站、无线通信设备、或用户装置等等。

[0078] 2、节点,可以是单个设备。本申请实施例所示的节点还可以是逻辑概念,例如为软件模块,本申请实施例对此不作具体限定。

[0079] 本申请实施例中,对于名词的数目,除非特别说明,表示“单数名词或复数名词”,即“一个或多个”。“至少一个”是指一个或者多个,“多个”是指两个或两个以上。“和/或”,描述关联对象的关联关系,表示可以存在三种关系,例如,A和/或B,可以表示:单独存在A,同时存在A和B,单独存在B的情况,其中A,B可以是单数或者复数。字符“/”一般表示前后关联对象是一种“或”的关系。例如,A/B,表示:A或B。“以下至少一项(个)”或其类似表达,是指的这些项中的任意组合,包括单项(个)或复数项(个)的任意组合。例如,a,b,或c中的至少一项(个),表示:a,b,c,a和b,a和c,b和c,或a和b和c,其中a,b,c可以是单个,也可以是多个。

[0080] 请参照图1,为本申请实施例适用的一种场景示意图。或者,图1也可理解为本申请实施例提供的一种数据处理系统的结构示意图。如图1所示,该场景包括终端设备110、运行在终端设备110中的客户端111和计算设备120。其中,终端设备110与计算设备120之间可以通过以太网或无线网络(如无线保真(wireless fidelity,WIFI)或第5代(5<sup>th</sup> generation,5G)通信)技术相互通信。

[0081] 客户端111可为软件模块或者程序。用户可以通过客户端111向计算设备120发起数据处理请求,数据处理请求例如为用于请求读取数据库中的数据的数据读取请求,或为用于请求向数据库中写入数据的数据写入请求。其中,计算设备120可以对数据库进行读操作或写操作。计算设备120可以根据数据处理请求,执行相应的数据处理。其中,计算设备120泛指具有处理能力的设备,例如服务器或终端设备等。

[0082] 请参照图2,为本申请实施例适用的一种场景示意图。或者,图2也可理解为本申请实施例提供的一种数据处理系统的结构示意图。与图1不同的是,图2中的数据处理系统220包括一个或多个计算设备221,数据处理系统220与客户端211通信,以处理客户端211的数据处理请求。

[0083] 如图2所示,该场景包括终端设备210、运行在终端设备210中的客户端211和数据处理系统220。终端设备210与数据处理系统220之间可以通过以太网或无线网络(如WIFI、第5代(5<sup>th</sup> generation,5G)通信)技术相互通信。

[0084] 在图2中是以数据处理系统220包括的计算设备221的数量为三个为例进行示例,实际不限制数据处理系统220包括的计算设备221的数量。其中,数据处理系统220包括的任

意两个计算设备221的结构可以是相同的。

[0085] 在一种可能的实施方式中,数据处理系统220可为分布式数据处理系统。其中,分布式数据处理系统可包括中心化的分布式数据处理系统和去中心化的分布式数据处理系统。

[0086] 请参照图3,为本申请实施例提供的一种分布式数据处理系统的结构示意图。图3例如可为一种中心化的分布式数据处理系统的结构示意图。如图3所示,数据处理系统300包括主节点310和一个或多个从节点320。主节点310也可以称为控制节点或管理节点等,从节点320也可以称为工作节点。主节点310可与客户端相互通信,主节点310和任一从节点320之间也可相互通信。

[0087] 主节点310用于接收来自客户端的数据处理请求,以及将客户端的数据处理请求分配给一个或多个从节点320中的某个从节点320处理。从节点320用于对数据处理请求进行处理。其中,客户端的含义可参照前文论述的内容,客户端例如为图2中的客户端211。其中,主节点310和从节点320均可通过计算设备实现,例如一个主节点310为一个计算设备,一个从节点320为一个计算设备。

[0088] 请参照图4,为本申请实施例提供的另一种分布式数据处理系统的结构示意图。或者,图4例如可为一种去中心化的分布式数据处理系统的结构示意图。与图3不同的是,图4中的任意两个节点的功能是相同的。

[0089] 如图4所示,数据处理系统400包括多个节点(如图4所示的第一节点410、第二节点420和第三节点430),第一节点410、第二节点420和第三节点430中的任一节点均可与客户端之间相互通信,并且用于接收来自客户端的数据处理请求,以及处理数据处理请求。其中,第一节点410、第二节点420和第三节点430中的任一节点可通过计算设备实现,例如任一节点为一个计算设备。

[0090] 请参照图5,为本申请实施例提供的一种计算设备的结构示意图。图5中的计算设备500可以是图1中的计算设备120、图2中的计算设备221、图3中的主节点310、图3中的从节点320、图4中的第一节点410、图4中的第二节点420或图4中的第三节点430。

[0091] 如图5所示,计算设备500包括处理器510、加速装置520、内存530和外存540。处理器510、加速装置520、内存530和外存540之间可通过总线550通信。其中,总线550在图5中以粗线表示,总线550可以为基于快捷外围部件互连标准(peripheral component interconnect express,PCIe)的线路。

[0092] 处理器510可以为中央处理器(central processing unit,CPU)、专用集成电路(application specific integrated circuit,ASIC)、现场可编程门阵列(field programmable gate array,FPGA)、人工智能(artificial intelligence,AI)芯片、片上系统(system on chip,SoC)、复杂可编程逻辑器件(complex programmable logic device,CPLD)或图形处理器(graphics processing unit,GPU)。

[0093] 加速装置520为用于专门处理数据的器件,可以为SOC、DPU或智能网卡等。

[0094] 内存530可以包括易失性存储器(volatile memory),例如随机存取存储器(random access memory,RAM)或动态随机存取存储器(dynamic random access memory,DRAM)等,也可以包括非易失性存储器(non-volatile memory),例如存储级内存(storage class memory,SCM)等,或者易失性存储器与非易失性存储器的组合等。

[0095] 内存530还可以包括操作系统等其他运行进程所需的软件模块。操作系统可以为LINUX<sup>TM</sup>、UNIX<sup>TM</sup>或WINDOWS<sup>TM</sup>等。内存530中还可以存储数据库中的数据,如内存530中所存储的数据可以包括数据库中最最近写入的数据。可选的,当内存230中的数据量达到一定阈值时,处理器510可以将内存530中的数据存储在至外存540中,以进行持久化存储。在需要读取数据库的数据时,从外存540中读取的数据可以先存储在内存530,也可以是说,内存530中所存储的数据也可以包括从外存540中读取的数据。

[0096] 外存540也可以称为辅助存储器,外存540可以为非易失性存储器(non-volatile memory),例如只读存储器(read-only memory,ROM)、磁盘或硬盘(hard disk)等。外存540可以用于持久化存储数据。

[0097] 在内存530存储有数据库中的数据的情况下,内存530中所存储的数据库中的数据和外存540中所存储的数据库中的数据的格式(或称为存储方式)可以相同,也可以不同。例如,内存530和外存540中所存储的数据库中的数据均可以为行格式或以列格式。其中,行格式是指以行为准的格式(或称为存储方式),列格式是指以列为准的格式(或称为存储方式)。又例如,内存530所存储的数据库中的数据为行格式,外存540中所存储的数据库中的数据是为列格式。又例如,内存530所存储的数据库中的数据为列格式,外存540中所存储的数据库中的数据为行格式。

[0098] 请参照图6,为本申请实施例提供的一种数据处理方法的流程示意图。图6所示的数据处理方法可以由加速装置执行,图6所示的实施例涉及的加速装置例如可为图5中的加速装置520。图6所示的实施例涉及的加速装置例如可以设置在图1的计算设备120、图2的计算设备221、图3的主节点310、图3的从节点320、图4的第一节点410、图4的第二节点420、或者图4的第三节点430中。图6所示的数据处理方法包括如下步骤:

[0099] S601、加速装置对第一数据集进行排序,获得第二数据集。

[0100] 例如,加速装置从处理器接收第一请求,第一请求用于请求对第一数据集进行处理。第一请求可为前文中的数据处理请求的一种示例。加速装置接收第一请求,从而确定对第一数据集进行处理。

[0101] 其中,第一数据集包括多个数据,第一数据集可以是计算设备的数据库中的部分或全部数据。加速装置可以对计算设备中的数据库进行读操作或写操作等,计算设备例如为图1的计算设备120、图2的计算设备221、图3的主节点310、图3的从节点320、图4的第一节点410、图4的第二节点420、或者图4的第三节点430。其中,第一数据集的格式可以为行格式和/或列格式。其中,行格式和列格式的含义可参照前文。第一数据集中的数据的类型可以是任意的类型,例如,字符型或整型等。

[0102] 示例性的,第一请求包括第一数据集,如此,加速装置接收第一请求,也就获得了第一数据集。或者,第一请求包括第一地址,第一地址为存储第一数据集的地址,如此,加速装置在获得第一请求之后,可以根据第一地址,获取第一数据集。

[0103] 例如,第一数据集为数据库中的全部数据,第一请求包括数据库的地址(即第一地址),加速装置根据第一地址,从而获得第一数据集。

[0104] 请参照表1,为本申请实施例提供的一种第一数据集的示例。

[0105] 表1

[0106]

学生姓名	成绩
------	----

张三	89
李四	90
王五	80

[0107] 如上述表1所示的第一数据集的格式为行格式。

[0108] 可选的,第一数据集中的数据的类型可以是任意的,例如,第一数据集包括的数据的类型可以为整型、字符型型、时间或日期一种或多种类型。

[0109] 作为一个示例,加速装置可处理多种数据类型的数据,例如,加速装置可以支持对整型、字符型型、时间或日期一种或多种类型的数据进行处理。相应的,加速装置可以支持对多种类型的数据库执行写操作和/或读操作,例如,加速装置支持对PostgreSQL或MySQL执行写操作和/或读操作。PostgreSQL和MySQL为两种数据库。

[0110] 请参照表2,为本申请实施例提供的加速装置支持的数据的类型,以及支持的数据类型对应的数据库的类型的一种示例。

[0111] 表2

数据的类型	数据库的类型	描述
smallint	PostgreSQL、MySQL	smallint 表示 16 位整型
integer (int)	PostgreSQL、MySQL	int/integer 表示 32 位整型
bigint	PostgreSQL、MySQL	bigint 表示 64 位整型
numeric/decimal	PostgreSQL、MySQL	numeric/decimal 表示变长精度
varchar	PostgreSQL、MySQL	varchar 表示变长字符型
char(n)	PostgreSQL、MySQL	char(n)表示定长字符型
文本 (text)	PostgreSQL、MySQL	文本
日期 (date)	PostgreSQL、MySQL	日期
时间 (time)	PostgreSQL、MySQL	时间
带时区的时间 (time with time zone)	PostgreSQL	时间, 带时区
时间戳 (timestamp)	PostgreSQL、MySQL	日期和时间
带时区的时间戳	PostgreSQL	日期和时间, 带时区
(timestamp with time zone)		

[0112] 作为一个示例,第一数据集的格式可能不符合加速装置处理的需求,和/或第一数据集的类型不符合加速装置的需求,因此加速装置获取第一数据集之后,可以对第一数据集执行转换操作,获得转换后的第一数据集。其中,转换操作包括格式转换操作和/或类型转换操作。其中,格式转换操作是指将一种格式的数据转换为另一种格式的数据,类型转换操作是指将一种类型的数据转换为另一种类型的数据。

[0113] 下面对加速装置对第一数据集执行转换操作的情况进行介绍。

[0114] 情况一、加速装置确定第一数据集的信息,根据第一数据集的信息,对从数据库中获取的第一数据集执行转换操作。第一数据集的信息包括第一数据集的格式和/或类型的信息。情况一适用于加速装置之前没有处理过数据库中的数据的情况。例如,情况一适用于

加速装置通过复制(copy)命令从数据库中获取第一数据集的情况。

[0117] 请参照图7,为本申请实施例提供的一种对第一数据集执行转换操作的原理示意图。

[0118] S7.1、加速装置解析数据库的数据定义语言(data definition language,DDL),获得第一数据集的信息。

[0119] S7.2、加速装置根据第一数据集的信息,对第一数据集执行转换操作,从而获得转换后的第一数据集。转换操作的内容可参照前文论述的内容。

[0120] 情况二、加速装置对第一数据集直接执行转换操作。情况二适用于加速装置之前已处理过数据库中的数据的情况。例如,情况二适用于加速装置通过插入(inset)命令从数据库中获取第一数据集的情况。

[0121] 具体的,加速装置之前已经处理过数据库中的数据(为便于区分,加速装置之前已经处理过的数据称为历史数据),加速装置后又需对该数据库中的第一数据集进行处理,这种情况下,加速装置相当于已经获得了历史数据的信息,也就相当于获取了第一数据集的信息,因此加速装置可直接执行S7.3,即加速装置直接对第一数据集执行转换操作。如此,无需加速装置获取第一数据集的信息,有利于减少加速装置的处理量。

[0122] 作为另一个示例,如果第一数据集的格式和/或类型符合加速装置的处理需求,那么加速装置可以无需对第一数据集进行转换操作。

[0123] 在加速装置对第一数据集执行转换操作的情况下,加速装置可以对转换后的第一数据集进行排序操作,获得第二数据集。或者,加速装置无需对第一数据集执行转换操作的情况下,加速装置可以对第一数据集执行排序操作,获得第二数据集。加速装置对转换后的第一数据集或第一数据集执行排序操作的方式是相同的,下面以加速装置对第一数据集执行排序操作为例进行介绍。

[0124] 示例性的,加速装置将第一数据集中的已排序的多个数据与第一数据集中的待排序的数据进行比较,将未排序的数据插入到已排序的数据中,直到加速装置对第一数据集中的所有数据均排完序,则获得第二数据集。可选的,加速装置可以将已排序的多个数据同步与第一数据集中的待排序的数据进行比较。

[0125] 具体而言,加速装置对第一数据集中的部分数据进行排序,获得第一数据组。第一数据集中的部分数据为第一数据集中的多个数据,部分数据例如可为第一数据集包括的两个数据。第一数据组可以视为已排序的多个数据。本申请实施例中的数据组可以理解为具有一定顺序的多个数据的排列结果,一个数据组也可视为一个向量。

[0126] 加速装置确定第二数据组,第二数据组中的每个元素均为目标数据。目标数据为第一数据集中除了部分数据之外的数据。其中,目标数据视为第一数据集中待排序的数据。加速装置将第一数据组中的每个位置(为了便于区分,称为第一位置)上的数据,与第二数据组中与所述每个第一位置上对应的数据进行比较,这相当于对第一数据组和第二数据组进行了向量化比较,从而确定在第一数据组中插入目标数据的位置,即目标位置。加速装置将目标数据插入目标位置,从而获得第二数据集。

[0127] 其中,加速装置将第一数据组中的每个位置(为了便于区分,称为第一位置)上的数据,与第二数据组中与所述每个第一位置上对应的数据进行比较,可以具体理解为加速装置将第一数据组中的第一个位置上的数据与第二数据组中的第一个位置上的数据进行

比较,将第一数据组中的第二个位置上的数据与第二数据组中的第二个位置上的数据进行比较,以此类推。

[0128] 上述是以待排序的目标数据为一个进行说明,在待排序的目标数据包括多个的情况下,加速装置可以分别根据多个目标数据,确定多个第二数据组,其中,每个第二数据组包括的元素均为多个目标数据中的一个。加速装置可以确定多个第二数据组中的每个第二数据组中的目标数据对应的目标位置,其中,加速装置确定一个目标位置的方式可参照前述的内容,此处不再列举。加速装置将多个目标数据分别插入第一数据组中对应的目标位置中,从而获得第二数据集。

[0129] 示例性的,加速装置可以根据多个目标数据中的一个目标数据,确定一个第二数据组,以及确定第一数据组。加速装置将第一数据组中的每个第一位置上的数据与第二数据组中与所述每个第一位置上的数据进行比较,确定所述一个目标数据对应的目标位置。加速装置将所述一个目标数据插入所述一个目标位置之后,获得更新后的第一数据组。加速装置可以确定多个目标数据中的另一个目标数据对应的第二数据组,确定另一个目标数据对应的目标位置。加速装置将另一个目标数据插入更新后的第一数据组中,以此类推,直到加速装置处理完多个目标数据,获得最终更新后的第一数据组,所述最终更新后的第一数据组即为第二数据集。

[0130] 可选的,加速装置可以按照第一规则,对第一数据组和第二数据组进行比较。其中,第一规则,例如为按照数据从大到小的规则进行排序,或者为按照数据从大到小的规则进行排序。第一规则可以是被预配置在加速装置中的,或者加速装置自行确定的,例如,第一请求指示第一规则,加速装置接收第一请求之后,也就获得了第一规则。例如,第一请求包括比较函数,加速装置可根据比较函数确定第一规则。其中,比较函数用于表示排序数据所采用的函数。

[0131] 可选的,加速装置包括第一寄存器和第二寄存器。寄存器可用于存储数据组。由于寄存器用于存储数据组,因此寄存器也可以称为向量(或数组)寄存器。

[0132] 示例性的,在对第一数据集执行排序操作的过程中,加速装置可以将第一数据组写入第一寄存器,以及将第二数据组写入第二寄存器,加速装置通过第一寄存器和第二寄存器,对第一数据组中的每个第一位置上的数据和第二数据组中的与每个第一位置上的数据进行比较。

[0133] 例如,请参照图8,为本申请实施例提供的一种第二数据组和第一数据组的示意图。在图8中是以第一数据集包括:1、4、5、4、7、9、3和10为例。如图8所示,加速装置确定的第一数据组具体为:1、4、5、7、9和10;加速装置确定第二数据组具体为:3、3、3、3、3和3,即目标数据为3。

[0134] 加速装置可以将第一数据组中的每个第一位置上的数据与第二数据组中的与所述每个第一位置上的数据进行比较,在图8中以虚线表示比较。具体的,加速装置将第一数据组中的第一个位置上的数据(即1)与第二数据组中的第一个位置上的数据(即3)比较,将第一数据组中的第二个位置上的数据(即4)与第二数据组中的第二个位置上的数据(即3)进行比较,将第一数据组中的第三个位置上的数据(即5)与第二数据组中的第三个位置上的数据(即3)进行比较,将第一数据组中的第四个位置上的数据(即7)与第二数据组中的第四个位置上的数据(即3)进行比较,将第一数据组中的第五个位置上的数据(即9)与第二数

据组中的第五个位置上的数据(即3)进行比较,以及将第一数据组中的第六个位置上的数据(即10)与第二数据组中的第六个位置上的数据(即3)进行比较。也就相当于加速装置将第一数据集中的待排序的数据并行与已排序的多个数据进行了比较,确定目标位置,即为第一数据组中的第二个位置。加速装置可以将目标数据插入第一数据组中的第二个位置,从而获得第二数据集,即第二数据集具体为:1、3、4、5、7、9和10。

[0135] S602、加速装置划分第二数据集,获得N个数据子集。其中,N为大于或等于2的整数。其中,N个数据子集中的每个数据子集包括第二数据集中连续排列的多个数据。

[0136] 划分可以理解为分组,或者可以理解为将第二数据集依次划分为多个数据子集。其中,N个数据子集中的任意两个数据子集不重叠,换言之,N个数据子集中的任意两个子集不包括同一个数据。可选的,N个数据子集中的任意两个数据子集包括的数据的个数是相同的。

[0137] 例如,第二数据集具体为:1、3、4、5、7、9、10、13、14、20、24、25、27、29、30、33、34和37。加速装置可以将第二数据集划分为两个数据子集,其中一个数据子集具体为:1、3、4、5、7、9、10、13和14,另一个数据子集具体为:20、24、25、27、29、30、33、34和37。

[0138] 示例性的,加速装置可以按照预设个数,划分第二数据集,以获得N个数据子集。预设个数为每个数据子集包括的数据的个数。其中,预设个数可以是被预存在加速装置中的,或者可以是加速装置根据第二数据集包括的数据的总个数确定的。

[0139] S603、加速装置根据N个数据子集,获得N个子树。

[0140] 示例性的,加速装置可以并行对N个数据子集进行处理,获得N个子树。其中,子树可以称为子树索引,可以理解为数据子集的数据索引结构。其中每个子树对应一个数据子集,一个数据子集对应的子树用于查找所述一个数据子集中的数据。

[0141] 其中,加速装置确定N个子树中的每个子树的方式相同,下面以加速装置确定N个子树中的一个子树为例进行介绍。

[0142] 加速装置可以被预配置有树的层级,或者用户在加速装置中设置有树的层级。加速装置确定了树的层级,也就相当于获得了子树的层级,子树的层级例如为树的层级减一。加速装置可以根据一个数据子集以及子树的层级,确定一个子树。

[0143] 示例性的,加速装置将该数据子集包括的每个数据作为一个叶子节点,可以获得多个叶子节点。加速装置确定多个叶子节点的上层节点的值,以此类推,获得满足所述子树的层级的子树。

[0144] 可选的,加速装置可按照第二规则,确定多个叶子节点的上层节点的值,第二规则例如为:沿上层节点的第一方向(如左边)的叶子节点的值小于上层节点的值,沿上层节点的第二方向(如右边)的叶子节点的值大于上层节点的值。

[0145] 例如,请参照图9,为本申请实施例提供的一种创建树的过程示意图。在图9中以第二数据集对应的第一数据子集包括1、2、3、8、9和10,第二数据子集包括13、15、16、19、23和26,以及以子树的层级为2级为例进行介绍。

[0146] 如图9所示,加速装置将第一数据子集中的每个数据作为一个叶子节点,并确定第一数据子集对应的叶子节点的上层节点的值,从而获得子树1。同理,加速装置将第二数据子集中的每个也作为一个叶子节点,并确定第一数据子集对应的叶子节点的上层节点的值,从而获得子树2。

[0147] S604、加速装置对N个子树进行合并操作,获得第二数据集的树。

[0148] 其中,树也可以称为树状索引,可以理解为第二数据集的索引结构。树用于查找第二数据集中的数据。

[0149] 加速装置可以按照N个数据子集的第一顺序,将这N个数据子集对应的N个子树进行合并操作,从而获得第二数据集的树。其中,N个子树中的任一子树可以理解为树(也可成为树状索引)的一部分。合并操作可以理解为按照所述第一顺序,依次连接这N个子树,并确定N个子树的根节点的值,从而获得树。其中,树包括根节点和N个子树。可选的,加速装置也可以按照第二规则,确定根节点的值。第二规则的内容可以参照前文。

[0150] 继续参照图9,加速装置可以将子树1和子树2进行合并,并确定这两个子树的根节点的值,从而获得如图9所示的树。

[0151] 在一种可能的实施方式中,加速装置可包括多个索引模块和计算模块。每个索引模块和计算模块可以通过硬件或软件实现。其中,每个索引模块可以包括控制子模块和运算子模块,计算模块也可以包括控制子模块和运算子模块。这多个索引模块中的每个索引模块可以用于创建N个子树中的一个子树。计算模块可以对多个索引模块输出的N个子树进行合并操作,从而获得树。在该实施方式中,加速装置可以通过多个索引模块并行创建多个子树,从而提高加速装置创建树的效率。

[0152] 其中,加速装置对第二数据集执行上述S602-S604的过程可以视为加速装置对第二数据集执行了创建树操作。

[0153] 请参照图10,为本申请实施例提供的一种创建树的原理示意图。如图10所示,加速装置包括多个索引模块,多个索引模块中的每个索引模块可以根据N个数据子集中的一个数据子集,获得一个子树,这多个索引模块可以对应输出N个子树(包括图10中的子树1、子树2和子树N)。计算模块对这N个子树进行合并操作,获得第二数据集的树。

[0154] 在本申请实施例中,加速装置确定第二数据集的树的过程中,加速装置或处理器无需生成指令以及译码指令等,有利于简化数据处理的过程,提高加速装置的处理数据的效率。并且,加速装置可以并行创建多个子树,也有利于提高创建第二数据集的树的效率,也就提高数据处理的效率。另外,本申请实施例还提供一种对第一数据集进行排序的方法,在对第一数据集进行排序的过程中,加速装置可以并行对第一数据组中的多个数据和第二数据组中的多个目标数据进行比较,有利于快速确定目标数据在第一数据组中的位置,提高对第一数据集进行排序的效率,也就能提高处理数据的效率。

[0155] 在一种可能的实施方式中,加速装置还可以对第二数据集执行分组操作。下面介绍加速装置执行分组操作的方式。

[0156] 加速装置可以将第二数据集中的多个数据与目标分组的分组键进行比较,确定与目标键匹配的至少一个数据。加速装置将这至少一个数据确定为目标分组。分组键可以是加速装置确定的,或者加速装置根据第一请求确定的,例如,第一请求包括分组键,加速装置接收第一请求,也就相当于获得了分组键。

[0157] 示例性的,加速装置根据目标分组的分组键,获得第三数据组。其中,第三数据组与第二数据集包括的数据的个数相同,且第三数据组中的每个元素均为目标键。加速装置可以将第二数据集中的每个位置(为了便于区分,称为第二位置)上的数据,与第三数据组中的与每个第二位置上对应的位置上的数据进行比较,从而确定与目标键匹配的至少

一个数据,将这至少一个数据划分为目标分组。换言之,目标分组包括的数据为所述至少一个数据。

[0158] 其中,与目标键匹配的数据可以为数据的索引与目标键相同的数据,可以是数据的索引小于或等于目标键的数据,可以是数据与目标键相同,或者数据小于或等于目标键。

[0159] 可选的,加速装置可包括第三寄存器和第四寄存器,加速装置可将第二数据集写入第三寄存器,将第三数据组写入第四寄存器,并通过第三寄存器和第四寄存器,对第二数据集和第三数据组进行比较。其中,寄存器的实现方式可参照前文论述的内容。

[0160] 例如,请参照图11,为本申请实施例提供的一种对数据进行分组的过程示意图。如图11所示,第二数据集包括1、1、1、2、2、2和2、以及目标键为1。加速装置根据目标键,确定第三数据组具体为:1、1、1、1、1、1和1。

[0161] 加速装置可以将第一数据集中的每个第二位置上的数据与第三数据组中与每个第二位置上的数据进行比较,从而确定与目标键匹配的至少一个数据。

[0162] 具体的,加速装置将第一个位置上的数据(即1)与第三数据组中的第一个位置上的数据(即1)进行比较,将第一数据集中的第二个位置上的数据(即1)与第三数据组中的第二个位置上的数据(即1)进行比较,将第一数据集中的第三个位置上的数据(即1)与第三数据组中的第三个位置上的数据(即1)进行比较,将第一数据集中的第四个位置上的数据(即2)与第三数据组中的第四个位置上的数据(即1)进行比较,将第一数据集中的第五个位置上的数据(即2)与第三数据组中的第五个位置上的数据(即1)进行比较,将第一数据集中的第六个位置上的数据(即2)与第三数据组中的第六个位置上的数据(即1)进行比较,以及将第一数据集中的第七个位置上的数据(即2)与第三数据组中的第七个位置上的数据(即1)进行比较,从而确定第一数据集中的第一个位置、第二个位置和第三个位置上的数据与目标键匹配。从而加速装置确定将第一数据集中的第一个位置、第二个位置和第三个位置上的数据确定为目标分组。

[0163] 加速装置获得目标分组之后,可以将目标分组的信息和目标分组写入外存。其中,外存的内容可以参照前文。目标分组的信息包括目标键。

[0164] 可选的,目标分组的信息还包括目标分组包括的数据(即至少一个数据)的个数(count)、目标分组包括的数据(即至少一个数据)中的最大值(max)、目标分组包括的数据(即至少一个数据)中的最小值(min)、或目标分组包括的数据(即至少一个数据)的求和结果(sum)中的一种或多种。

[0165] 可选的,加速装置可以根据多个目标键,确定多个目标分组,其中,加速装置确定每个目标分组的方式可参照前文论述的内容。

[0166] 在加速装置确定多个目标分组的情况下,加速装置可以以页(page)的形式存储这多个目标分组的信息。其中,一个目标分组的信息以及一个目标分组可以对应存储在一个页或多个页中。

[0167] 可选的,一个页除了存储目标分组的信息和目标分组之外,所述一个页还包括所述一个页的信息。所述一个页的信息包括所述一个页的标识。

[0168] 其中,所述一个页的信息还包括所述一个页的页头(page head)、目标分组所属的表的标识、用于存放与所述一个页对应的目标分组的页的个数、所述一个页的索引、目标分组的总个数、所述一个页包括的目标分组的数量、以及所述一个页包括的分组键的个数中

的一种或多种。

[0169] 例如,请参照图12,为本申请实施例提供的一种页的结构示意图。如图12所示,该页包括该页的信息和该页包含的目标分组的信息。如图12所示,页的信息包括页的标识、页的索引和下一页的索引。目标分组的信息包括分组的标识(如分组1或分组2)、索引1至索引n、分组包括的数据的个数、分组中的最大值、最小值、以及求和结果等。

[0170] 在一种可能的实施方式中,加速装置可以对数据进行更新。例如,加速装置可以对第二数据集进行更新、插入或删除操作等,这种情况下,加速装置可以对第二数据集进行更新,还可对第二数据集对应的目标分组的信息进行更新。例如,加速装置可以根据更新的数据对应的分组键,对该分组键对应的分组的信息进行更新,并对该分组所在的页的信息进行更新。

[0171] 请参照图13,为本申请实施例提供的一种数据处理方法的流程示意图。图13所示的实施例涉及的数据处理方法例如应用于图1所示的应用场景。图13所示的实施例涉及的计算设备例如为图1的计算设备120、图2的计算设备221、图3的主节点310、图3的从节点320、图4的第一节点410、图4的第二节点420、或者图4的第三节点430,图13所示的实施例涉及的计算设备的结构示意图例如为图5所示的计算设备的结构示意图,图13所示的实施例涉及的处理器例如为图5所示的处理器510,以及图13所示的实施例涉及的加速装置例如为图5所示的加速装置520。图13所示的数据处理方法包括如下处理步骤:

[0172] S1301、处理器向加速装置发送第一请求。相应的,加速装置接收来自处理器的第一请求。

[0173] 第一请求的含义可以参照前文。在本申请实施例中,以第一请求包括第一地址,且地址具体指示为外存中某个地址。可选的,第一请求还包括结构化查询语言(structured query language,SQL)语句。

[0174] S1302、加速装置从外存中获取第一数据集。

[0175] 加速装置可以根据第一地址,获取第一数据集。其中,第一数据集的内容可以参照前文。

[0176] S1301-S1302为加速装置获取第一数据集的方式,如果加速装置采用其他方式获取第一数据集时,加速装置可以无需执行S1301-S1302的步骤,即S1301-S1302为可选的步骤。

[0177] S1303、加速装置确定目标执行计划。

[0178] 加速装置可以确定用于处理第一数据集的目标执行计划。其中,目标执行计划包括加速装置对第一数据集需执行的操作,例如,创建树操作、转换操作、排序操作或分组操作中的一种或多种。

[0179] 可选的,加速装置可从多个优化器中确定与第一数据集匹配的目标优化器,并从目标优化器接收目标执行计划。其中,优化器又可以称为查询优化器,用于确定执行计划。目标优化器可根据第一请求中的结构化查询语言(structured query language,SQL)语句,确定与第一数据集对应的目标执行计划。在该方式中,无需处理器确定目标执行计划,有利于减少处理器的处理量。

[0180] 在本申请实施例中,是以目标执行计划包括对第一数据集进行排序操作和创建树操作为例进行介绍。

- [0181] S1304、加速装置对第一数据集进行排序,获得第二数据集。
- [0182] 其中,加速装置对第一数据集进行排序的内容可以参照前文论述的内容,此处不再赘述。
- [0183] S1305、加速装置划分第二数据集,获得N个数据子集。
- [0184] 加速装置划分第二数据集的方式、以及数据子集的内容均可以参照前文论述的内容,此处不再赘述。
- [0185] S1306、加速装置根据N个数据子集,确定N个子树。
- [0186] 加速装置确定N个子树的方式可以参照前文论述的内容,此处不再赘述。
- [0187] S1307、加速装置对N个子树进行合并操作,获得第二数据集的树。
- [0188] 其中,加速装置对N个子树进行合并操作的具体内容可以参照前文论述的内容,此处不再赘述。
- [0189] S1308、加速装置对第二数据集进行分组,获得目标分组。
- [0190] 其中,目标分组、分组的具体过程可以参照前文论述的内容,此处不再赘述。
- [0191] S1309、加速装置将目标分组的信息和目标分组写入外存。外存例如为图5所示的外存540。
- [0192] 其中,目标分组的信息的内容可以参照前文论述的内容。
- [0193] 请参照图14,为本申请实施例提供的一种数据处理方法的流程示意图。图14所示的实施例涉及的数据处理方法例如应用于图2至图5任一所示的应用场景。图14所示的实施例涉及的第一节点和第二节点例如为图2中的两个计算设备221,或者,图14所示的实施例涉及的第一节点例如为图3中的主节点310,第二节点例如为图3中的从节点320,或者,图14所示的实施例涉及的第一节点例如为图4中的第一节点410,第二节点例如为图4中的第二节点420或第三节点430。图14所示的实施例以第一节点包括第一处理器和第一加速装置,以及第二节点包括第二处理器、第二加速装置和外存为例。图14所示的实施例涉及的第一加速装置或第二加速装置例如为图5所示的加速装置520。图14所示的数据处理方法包括如下处理步骤:
- [0194] S1401、第一处理器接收第二请求。第二请求用于请求对第一数据集进行处理。
- [0195] 第一处理器可以从客户端接收第二请求,客户端例如为图1中的客户端111,或者为图2中的客户端211。
- [0196] S1402、第一处理器向第一加速装置发送第二请求。相应的,第一加速装置接收来自第一处理器的第二请求。
- [0197] S1403、第一加速装置确定目标执行计划。
- [0198] 其中,第一加速装置确定目标执行计划的内容可以参照前文论述的内容。
- [0199] S1404、第一加速装置向第二处理器发送目标执行计划。相应的,第二处理器接收来自第一加速装置的目标执行计划。
- [0200] S1405、第二处理器向第二加速装置发送第一请求。相应的,第二加速装置接收来自第二处理器的第一请求。
- [0201] 可选的,第一请求包括目标执行计划。
- [0202] S1401-S1405为第二加速装置获取第一数据集的一种方式,当第二加速装置采用其他方式确定第一数据集时,则无需执行S1401-S1405,即S1401-S1405为可选的步骤。

- [0203] S1406、第二加速装置获取第一数据集。
- [0204] 示例性的,第二加速装置根据第一请求,获得第一数据集。
- [0205] S1407、第二加速装置对第一数据集进行排序,获得第二数据集。
- [0206] 对第一数据集进行排序的内容可以参照前文论述的内容,此处不再赘述。
- [0207] 在一种可能的实施方式中,第二加速装置对第一数据集执行转换操作,获得转换后的第一数据集。其中,转换操作的内容可以参照前文,此处不再赘述。
- [0208] S1408、第二加速装置划分第二数据集,获得N个数据子集。
- [0209] 其中,第二加速装置获得N个数据子集的方式可以参照前文论述的内容,此处不再赘述。
- [0210] S1409、第二加速装置根据N个数据子集,确定N个子树。
- [0211] 第二加速装置确定N个子树的内容可以参照前文论述的内容。
- [0212] S1410、第二加速装置对N个子树进行合并操作,获得第二数据集的树。
- [0213] 其中,树的含义、以及第二加速装置对N个子树进行合并操作的内容可以参照前文。
- [0214] S1411、第二加速装置将第二数据集的树和第二数据集写入外存。
- [0215] S1412、第二加速装置对第二数据集进行分组,获得目标分组,以及目标分组的信息。
- [0216] 其中,目标分组以及目标分组的信息可以参照前文论述的内容。
- [0217] S1413、第二加速装置将目标分组和目标分组的信息写入外存。
- [0218] 作为一个示例,S1411-S1413为可选的步骤,在图14中以虚线示意。
- [0219] 在本申请实施例中,可以由第一加速装置确定目标执行计划,而无需由第一处理器确定目标执行计划,有利于减少第一处理器的处理量。并且,第二加速装置可以对第一数据集执行排序操作、转换操作、分组操作和创建树操作中的一种或多种,且执行过程中无需第一处理器译码和下发指令,有利于简化处理数据的过程,提高处理数据的效率。
- [0220] 本申请实施例还提供一种数据处理方法。请参照图15,为该方法的一种流程示意图。图15所示的方法可以由加速装置执行,加速装置例如可为图5中的加速装置520。图15所示的实施例涉及的加速装置例如可以设置在图1的计算设备120、图2的计算设备221、图3的主节点310、图3的从节点320、图4的第一节点410、图4的第二节点420、或者图4的第三节点430中。图15所示的数据处理方法包括如下处理步骤:
- [0221] S1501、加速装置对第一数据集进行排序,获得第二数据集。
- [0222] 加速装置确定第二数据集的方式可以参照前文论述的内容。
- [0223] S1502、加速装置确定第三数据组。
- [0224] 第三数据组的内容,以及加速装置确定第三数据组的内容可以参照前文论述的内容。
- [0225] S1503、加速装置将第二数据集中的每个第二位置上的数据,与第三数据组中的与每个第二位置对应的位置上的数据进行比较,确定第二数据集中与目标键匹配的至少一个数据。
- [0226] 加速装置确定目标键匹配的至少一个数据的方式可以参照前文论述的内容。
- [0227] S1504、将至少一个数据确定为目标分组,并将目标分组的信息和目标分组写入外

存中。

[0228] 其中,目标分组的信息的内容可以参照前文论述的内容。

[0229] 在本申请实施例中,加速装置可以并行对第三数据组中的多个数据,以及第二数据集中的多个数据进行比较,提高了加速装置对第二数据集进行分组的效率。并且,无需加速装置或处理器译码或下发指令等,也有利于提高加速装置执行分组操作的效率。

[0230] 在一种可能的实施方式中,加速装置可以对第一数据集进行创建树操作和/或转换操作,执行创建树操作和转换操作的内容可以参照前文论述的内容,此处不再赘述。

[0231] 本申请实施例还提供一种数据处理方法。请参照图16,为该方法的一种流程示意图。图16所示的方法可以由加速装置执行,加速装置例如可为图5中的加速装置520。图16所示的实施例涉及的加速装置例如可以设置在图1的计算设备120、图2的计算设备221、图3的主节点310、图3的从节点320、图4的第一节点410、图4的第二节点420、或者图4的第三节点430中。图16所示的数据处理方法包括如下处理步骤:

[0232] S1601、加速装置获取第一数据集。

[0233] 加速装置获得第一数据集的内容可以参照前文论述的内容。

[0234] S1602、加速装置确定第一数据组。

[0235] 第一数据组的内容、确定第一数据组的方式可参照前文论述的内容。

[0236] 在一种可能的实施方式中,加速装置可以对第一数据集执行转换操作。执行转换操作的内容可以参照前文论述的内容。

[0237] S1603、加速装置将第一数据组中的每个第一位置上的数据,与第二数据组中每个第一位置对应的位置上的目标数据进行比较,确定目标位置。

[0238] 加速装置确定目标位置的方式也可参照前文论述的内容。

[0239] S1604、加速装置将目标数据插入到目标位置中,获得第二数据集。

[0240] 加速装置获得第二数据集的内容也可参照前文论述的内容。

[0241] 在一种可能的实施方式中,加速装置可以对第一数据集进行创建树操作和/或分组操作,执行创建树操作和分组操作的内容可以参照前文论述的内容,此处不再赘述。

[0242] 请参照图17,为本申请的实施例提供的加速装置的结构示意图。如图17所示,加速装置1700包括数据获取模块1701、分组模块1703、排序模块1704和树创建模块1706。可选的,加速装置1700还包括转换模块1702、执行计划确定模块1705、驱动模块1708和接口模块1709。其中,加速装置1700中的任意两个模块之间可以通信总线1707连接。

[0243] 可选的,图17中所示的多个模块可以是通过软件实现,也可以是通过硬件实现,本申请实施例对此不做限定。

[0244] 在本申请实施例中,加速装置1700可以用于实现前文任一的数据处理方法,例如图6、图13至图16中任一的数据处理方法。

[0245] 在一种可能的实施例中,加速装置1700可用于实现前文图6、图13和图14所示的数据处理方法。

[0246] 例如,加速装置1700可用于实现前文图6所示的数据处理方法。具体的,排序模块1704用于执行S601,树创建模块1706用于执行S602-S604的步骤。

[0247] 又例如,加速装置1700中的可用于实现前文图13中任一的数据处理方法。具体的,排序模块1704用于执行S1304的步骤,树创建模块1706用于执行S1304-S1307的步骤。可选

的,数据获取模块1701用于执行S1301-S1302的步骤,执行计划确定模块1705用于执行S1303的步骤,分组模块1703用于执行S1308的步骤。

[0248] 又例如,加速装置1700中的可用于实现前文图14中任一的数据处理方法。具体的,数据获取模块1701用于执行S1406的步骤,排序模块1704用于执行S1407的步骤,树创建模块1706用于执行S1408至S1410的步骤。

[0249] 在一种可能的实施例中,加速装置1700可用于实现前文图15中任一的数据处理方法。具体的,排序模块1704用于执行S1501的步骤,分组模块1703用于执行S1502-S1504的步骤。

[0250] 在一种可能的实施例中,加速装置1700可用于实现前文图16中任一的数据处理方法。具体的,数据获取模块1701用于执行S1601的步骤,排序模块1704用于执行S1602-S1604的步骤。

[0251] 在一种可能的实施方式中,加速装置1700可以包括多个转换模块1702,每个转换模块1702用于将一种格式的数据转换为另一种格式的数据,和/或用于将一种数据类型的数据转换为另一种数据类型的数据。加速装置1700可以根据第一数据集的信息,从多个转换模块1702中确定用于处理第一数据集的转换模块1702。在该实施方式中,加速装置1700可以利用多个转换模块1702,并行对多个数据进行处理,有利于提高加速装置1700进行格式转换的效率。

[0252] 在一种可能的实施方式中,加速装置1700可包括多个树创建模块1706,每个树创建模块1706例如可用于实现前文中论述的索引模块的功能。索引模块的内容可以参照前文图10的内容。其中,树创建模块1706可以包括控制子模块和运算符模块。这多个多个树创建模块1706中的每个树创建模块1706可以用于创建一个子树。在该实施方式中,加速装置1700可以通过多个树创建模块1706并行创建多个子树,从而提高加速装置1700创建树的效率,也就能提高处理数据的效率。

[0253] 可选的,加速装置1700可以通过驱动模块1708调用加速装置1700的硬件资源,接口模块1709可以用于与外部设备(如处理器)进行通信。

[0254] 请参照图18,为本申请实施例提供的一种加速装置的结构示意图。如图18所示,加速装置1800包括处理器1801和供电电路1802。供电电路1802用于为处理器1801供电。其中,处理器1801可实现前文任一的数据处理方法,例如实现前文图6、图13至图16中任一的数据处理方法。可选的,加速装置1800可用于实现图17中的加速装置1700的功能。

[0255] 可选的,加速装置1800还包括寄存器,寄存器例如包括图6所示的实施例涉及的第一寄存器、第二寄存器、第三寄存器和第四寄存器。

[0256] 本申请实施例提供一种计算设备,所述计算设备包括加速装置和处理器。处理器可以用于向加速装置发送第一请求,第一请求的含义可以参照前文。加速装置根据第一请求,获得第一数据集,并对第一数据集执行前文论述的任一数据处理方法,例如实现前文图6、图13至图16中任一数据处理方法。其中,处理器和加速装置的具体实现形式可以参照前文论述的内容。可选的,加速装置的结构例如可为图17所示的加速装置1700或图18所示的加速装置1800。

[0257] 本申请实施例提供一种芯片系统,该芯片系统包括:处理器和接口。其中,该处理器用于从该接口调用并运行指令,当该处理器执行该指令时,实现前文任一数据处理方法,

例如实现前文图6、图13至图16中任一数据处理方法。

[0258] 本申请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用于存储计算机程序或指令,当其被运行时,实现前文任一数据处理方法,例如实现前文图6、图13至图16中任一数据处理方法。

[0259] 本申请实施例提供一种包含指令的计算机程序产品,当其在计算机上运行时,实现前文任一数据处理方法,例如实现前文图6、图13至图16中任一数据处理方法。

[0260] 本申请的实施例中的方法步骤可以通过硬件的方式来实现,也可以由处理器执行软件指令的方式来实现。软件指令可以由相应的软件模块组成,软件模块可以被存放于随机存取存储器、闪存、只读存储器、可编程只读存储器、可擦除可编程只读存储器、电可擦除可编程只读存储器、寄存器、硬盘、移动硬盘、CD-ROM或者本领域熟知的任何其它形式的存储介质中。一种示例性的存储介质耦合至处理器,从而使处理器能够从该存储介质读取信息,且可向该存储介质写入信息。当然,存储介质也可以是处理器的组成部分。处理器和存储介质可以位于ASIC中。另外,该ASIC可以位于基站或终端中。当然,处理器和存储介质也可以作为分立组件存在于基站或终端中。

[0261] 在上述实施例中,可以全部或部分地通过软件、硬件、固件或者其任意组合来实现。当使用软件实现时,可以全部或部分地以计算机程序产品的形式实现。所述计算机程序产品包括一个或多个计算机程序或指令。在计算机上加载和执行所述计算机程序或指令时,全部或部分地执行本申请实施例所述的流程或功能。所述计算机可以是通用计算机、专用计算机、计算机网络、网络设备、用户设备或者其它可编程装置。所述计算机程序或指令可以存储在计算机可读存储介质中,或者从一个计算机可读存储介质向另一个计算机可读存储介质传输,例如,所述计算机程序或指令可以从一个网站站点、计算机、服务器或数据中心通过有线或无线方式向另一个网站站点、计算机、服务器或数据中心进行传输。所述计算机可读存储介质可以是计算机能够存取的任何可用介质或者是集成一个或多个可用介质的服务器、数据中心等数据存储设备。所述可用介质可以是磁性介质,例如,软盘、硬盘、磁带;也可以是光介质,例如,数字视频光盘;还可以是半导体介质,例如,固态硬盘。该计算机可读存储介质可以是易失性或非易失性存储介质,或可包括易失性和非易失性两种类型的存储介质。

[0262] 在本申请的各个实施例中,如果没有特殊说明以及逻辑冲突,不同的实施例之间的术语和/或描述具有一致性、且可以相互引用,不同的实施例中的技术特征根据其内在的逻辑关系可以组合形成新的实施例。

[0263] 可以理解的是,在本申请的实施例中涉及的各种数字编号仅为描述方便进行的区分,并不用来限制本申请的实施例的范围。上述各过程的序号的大小并不意味着执行顺序的先后,各过程的执行顺序应以其功能和内在逻辑确定。

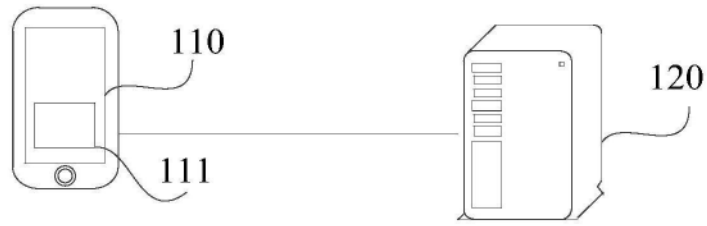


图1

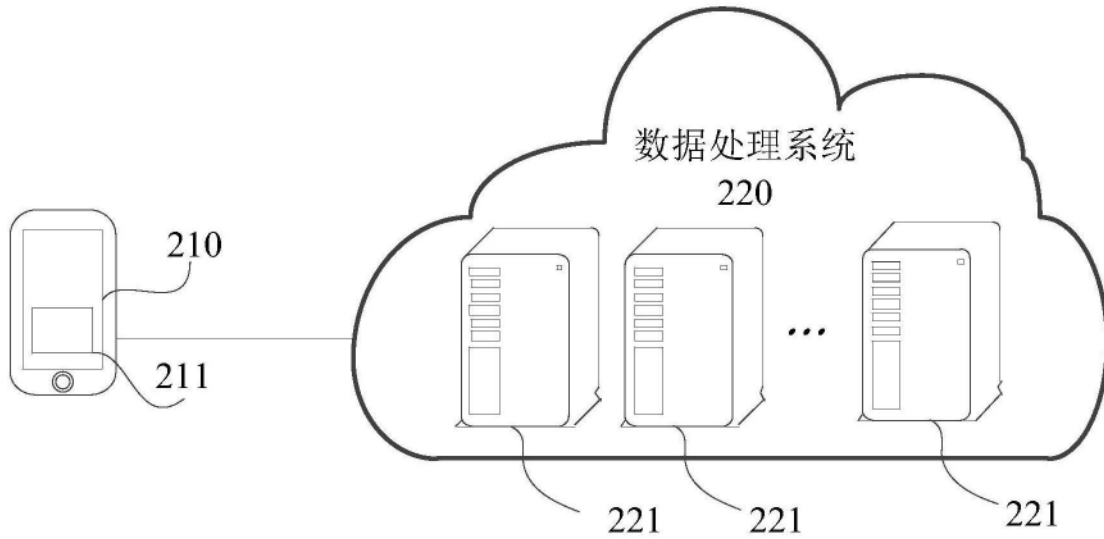


图2

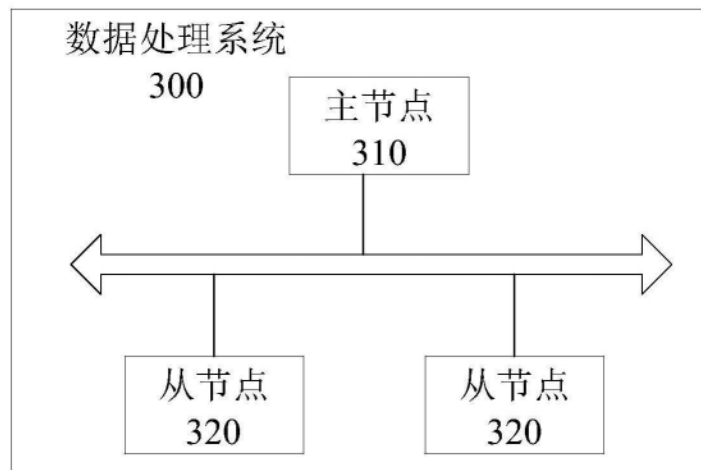


图3

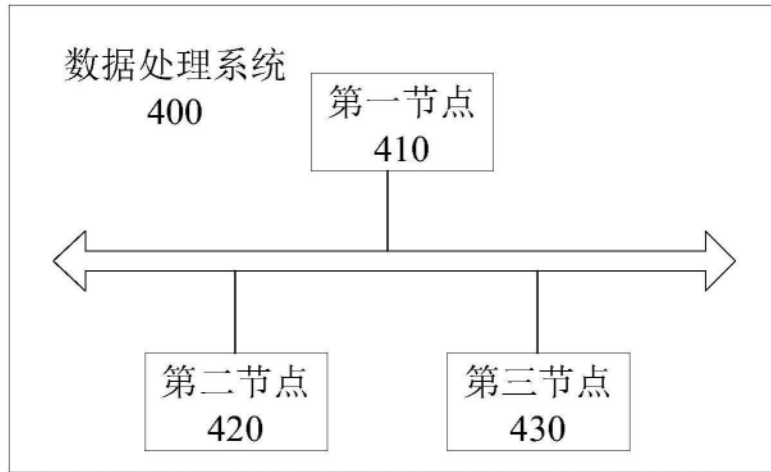


图4

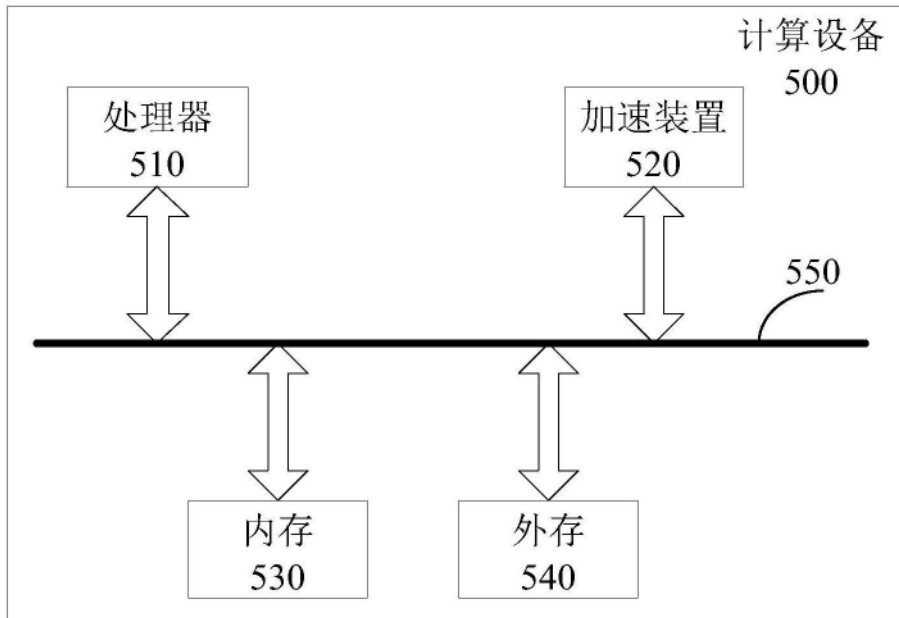


图5

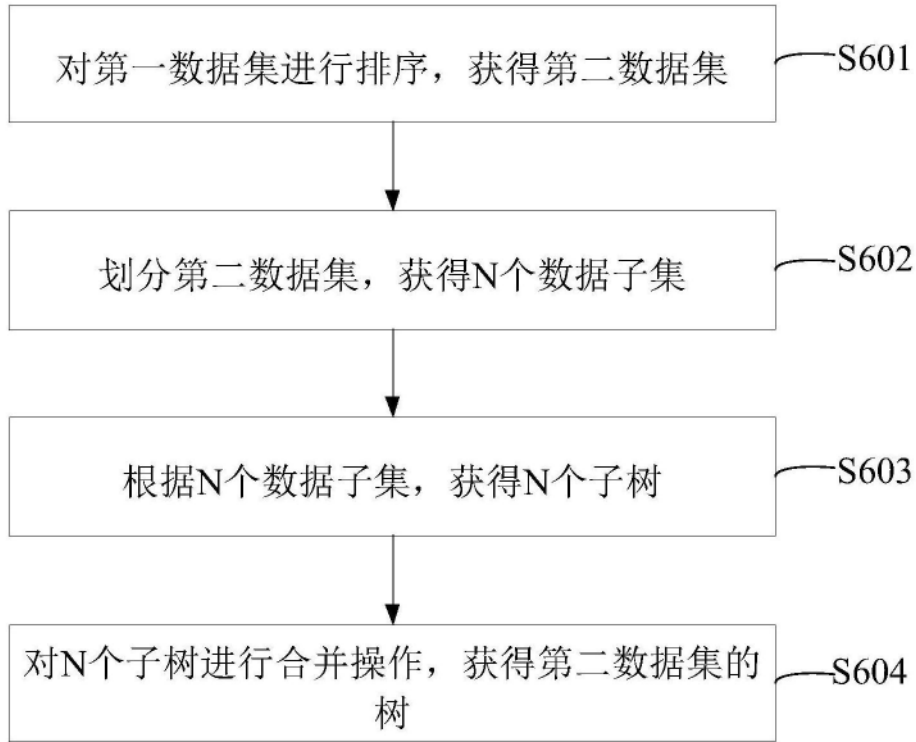


图6

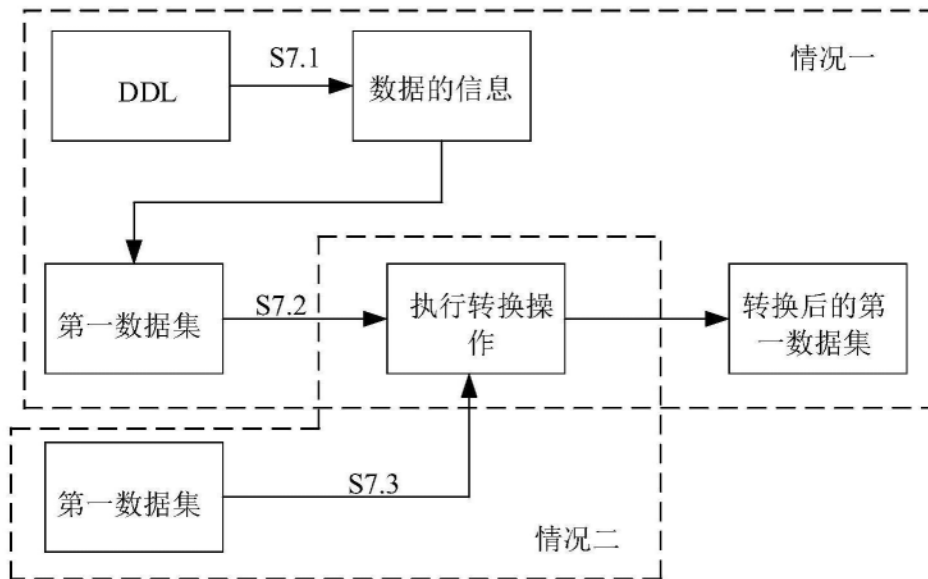


图7

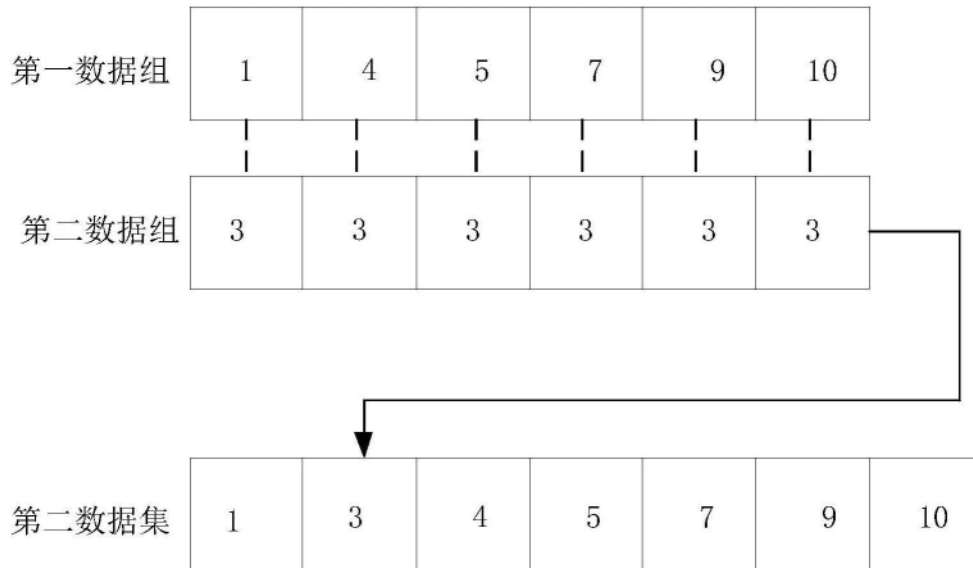


图8

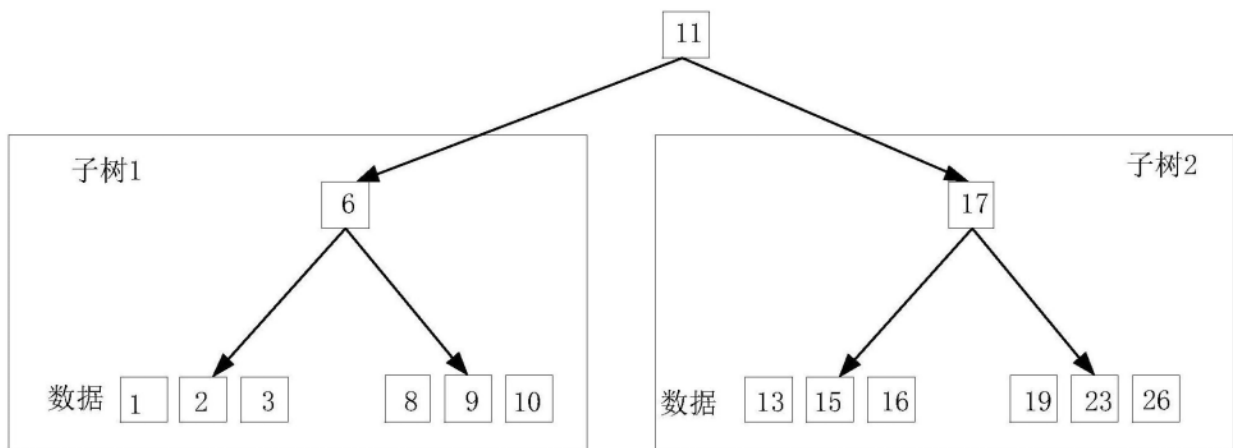


图9

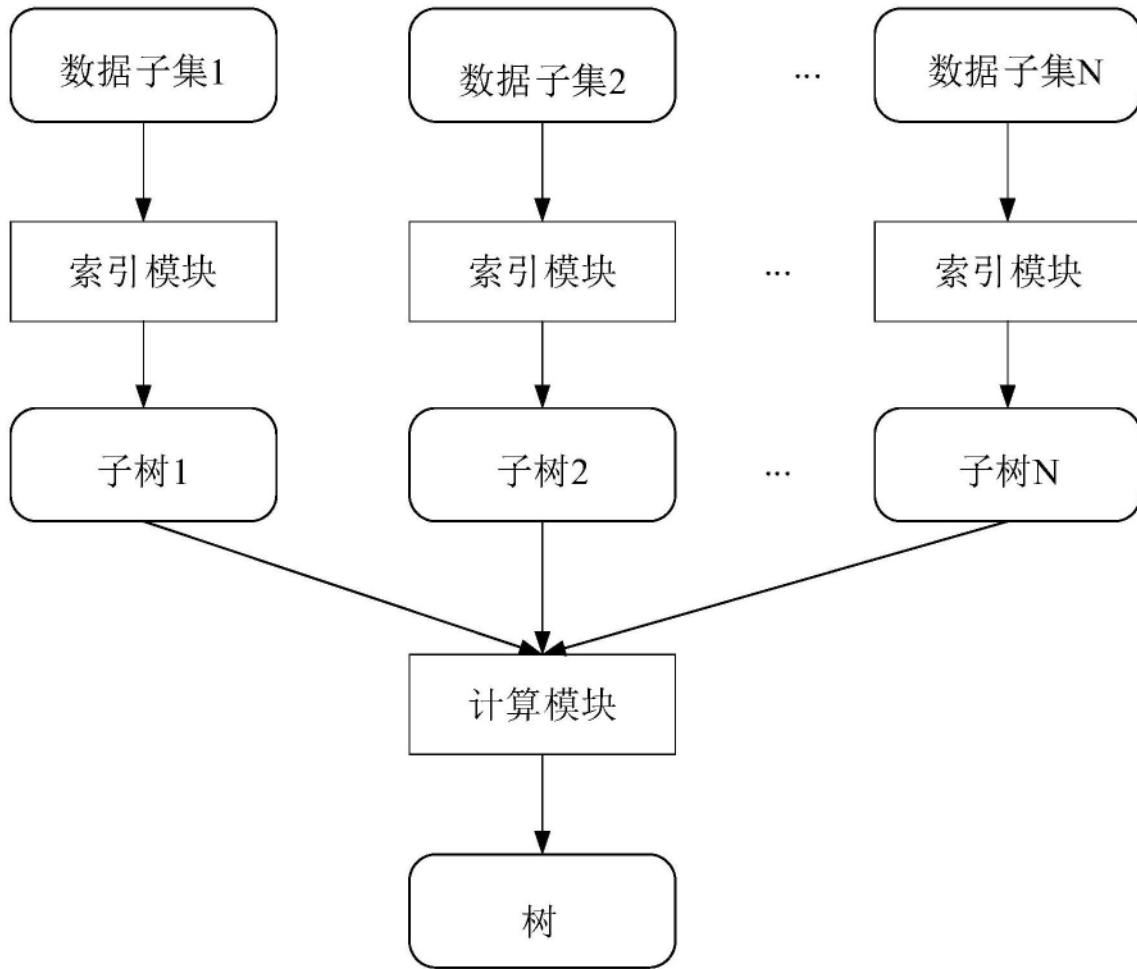


图10

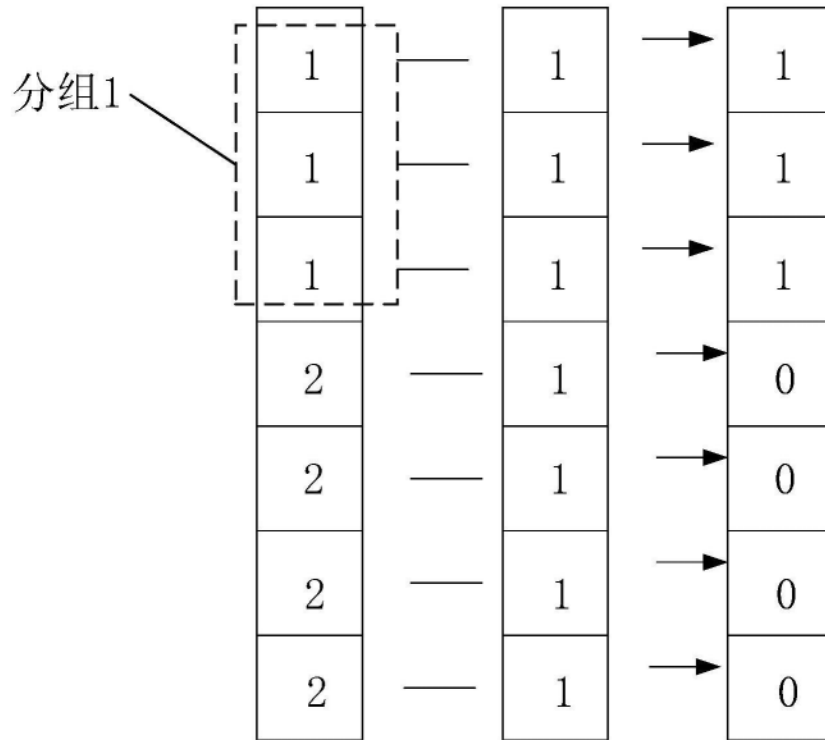


图11

页头						
表的标识		页的标识		页的索引		下一页的索引
分组1	分组2	分组键1	数据类型1	...	数据类型n	
索引1	...	索引n	个数	最小值	最大值	求和结果

图12

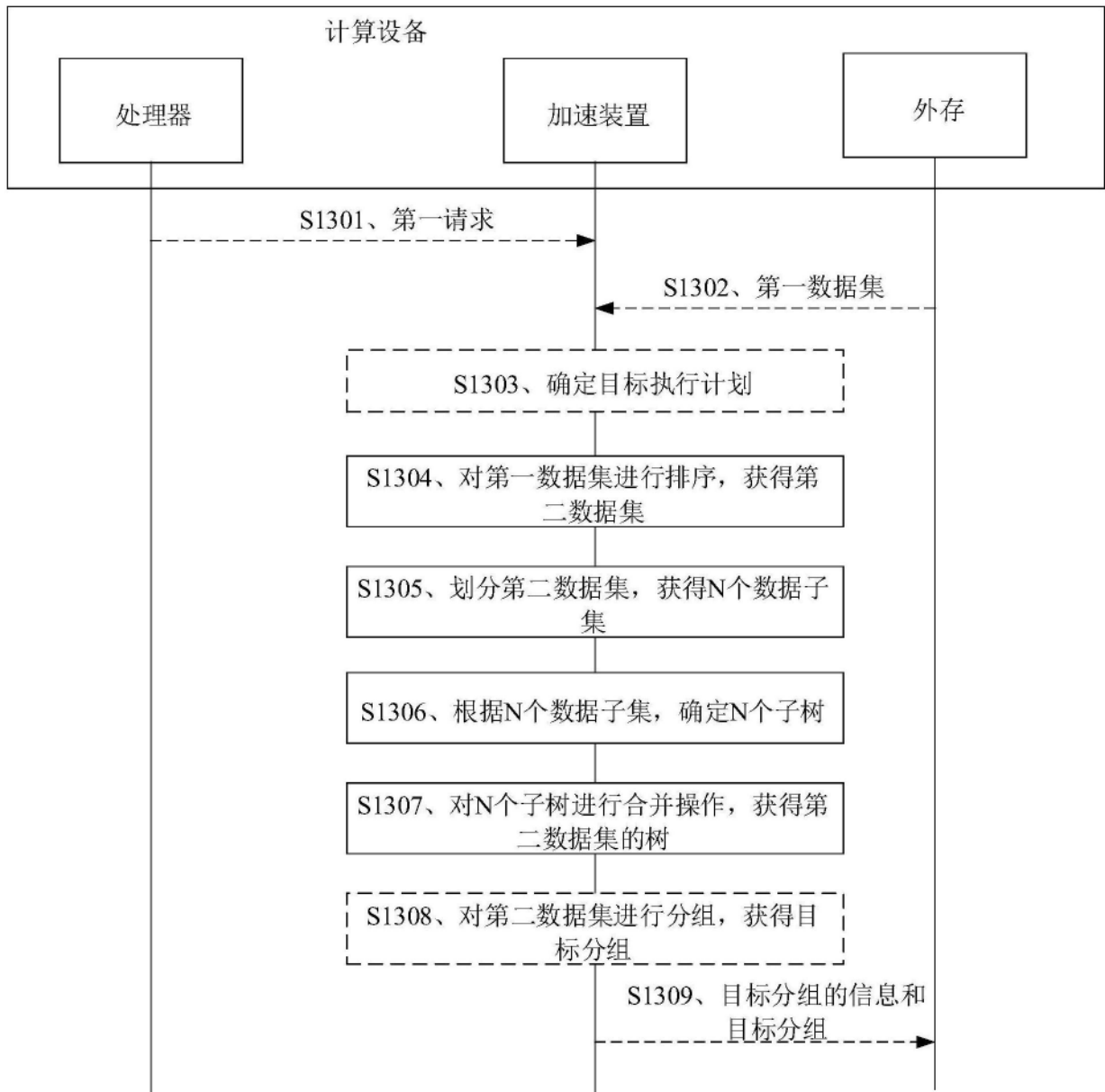


图13

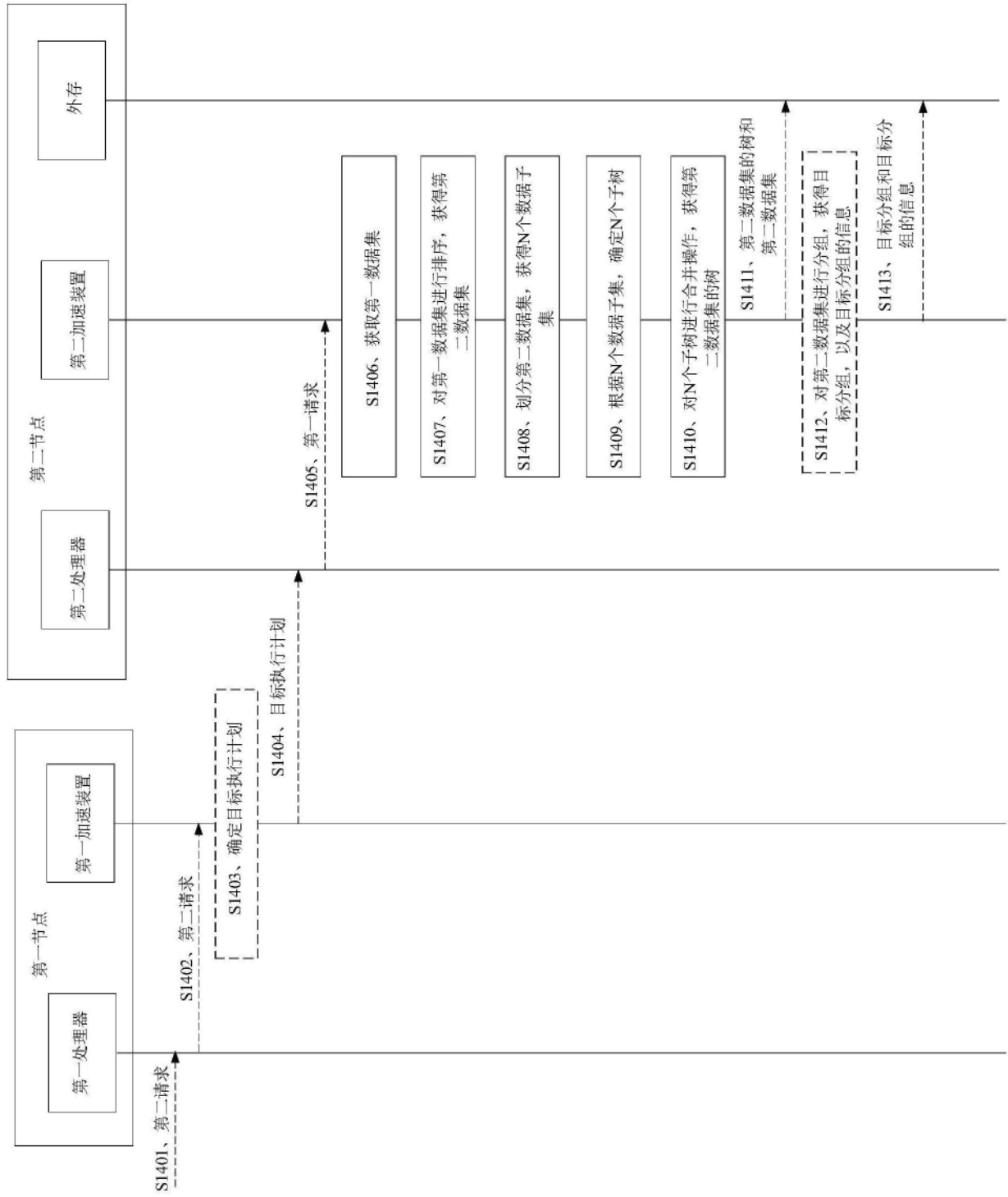


图14

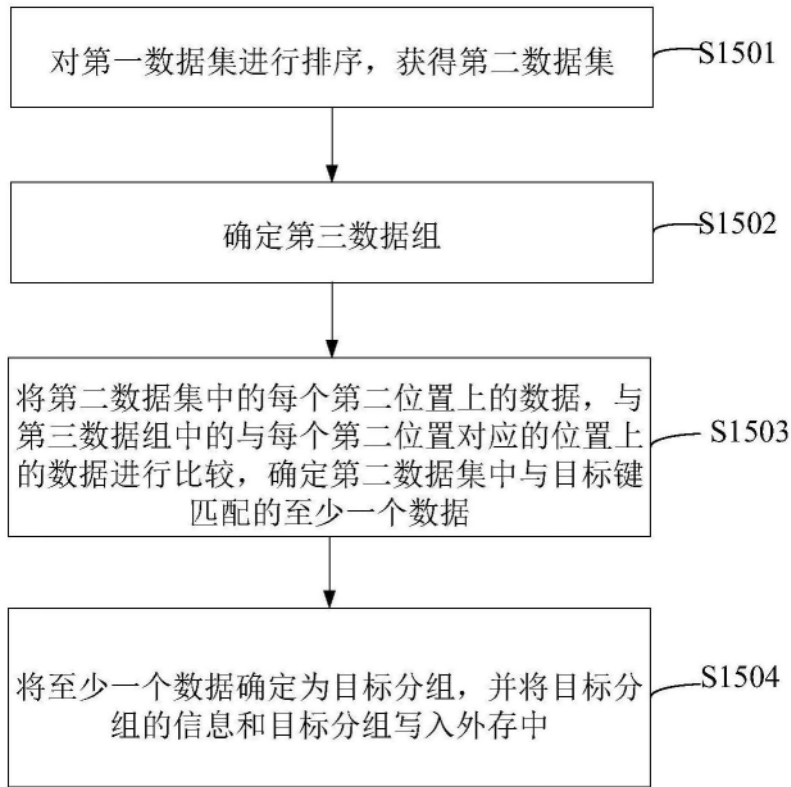


图15

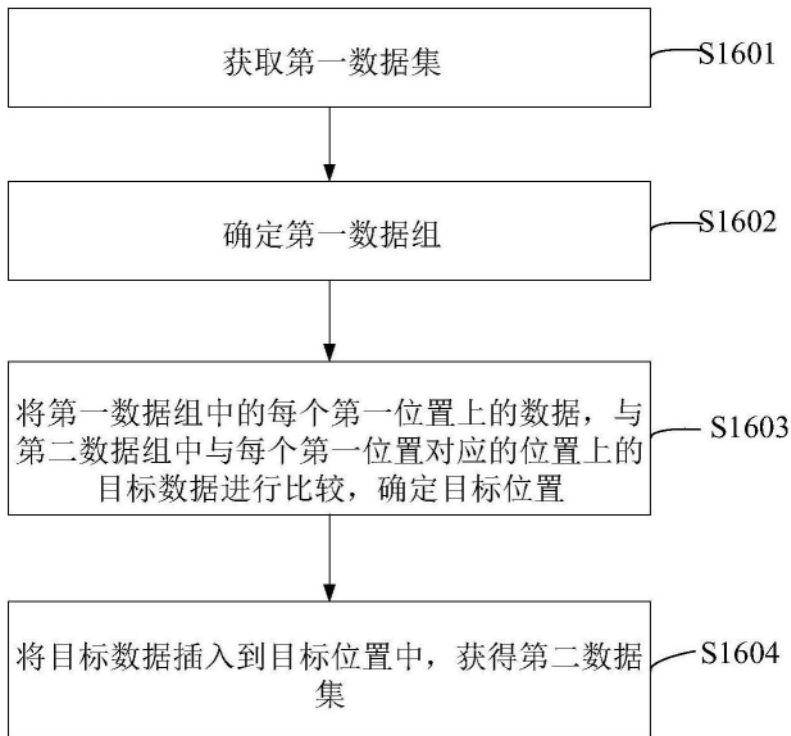


图16

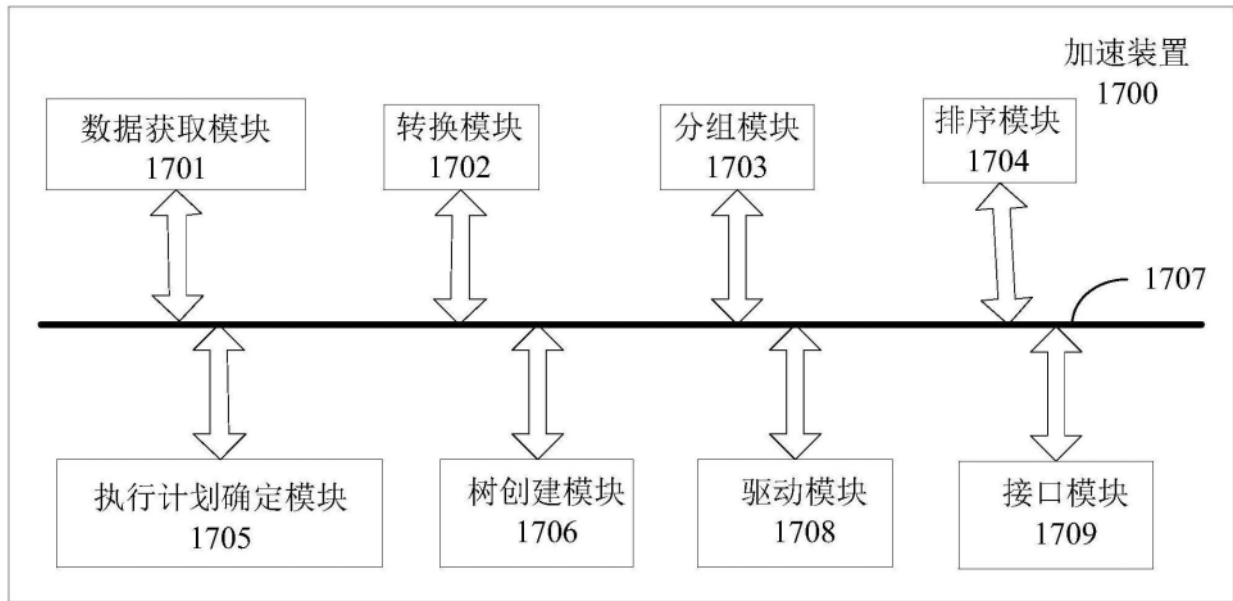


图17

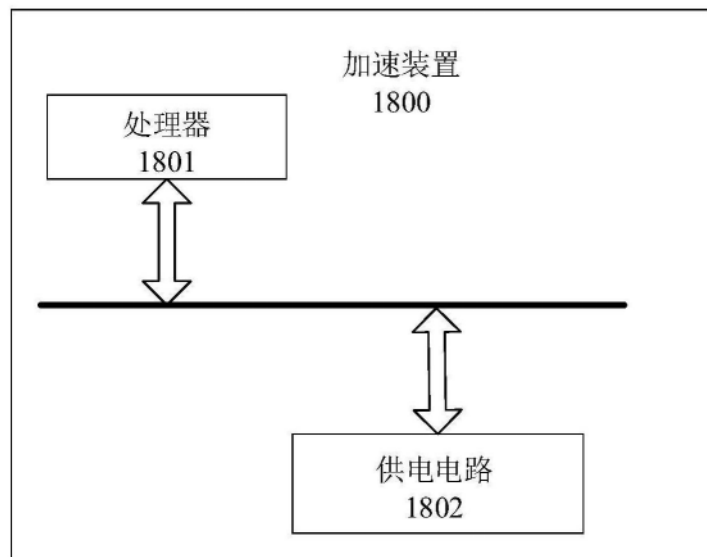


图18