



(12) 发明专利申请

(10) 申请公布号 CN 113297204 A

(43) 申请公布日 2021.08.24

(21) 申请号 202010682420.8

(22) 申请日 2020.07.15

(71) 申请人 阿里巴巴集团控股有限公司
地址 英属开曼群岛大开曼资本大厦一座四
层847号邮箱

(72) 发明人 李茂

(74) 专利代理机构 北京智信禾专利代理有限公
司 11637
代理人 吴肖肖

(51) Int. Cl.
G06F 16/22 (2019.01)

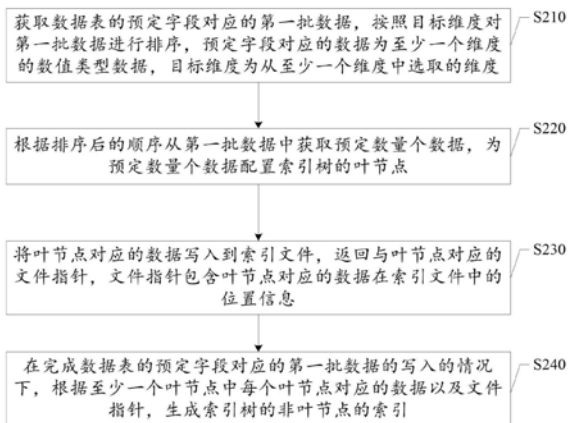
权利要求书3页 说明书15页 附图5页

(54) 发明名称

索引生成方法及装置

(57) 摘要

本说明书提供一种索引生成方法及装置,该方法包括:获取数据表的预定字段对应的第一批数据,根据目标维度对第一批数据进行排序,每个数据为至少一个维度的数值类型数据;根据排序后的顺序从第一批数据中获取预定数量个数据,为预定数量个数据配置索引树的叶节点;将叶节点对应的数据写入到索引文件,返回与叶节点对应的文件指针,该文件指针包含叶节点对应的数据在索引文件中的位置信息;在完成第一批数据的写入的情况下,根据至少一个叶节点中每个叶节点对应的数据以及文件指针,生成索引树的非叶节点的索引。根据本说明书实施例的技术方案,能够降低I/O操作的次数,进而能够提高数据库的性能。



1. 一种索引生成方法,所述方法包括:

获取数据表的预定字段对应的第一批数据,根据目标维度对所述第一批数据进行排序,所述数据为至少一个维度的数值类型数据,所述目标维度为从所述至少一个维度中选取的维度;

根据排序后的顺序从所述第一批数据中获取预定数量个数据,为所述预定数量个数据配置索引树的叶节点;

将所述叶节点对应的数据写入到索引文件,返回与所述叶节点对应的文件指针,所述文件指针包含所述叶节点对应的数据在所述索引文件中的位置信息;

在完成所述第一批数据的写入的情况下,根据至少一个叶节点中每个叶节点对应的数据以及文件指针,生成所述索引树的非叶节点的索引,所述非叶节点的索引包含所述非叶节点对应的叶节点的数据在所述目标维度上的取值范围以及对应的叶节点的所述文件指针。

2. 根据权利要求1所述的方法,其中,所述根据至少一个叶节点中每个叶节点对应的数据以及文件指针,生成所述索引树的非叶节点的索引,包括:

确定所述索引树的当前非叶节点对应的叶节点;

确定所述当前非叶节点对应的叶节点的数据在所述目标维度上的最大值与最小值;

基于所述当前非叶节点对应的叶节点的数据在所述目标维度上的最大值与最小值以及所述文件指针,生成所述当前非叶节点的索引。

3. 根据权利要求1或2所述的方法,其中,在所述根据目标维度对所述第一批数据进行排序之前,所述方法还包括:

确定所述第一批数据的各个维度的最大值与最小值;

确定各个所述维度的最大值与最小值中差值最大的维度,将所述差值最大的维度作为所述目标维度。

4. 根据权利要求1所述的方法,其中,所述取值范围为所述索引树的当前节点的索引区间,所述方法还包括:

若接收到对所述目标维度的待查询区间的查询请求,则将所述待查询区间与所述索引树的当前节点的索引区间进行比较;

若所述当前节点的索引区间在所述待查询区间内,则递归获取所述当前节点的各个子节点的数据,直到获取到所述当前节点对应的叶节点的数据;

若所述当前节点的索引区间在所述待查询区间之外,则返回;

若所述当前节点的索引区间与所述待查询区间相交,则基于所述当前节点的节点类型进行查询,所述节点类型包括叶节点和非叶节点。

5. 根据权利要求4所述的方法,其中,所述索引树为平衡二叉树,所述非叶节点的索引包括所述非叶节点的左子节点的指针与右子节点的指针,所述基于所述当前节点的节点类型进行查询,包括:

若所述当前节点为叶节点,则将所述当前节点的每个数据的值与所述待查询区间比较,获取满足所述待查询区间的数据;

若所述当前节点为非叶节点,则确定所述当前节点的左子节点和右子节点,并分别以所述左子节点和所述右子节点作为所述当前节点,基于所述当前节点的节点类型进行查

询。

6. 根据权利要求4所述的方法,其特征在于,所述将所述待查询区间与所述索引树的当前节点的索引区间进行比较,包括:

从所述索引树的根节点开始,递归获取所述索引树的下一节点;

将所述待查询区间与递归到的当前节点的索引区间进行比较。

7. 根据权利要求1所述的方法,其中,所述方法还包括:

若接收到对目标维度的待查询数据的查询请求,则将所述待查询数据与所述索引树的当前节点的索引区间进行比较;

若所述待查询数据在所述索引区间,则从所述当前节点对应的叶节点查询所述待查询数据;

若所述待查询数据不在所述索引区间,则返回。

8. 根据权利要求1所述的方法,其特征在于,所述方法还包括:

获取所述数据表的所述预定字段对应的第二批数据,根据所述目标维度对所述第二批数据进行排序;

在生成所述第二批数据的索引的过程中,若所述第二批数据中存在与所述第一批数据中的数据具有相同大小的数据;则对所述第二批数据与所述第一批数据中具有相同大小的数据进行合并处理。

9. 根据权利要求1所述的方法,其中,所述索引树为平衡二叉树,所述生成所述索引树的非叶节点的索引,包括:

根据所述平衡二叉树的生成过程,对所述索引树的非叶节点进行左右旋转或者双向旋转,以使所述索引树保持平衡。

10. 根据权利要求1至2或4至9中任一项所述的方法,其中,所述数据为键值对数据,所述键值对包括字段值和对应的行号。

11. 根据权利要求1至2或4至9中任一项所述的方法,其中,所述索引树为一维BKD树。

12. 一种索引生成装置,所述装置包括:

排序模块,用于获取数据表的预定字段对应的第一批数据,根据目标维度对所述第一批数据进行排序,所述数据为至少一个维度的数值类型数据,所述目标维度为从所述至少一个维度中选取的维度;

叶节点生成模块,用于根据排序后的顺序从所述第一批数据中获取预定数量个数据,为所述预定数量个数据配置索引树的叶节点;

索引写入模块,用于将所述叶节点对应的数据写入到索引文件,返回与所述叶节点对应的文件指针,所述文件指针包含所述叶节点对应的数据在所述索引文件中的位置信息;

非叶节点生成模块,用于在完成所述第一批数据的写入的情况下,根据至少一个叶节点中每个叶节点对应的数据以及文件指针,生成所述索引树的非叶节点的索引,所述非叶节点的索引包含所述非叶节点对应的叶节点的数据在所述目标维度上的取值范围以及对应的叶节点的所述文件指针。

13. 根据权利要求12所述的装置,其中,所述非叶节点生成模块具体包括:

节点确定单元,用于确定所述索引树的当前非叶节点对应的叶节点;

值确定单元,用于确定所述当前非叶节点对应的叶节点的数据在所述目标维度上的最

大值与最小值；

索引生成单元,用于基于所述当前非叶节点对应的叶节点的数据在所述目标维度上的最大值与最小值以及所述文件指针,生成所述当前非叶节点的索引。

14. 一种计算设备,包括:存储器和处理器;

所述存储器用于存储计算机可执行指令,所述计算机可执行指令被所述处理器执行时实现权利要求1至11中任意一项所述的索引生成方法。

15. 一种计算机可读存储介质,其存储有计算机指令,所述计算机指令被处理器执行时实现权利要求1至11中任意一项所述的索引生成方法。

索引生成方法及装置

技术领域

[0001] 本说明书涉及大数据技术领域,特别涉及一种索引生成方法。本说明书同时涉及一种索引生成装置,一种计算设备,以及一种计算机可读存储介质。

背景技术

[0002] 索引是影响数据库性能的关键因素。目前,业务系统的数据库的数据规模越来越大,如何对数据库中的海量数据进行索引成为了关注的焦点。

[0003] 传统的数据库采用树形索引例如B+树索引,B+树索引一般适用于行式存储数据库,但是在分析型数据库中,经常会使用列式存储。在列式存储数据库中,使用传统的数据库索引例如B+树索引,在处理海量数据时,由于I/O操作次数较多,会导致数据库性能的降低的问题。

[0004] 因此,如何降低数据库的I/O操作次数,提高数据库的性能成为了亟待解决的技术难题。

发明内容

[0005] 有鉴于此,本说明书实施例提供了一种索引生成方法。本说明书同时涉及一种索引生成装置,一种计算设备,以及一种计算机可读存储介质,以解决现有技术中存在的技术缺陷。

[0006] 根据本说明书实施例的第一方面,提供了一种索引生成方法,所述方法包括:

[0007] 获取数据表的预定字段对应的第一批数据,根据目标维度对所述第一批数据进行排序,所述数据为至少一个维度的数值类型数据,所述目标维度为从所述至少一个维度中选取的维度;

[0008] 根据排序后的顺序从所述第一批数据中获取预定数量个数据,为所述预定数量个数据配置索引树的叶节点;

[0009] 将所述叶节点对应的数据写入到索引文件,返回与所述叶节点对应的文件指针,所述文件指针包含所述叶节点对应的数据在所述索引文件中的位置信息;

[0010] 在完成所述第一批数据的写入的情况下,根据至少一个叶节点中每个叶节点对应的数据以及文件指针,生成所述索引树的非叶节点的索引,所述非叶节点的索引包含所述非叶节点对应的叶节点的数据在所述目标维度上的取值范围以及对应的叶节点的所述文件指针。

[0011] 在本说明书的一些示例实施例中,所述根据至少一个叶节点中每个叶节点对应的数据以及文件指针,生成所述索引树的非叶节点的索引,包括:

[0012] 确定所述索引树的当前非叶节点对应的叶节点;

[0013] 确定所述当前非叶节点对应的叶节点的数据在所述目标维度上的最大值与最小值;

[0014] 基于所述当前非叶节点对应的叶节点的数据在所述目标维度上的最大值与最小

值以及所述文件指针,生成所述当前非叶节点的索引。

[0015] 在本说明书的一些示例实施例中,在所述根据目标维度对所述第一批数据进行排序之前,所述方法还包括:

[0016] 确定所述第一批数据的各个维度的最大值与最小值;

[0017] 确定各个所述维度的最大值与最小值中差值最大的维度,将所述差值最大的维度作为所述目标维度。

[0018] 在本说明书的一些示例实施例中,所述取值范围为所述索引树的当前节点的索引区间,所述方法还包括:

[0019] 若接收到对所述目标维度的待查询区间的查询请求,则将所述待查询区间与所述索引树的当前节点的索引区间进行比较;

[0020] 若所述当前节点的索引区间在所述待查询区间内,则递归获取所述当前节点的各个子节点的数据,直到获取到所述当前节点对应的叶节点的数据;

[0021] 若所述当前节点的索引区间在所述待查询区间之外,则返回;

[0022] 若所述当前节点的索引区间与所述待查询区间相交,则基于所述当前节点的节点类型进行查询,所述节点类型包括叶节点和非叶节点。

[0023] 在本说明书的一些示例实施例中,所述索引树为平衡二叉树,所述非叶节点的索引包括所述非叶节点的左子节点的指针与右子节点的指针,所述基于所述当前节点的节点类型进行查询,包括:

[0024] 若所述当前节点为叶节点,则将所述当前节点的每个数据的值与所述待查询区间比较,获取满足所述待查询区间的数据;

[0025] 若所述当前节点为非叶节点,则确定所述当前节点的左子节点和右子节点,并分别以所述左子节点和所述右子节点作为所述当前节点,基于所述当前节点的节点类型进行查询。

[0026] 在本说明书的一些示例实施例中,所述将所述待查询区间与所述索引树的当前节点的索引区间进行比较,包括:

[0027] 从所述索引树的根节点开始,递归获取所述索引树的下一节点;

[0028] 将所述待查询区间与递归到的当前节点的索引区间进行比较。

[0029] 在本说明书的一些示例实施例中,所述方法还包括:

[0030] 若接收到对目标维度的待查询数据的查询请求,则将所述待查询数据与所述索引树的当前节点的索引区间进行比较;

[0031] 若所述待查询数据在所述索引区间,则从所述当前节点对应的叶节点查询所述待查询数据;

[0032] 若所述待查询数据不在所述索引区间,则返回。

[0033] 在本说明书的一些示例实施例中,所述方法还包括:

[0034] 获取所述数据表的所述预定字段对应的第二批数据,根据所述目标维度对所述第二批数据进行排序;

[0035] 在生成所述第二批数据的索引的过程中,若所述第二批数据中存在与所述第一批数据中的数据具有相同大小的数据;则对所述第二批数据与所述第一批数据中具有相同大小的数据进行合并处理。

[0036] 在本说明书的一些示例实施例中,所述索引树为平衡二叉树,所述生成所述索引树的非叶节点的索引,包括:

[0037] 根据所述平衡二叉树的生成过程,对所述索引树的非叶节点进行左右旋转或者双向旋转,以使所述索引树保持平衡。

[0038] 在本说明书的一些示例实施例中,所述数据为键值对数据,所述键值对包括字段值和对应的行号。

[0039] 在本说明书的一些示例实施例中,所述索引树为一维BKD树。

[0040] 根据本说明书实施例的第二方面,提供了一种索引生成装置,所述装置包括:

[0041] 排序模块,用于获取数据表的预定字段对应的第一批数据,根据目标维度对所述第一批数据进行排序,所述数据为至少一个维度的数值类型数据,所述目标维度为从所述至少一个维度中选取的维度;

[0042] 叶节点生成模块,用于根据排序后的顺序从所述第一批数据中获取预定数量个数据,为所述预定数量个数据配置索引树的叶节点;

[0043] 索引写入模块,用于将所述叶节点对应的数据写入到索引文件,返回与所述叶节点对应的文件指针,所述文件指针包含所述叶节点对应的数据在所述索引文件中的位置信息;

[0044] 非叶节点生成模块,用于在完成所述第一批数据的写入的情况下,根据至少一个叶节点中每个叶节点对应的数据以及文件指针,生成所述索引树的非叶节点的索引,所述非叶节点的索引包含所述非叶节点对应的叶节点的数据在所述目标维度上的取值范围以及对应的叶节点的所述文件指针。

[0045] 在本说明书的一些示例实施例中,所述非叶节点生成模块具体包括:

[0046] 节点确定单元,用于确定所述索引树的当前非叶节点对应的叶节点;

[0047] 值确定单元,用于确定所述当前非叶节点对应的叶节点的数据在所述目标维度上的最大值与最小值;

[0048] 索引生成单元,用于基于所述当前非叶节点对应的叶节点的数据在所述目标维度上的最大值与最小值以及所述文件指针,生成所述当前非叶节点的索引。

[0049] 在本说明书的一些示例实施例中,在所述根据目标维度对所述第一批数据进行排序之前,所述装置还包括:

[0050] 维度值确定模块,用于确定所述第一批数据的各个维度的最大值与最小值;

[0051] 维度确定模块,用于确定各个所述维度的最大值与最小值中差值最大的维度,将所述差值最大的维度作为所述目标维度。

[0052] 在本说明书的一些示例实施例中,所述取值范围为所述索引树的当前节点的索引区间,所述装置还包括:

[0053] 区间比较模块,用于若接收到对所述目标维度的待查询区间的查询请求,则将所述待查询区间与所述索引树的当前节点的索引区间进行比较;

[0054] 区间内处理模块,用于若所述当前节点的索引区间在所述待查询区间内,则递归获取所述当前节点的各个子节点的数据,直到获取到所述当前节点对应的叶节点的数据;

[0055] 区间外处理模块,用于若所述当前节点的索引区间在所述待查询区间之外,则返回;

[0056] 区间相交处理模块,用于若所述当前节点的索引区间与所述待查询区间相交,则基于所述当前节点的节点类型进行查询,所述节点类型包括叶节点和非叶节点。

[0057] 在本说明书的一些示例实施例中,所述索引树为平衡二叉树,所述非叶节点的索引包括所述非叶节点的左子节点的指针与右子节点的指针,所述基区间相交处理模块具体还用于:

[0058] 若所述当前节点为叶节点,则将所述当前节点的每个数据的值与所述待查询区间比较,获取满足所述待查询区间的数据库;

[0059] 若所述当前节点为非叶节点,则确定所述当前节点的左子节点和右子节点,并分别以所述左子节点和所述右子节点作为所述当前节点,基于所述当前节点的节点类型进行查询。

[0060] 在本说明书的一些示例实施例中,所述区间比较模块具体还用于:

[0061] 从所述索引树的根节点开始,递归获取所述索引树的下一节点;

[0062] 将所述待查询区间与递归到的当前节点的索引区间进行比较。

[0063] 在本说明书的一些示例实施例中,所述装置还包括:

[0064] 值比较模块,用于若接收到对目标维度的待查询数据的查询请求,则将所述待查询数据与所述索引树的当前节点的索引区间进行比较;

[0065] 等值查询模块,用于若所述待查询数据在所述索引区间,则从所述当前节点对应的叶节点查询所述待查询数据;若所述待查询数据不在所述索引区间,则返回。

[0066] 在本说明书的一些示例实施例中,所述装置还包括:

[0067] 合并处理模块,用于获取所述数据表的所述预定字段对应的第二批数据,根据所述目标维度对所述第二批数据进行排序;在生成所述第二批数据的索引的过程中,若所述第二批数据中存在与所述第一批数据中的数据具有相同大小的数据;则对所述第二批数据与所述第一批数据中具有相同大小的数据进行合并处理。

[0068] 在本说明书的一些示例实施例中,所述索引树为平衡二叉树,所述非叶节点生成模块包括:

[0069] 旋转单元,用于根据所述平衡二叉树的生成过程,对所述索引树的非叶节点进行左右旋转或者双向旋转,以使所述索引树保持平衡。

[0070] 在本说明书的一些示例实施例中,所述数据为键值对数据,所述键值对包括字段值和对应的行号。

[0071] 在本说明书的一些示例实施例中,所述索引树为一维BKD树。

[0072] 根据本说明书实施例的第三方面,提供了一种计算设备,包括:

[0073] 存储器和处理器;

[0074] 所述存储器用于存储计算机可执行指令,所述计算机可执行指令被所述处理器执行时实现如上述第一方面所述的索引生成方法。

[0075] 根据本说明书实施例的第四方面,提供了一种计算机可读存储介质,其存储有计算机可执行指令,所述计算机可执行指令被处理器执行时实现上述第一方面所述的索引生成方法。

[0076] 本说明书实施例提供的技术方案,一方面,由于对数据表的预定字段的第一批数据排序后进行存储,预定字段的数据类型相同,从而能够对第一批数据进行高效地压缩,降

低数据库索引占用的存储空间；另一方面，由于第一批数据在磁盘中按照顺序存储，数据的聚集性较好，不仅提高了数据写入速度，而且能够在查询该预定字段时顺序读取数据，从而能够降低I/O操作的次数；再一方面，索引树的非叶节点存储对应的叶节点的指针以及索引范围，能够通过索引范围以及指针快速定位到待查询的数据，从而能够进一步地降低I/O操作的次数，进而能够提高数据库的性能。

附图说明

- [0077] 图1是本说明书一实施例提供的索引生成方法的应用场景的示意图；
- [0078] 图2是本说明书一实施例提供的一种索引生成方法的流程示意图；
- [0079] 图3是根据本说明书一实施例提供的索引生成方法生成的索引树的示意图；
- [0080] 图4是本说明书一实施例提供的一种范围查询过程的流程示意图；
- [0081] 图5是本说明书一实施例提供的一种等值查询过程的流程示意图；
- [0082] 图6是本说明书一实施例提供的一种索引生成装置的结构示意图；
- [0083] 图7是本说明书一实施例提供的一种非叶节点生成模块的结构示意图；
- [0084] 图8是本说明书一实施例提供的一种计算设备的结构框图。

具体实施方式

[0085] 在下面的描述中阐述了很多具体细节以便于充分理解本说明书。但是本说明书能够以很多不同于在此描述的其它方式来实施，本领域技术人员可以在不违背本说明书内涵的情况下做类似推广，因此本说明书不受下面公开的具体实施的限制。

[0086] 在本说明书一个或多个实施例中使用的术语是仅仅出于描述特定实施例的目的，而非旨在限制本说明书一个或多个实施例。在本说明书一个或多个实施例和所附权利要求书中所使用的单数形式的“一种”、“所述”和“该”也旨在包括多数形式，除非上下文清楚地表示其他含义。还应当理解，本说明书一个或多个实施例中使用的术语“和/或”是指并包含一个或多个相关联的列出项目的任何或所有可能组合。

[0087] 应当理解，尽管在本说明书一个或多个实施例中可能采用术语第一、第二等来描述各种信息，但这些信息不应限于这些术语。这些术语仅用来将同一类型的信息彼此区分开。例如，在不脱离本说明书一个或多个实施例范围的情况下，第一也可以被称为第二，类似地，第二也可以被称为第一。取决于语境，如在此所使用的词语“如果”可以被解释成为“在……时”或“当……时”或“响应于确定”。

[0088] 首先，对本说明书一个或多个实施例涉及的名词术语进行解释。

[0089] 索引：是一种单独的、物理的对数据库表中一列或多列的值进行排序的一种存储结构，索引提供指向存储在表的指定列中的数据值的指针。索引树：指的是树形索引，包括根节点、非叶节点以及叶节点。

[0090] 叶节点：存储数据的节点，一个叶节点可以存储一个块block大小的数据，叶节点存储的数据有序。

[0091] 非叶节点：作为索引数据的索引，非叶节点的索引主要存该非叶节点对应的子节点的最小值、最大值以及子节点的指针。

[0092] 块：block，指的是固定数量的键值对(key,value)，比如一个block定义大小为

128,就是包含128对键值对,每对键值对是定长的。

[0093] BKD(Block K-Dimension Balanced,块K维平衡)树:一组KDB树(K-Dimension Balanced Tree,K维平衡树)的集合,BKD树是一种保存多维空间点的数据结构,能够用于数值类型(包括空间点)的快速查找。

[0094] 页:是计算机管理存储器的逻辑块,硬件及操作系统往往将主存和磁盘存储区分割为连续的大小相等的块,每个存储块称为一页(在许多操作系统中,页的大小通常为4k)

[0095] 传统的数据库索引采用树形索引例如B+树索引,一般适用于行式存储数据库,在列式存储的数据库中,使用传统的数据库索引,在处理海量数据时,由于I/O操作次数较多,会降低数据库的性能。

[0096] 基于上述内容,本说明书实施例的基本思想在于,采用适于列式存储的树形索引,按照目标维度对数据表的预定字段或列对应的第一批数据进行排序,将排序后的数据存储到索引树的叶节点,索引树的非叶节点存储对应的叶节点的指针以及索引范围。根据本说明书实施例的技术方案,一方面,由于对数据表的预定字段的第一批数据排序后进行存储,预定字段的数据类型相同,从而能够对第一批数据进行高效地压缩,降低数据库索引占用的存储空间;另一方面,由于第一批数据在磁盘中按照顺序存储,数据的聚集性较好,不仅提高了数据写入速度,而且能够在查询该预定字段时顺序读取数据,从而能够降低I/O操作的次数;再一方面,索引树的非叶节点存储对应的叶节点的指针以及索引范围,能够通过索引范围以及指针快速定位到待查询的数据,从而能够进一步地降低I/O操作的次数,进而能够提高数据库的性能。

[0097] 图1是本说明书一实施例提供的索引生成方法的应用场景的示意图。参照图1所示,该应用场景包括数据处理装置110以及一个或第一批数据库120,数据处理装置110与一个或第一批数据库120通过有线或无线网络130进行通信,数据库120中存储有待生成索引的数据。其中,数据处理装置110用于获取数据库120的数据表中的预定字段或列对应的数据,按照目标维度对获取的第一批数据进行排序,将排序后的数据存储到索引树的多个叶节点,生成索引树的非叶节点的索引,该预定字段对应的数据为至少一个维度的数据类型数据,非叶节点的索引包含非叶节点对应的叶节点的数据在目标维度上的取值范围以及对应的叶节点的文件指针。

[0098] 需要说明的是,数据处理装置110可以为台式计算机、膝上型计算机以及笔记本电脑等,本说明书实施例对此不进行特殊限定。数据库120可以为列式存储数据库例如分析型数据库。

[0099] 下面结合图1的应用场景,参考图2来描述根据本说明书的示例性实施例的索引生成方法。需要注意的是,上述应用场景仅是为了便于理解本说明书实施例的精神和原理而示出,本说明书的实施例在此方面不受任何限制。相反,本说明书的实施例可以应用于适用的任何场景。

[0100] 图2是本说明书一实施例提供的一种索引生成方法的流程示意图。该索引生成方法可以应用于图1的数据处理装置120侧,该处理方法包括步骤S210至步骤S240,下面对图2的示例实施例进行详细的说明。

[0101] 在步骤S210中,获取数据表的预定字段对应的第一批数据,按照目标维度对第一批数据进行排序,预定字段对应的数据为至少一个维度的数值类型数据,目标维度为从至

少一个维度中选取的维度。

[0102] 在示例实施例中,数据表为列式存储数据库中的数据表,预定字段或预定列对应的数据为至少一个维度的数值类型数据。数据类型数据可以为int型数据,也可以为float型数据或double型数据,还可以为其他可以转化为数值类型数据的数据,本说明书实施例对此不进行特殊限定。

[0103] 需要说明的是,预定字段对应的数据可以为二维数据例如整型数据,也可以为二维数据例如经纬度数据,还可以为其他适当维度的数据例如三维空间坐标数据。

[0104] 在示例实施例中,目标维度为从预定字段对应的数据的至少一个维度中选择的维度,例如,若预定字段对应的数据为经纬度数据,则目标维度可以为经度,也可以为纬度,例如预定字段对应的数据为各个城市的经纬度数据,北京市市中心经纬度(116.41667, 39.91667)、上海市区经纬度(121.43333, 34.50000)、天津市区经纬度(117.20000, 39.13333)、香港经纬度:(114.10000, 22.20000)、广州经纬度:(113.23333, 23.16667)等。

[0105] 进一步地,在示例实施例中,根据选取的目标维度,对第一批数据按照从小到大的顺序进行排序。以目标维度为经度为例,可以按照经度的从小到大的顺序,对多个经纬度数据进行排序。

[0106] 在步骤S220中,根据排序后的顺序从第一批数据中获取预定数量个数据,为预定数量个数据配置索引树的叶节点。

[0107] 在示例实施例中,一个叶节点对应预定数量个数据,预定数量个数据可以为根据数据库的数据规模确定的数量,也可以为根据预定大小的存储区域确定的数据,预定大小的存储区域可以为磁盘页的整数倍大小的存储区域。

[0108] 具体而言,在示例实施例中,预定数量个数据为一个block大小的数据,block指的是固定数量的数据。例如,block可以为固定数量的键值对(key, value)数据,比如一个block定义大小为128,就是包含128对键值对,其中,key值为字段值,value值为行号,即字段值到行号的映射,每对键值对的大小是定长的。

[0109] 进一步地,在示例实施例中,对第一批数据进行排序,按照排序的顺序从第一批数据中获取预定数量个数据,为预定数量个数据配置索引树的叶节点。例如,设预定数量为128,从第一批数据中获取128个数据,为获取的128个数据的配置索引树的叶节点,具体而言,设索引树的一个叶节点对应一个128维的数组例如数组名为leafblock,将128个数据写入到该128维的数组中,该叶节点对应的数组名leafblock与这128个数据具有对应关系。

[0110] 在步骤S230中,将叶节点对应的数据写入到索引文件,返回与叶节点对应的文件指针,文件指针包含叶节点对应的数据在索引文件中的位置信息。

[0111] 在示例实施例中,在生成叶节点之后,将该叶节点对应的预定数量个数据写入到索引文件,返回与该叶节点对应的文件指针,文件指针包含叶节点对应的数据在索引文件中的位置信息,例如数据在索引文件中的偏移信息。

[0112] 具体而言,设索引树的一个叶节点对应一个128维的数组例如数组名为leafblock,根据该数组名称顺序读取该叶节点对应的数据,将读取的数据写入到索引文件,该叶节点写入完成之后,返回与该叶节点对应的文件指针,该文件指针包含叶节点对应的数据在索引文件中的位置信息例如偏移信息。

[0113] 进一步地,在示例实施例中,预定数量个数据为固定数量的键值对数据,比如128

对键值对,其中,key值为字段值,value值为行号。索引文件中可以存储叶节点对应的键值对数据。

[0114] 在步骤S240中,在完成数据表的预定字段对应的第一批数据的写入的情况下,根据至少一个叶节点中每个叶节点对应的数据以及文件指针,生成索引树的非叶节点的索引,非叶节点的索引包含非叶节点对应的叶节点的数据在目标维度上的取值范围以及对应的叶节点的文件指针。

[0115] 在示例实施例中,在第一批数据写入完成时,根据当前非叶节点对应的叶节点的数据的大小,确定当前非叶节点对应的叶节点的取值范围,根据当前非叶节点对应的叶节点的取值范围以及对应的叶节点的文件指针,生成该当前非叶节点的索引。

[0116] 具体而言,在示例实施例中,设当前非叶节点的索引为一个类对象即Non-Leaf,该类对象包含对应的叶节点的取值范围以及对应的叶节点的文件指针,确定索引树的当前非叶节点对应的叶节点;确定当前非叶节点对应的叶节点的数据在目标维度上的最大值与最小值即取值范围;将当前非叶节点对应的叶节点在目标维度上的最大值与最小值以及文件指针赋值给该类对象的元素,将该类对象写入到索引文件中。例如,设当前非叶节点对应两个叶节点,这两个叶节点的目标维度的取值范围分别为[1,20]、[21,40],则当前非叶节点的索引中的取值范围为[1,40]。

[0117] 根据图2的示例实施例中的技术方案,一方面,由于对数据表的预定字段的第一批数据排序后进行存储,预定字段的数据类型相同,从而能够对第一批数据进行高效地压缩,降低数据库索引占用的存储空间;另一方面,由于第一批数据在磁盘中按照顺序存储,数据的聚集性较好,不仅提高了数据写入速度,而且能够在查询该预定字段时顺序读取数据,从而能够降低I/O操作的次数;再一方面,索引树的非叶节点存储对应的叶节点的指针以及索引范围,能够通过索引范围以及指针快速定位到待查询的数据,从而能够进一步地降低I/O操作的次数,进而能够提高数据库的性能。

[0118] 进一步地,在示例实施例中,在根据目标维度对数据表的预定字段的第一批数据进行排序之前,该索引生成方法还包括:确定第一批数据的各个维度的最大值与最小值;确定各个维度的最大值与最小值中差值最大的维度,将所述差值最大的维度作为所述目标维度。此外,在一些示例实施例中,还可以根据各个维度的数据的方差的大小,确定目标维度。通过确定数据表的预定字段的目标维度,能够更好地对数据表的该预定字段的数据进行划分,从而使得生成的索引能够更合理,从而能够降低索引树的高度,减少磁盘的I/O操作,提高数据的查询效率。

[0119] 此外,在示例实施例中,该索引生成方法还包括:获取所述数据表的所述预定字段对应的第二批数据,根据所述目标维度对所述第二批数据进行排序;在生成所述第二批数据的索引的过程中,若所述第二批数据中存在与所述第一批数据中的数据具有相同大小的数据;则对所述第二批数据与所述第一批数据中具有相同大小的数据进行合并处理。

[0120] 需要说明的是,数据表的预定字段对应的数据可以有多个批次例如3个以上批次的的数据,每次获取该预定字段对应的一批数据,生成对应的索引树,各个批次的索引树最终可以形成一个总的索引树。若不同批次的数据中存在具有相同大小的数据,则将后一批次数据与之前批次数据中具有相同大小的数据进行合并处理。

[0121] 具体而言,在示例实施例中,预定数量个数据为固定数量的键值对数据,比如128

对键值对,其中,key值为字段值,value值为行号。在生成索引树的过程中,若第一批数据的索引树的叶节点对应的数据与第二批数据的索引树的叶节点对应的数据存在大小相同的数据,即key值相同,则将第二批数据的索引树的该key值相同的数据的行号合并或添加到第一批数据的索引树的对应数据的value值里。

[0122] 通过对不同批次中具有相同大小的数据进行合并处理,能够减少索引树的节点的数据量,从而能够减少磁盘的I/O操作,提高数据的查询效率。

[0123] 进一步地,在示例实施例中,索引树为平衡二叉树,生成索引树的非叶节点的索引,包括:根据平衡二叉树的生成过程,对索引树的非叶节点进行左右旋转或者双向旋转,以使索引树保持平衡。通过对索引树进行平衡处理,能够降低索引树的高度,减少磁盘的I/O操作,提高数据的查询效率。

[0124] 图3是根据本说明书一实施例提供的索引生成方法生成的索引树的示意图。

[0125] 参照图3所示,该索引树为一维BKD树,根节点的索引为[1,80],两个非叶节点的索引分别为[1,40]、[41,80],非叶节点[1,40]包含两个叶子节点,这两个叶子节点分别存储索引为[1,20]、[21,40]的数据,非叶节点[41,80]包含两个叶子节点,这两个叶子节点分别存储索引为[41,60]、[61,80]的数据。

[0126] 进一步地,该索引树具有以下特点:(1)叶节点存储了所有的数据;(2)非叶节点作为BKD树的索引,非叶节点的索引主要存当前节点的所有子节点的最小值min、最大值max,以及当前节点左右子节点的指针;(3)一个叶节点可以存一个块block的数据,并不只存单个数据,叶节点的数据有序;(4)整个BKD树是有序的。

[0127] 根据图3所示的索引树,一方面,由于索引树对应的第一批数据有序,第一批数据在磁盘中按照顺序存储,不仅提高了数据写入速度,而且能够在查询该预定字段时顺序读取数据,从而能够降低I/O操作的次数;再一方面,索引树的非叶节点存储对应的叶节点的指针以及索引范围,能够通过索引范围以及指针快速定位到待查询的数据,从而能够进一步地降低I/O操作的次数,进而能够提高数据库的性能。

[0128] 图4是本说明书一实施例提供的一种范围查询过程的流程示意图。

[0129] 参照图4所示,在步骤S410中,接收对目标维度的待查询区间的查询请求。

[0130] 在示例实施例中,接收到对目标维度的待查询区间的查询请求时,采用正则表达式从该查询请求中提取该待查询区间。例如,设数据表的预定字段为一维字段,待查询区间range为[20,40],根据正则表达式提取字符串“range”对应的区间[20,40]。

[0131] 在步骤S420中,将待查询区间与索引树的当前节点的索引区间进行比较。

[0132] 在示例实施例中,从索引树的根节点开始,递归获取所述索引树的下一节点,将获取到的当前节点的索引区间与待查询区间进行比较。

[0133] 具体而言,参照图3所示,设索引树的当前节点为非叶节点,该非叶节点的索引区间为[1,40],若待查询区间为[1,60],则确定当前节点的索引区间在该待查询区间内即CELL_INSIDE_QUERY,也就是说,当前节点对应的所有数据均满足查询条件;若待查询区间为[41,60],则确定当前节点的索引区间在该待查询区间之外即CELL_OUTSIDE_QUERY,也就是说,当前节点对应的所有数据均不满足查询条件;若待查询区间为[20,60],则确定当前节点的索引区间与待查询区间相交即CELL_CROSSES_QUERY,也就是说,当前节点有部分数据满足待查询区间的查询条件。

[0134] 在步骤S430中,若当前节点的索引区间在待查询区间内,则递归获取当前节点的各个子节点的数据,直到获取到当前节点对应的叶节点的数据。

[0135] 在示例实施例中,若当前节点的索引区间在待查询区间内,则表示当前节点对应的所有数据均满足查询条件,则递归获取当前节点的各个子节点的数据,直到获取当前节点对应的每个叶节点的数据为止。例如,该非叶节点的索引区间为[1,40],若待查询区间为[1,60],获取到非叶节点对应的叶节点的[1,20]、[21,40]对应的数据。

[0136] 在步骤S440中,若当前节点的索引区间在待查询区间之外,则返回。

[0137] 在示例实施例中,若当前节点的索引区间在待查询区间之外,则表示当前节点对应的所有数据均不满足条件,则返回。例如,该非叶节点的索引区间为[1,40],若待查询区间为[41,60],则确定当前节点的索引区间在该待查询区间之外即CELL_OUTSIDE_QUERY,返回从该索引树查找不到对应的数据的信息,从下一个索引树开始重新查找。

[0138] 在步骤S450中,若当前节点的索引区间与待查询区间相交,则基于当前节点的节点类型进行查询,节点类型包括叶节点和非叶节点。

[0139] 在示例实施例中,若当前节点的索引区间与待查询区间相交,则确定当前节点的部分数据满足查询条件,若当前节点为叶节点,则将当前节点的每个数据的值与待查询区间比较,获取满足待查询区间的查询数据。进一步地,若当前节点为非叶节点,则确定当前节点的左子节点和右子节点,并分别以左子节点和右子节点作为当前节点,基于所述当前节点的节点类型进行查询。

[0140] 具体而言,设当前节点的索引区间为[1,40],若待查询区间为[10,60],则确定当前节点的索引区间与待查询区间相交即CELL_CROSSES_QUERY,也就是说,当前节点有部分数据满足待查询区间的查询条件,若当前节点为非叶节点,则获取当前节点的左子节点和右子节点即节点[1,20]与节点[21,40],将当前节点的左子节点和右子节点分别作为当前节点,基于当前节点的节点类型进行查询。例如,当前节点[1,20]与节点[21,40]为叶节点,则将当前节点的每个数据的值与待查询区间[10,60]比较,获取满足待查询区间的查询数据。

[0141] 根据图4的示例实施例中的技术方案,一方面,由于对数据表的预定字段的第一批数据排序后进行存储,第一批数据在磁盘中按照顺序存储,数据的聚集性较好,从而能够在查询该预定字段时顺序读取数据,进而能够降低I/O操作的次数;另一方面,索引树的非叶节点存储对应的叶节点的指针以及索引范围,能够通过索引范围以及指针快速定位到待查询的数据,从而能够进一步地降低I/O操作的次数,进而能够提高数据库的性能。

[0142] 图5是本说明书一实施例提供的一种等值查询过程的流程示意图。

[0143] 参照图5所示,在步骤S510中,接收对目标维度的待查询数据的查询请求,确定待查询数据。

[0144] 在示例实施例中,接收到对目标维度的待查询数据的查询请求时,采用正则表达式从该查询请求中提取该待查询数据。例如,设数据表的预定字段为一维字段,待查询数据为30,根据正则表达式提取字符串“value”对应的值30。

[0145] 在步骤S520中,将待查询数据与所述索引树的当前节点的索引区间进行比较,若待查询数据与当前节点的索引区间相交则进行至步骤S540,若待查询数据与当前节点的索引区间不相交,则进行至步骤S550。

[0146] 在步骤S540中,若待查询数据在索引区间,则从当前节点对应的叶节点查询待查

询数据。

[0147] 在示例实施例中,设该当前节点的索引区间为[1,40],若待查询数据为30,则确定待查询数据在当前节点的索引区间内,从当前节点对应的叶节点中查询该待查询数据。例如,若当前节点为非叶节点,则根据该非叶节点对应的叶节点的文件指针,确定对应的叶节点的数据,从该叶节点的数据中获取该待查询数据。

[0148] 在步骤S550中,若待查询数据不在索引区间,则返回。

[0149] 在示例实施例中,若待查询数据不在索引区间,则表示当前节点对应的所有数据均不满足条件,则返回。例如,该非叶节点的索引区间为[1,40],若待查询数据为50,则确定当前节点的索引区间在该待查询数据之外即CELL_OUTSIDE_QUERY,返回从该索引树查找不到对应的数据的信息,从下一个索引树开始重新查找。

[0150] 根据图5的示例实施例中的技术方案,一方面,由于对数据表的预定字段的第一批数据排序后进行存储,第一批数据在磁盘中按照顺序存储,数据的聚集性较好,从而能够在查询该预定字段时顺序读取数据,进而能够降低I/O操作的次数;另一方面,索引树的非叶节点存储对应的叶节点的指针以及索引范围,能够通过索引范围以及指针快速定位到待查询的数据,从而能够进一步地降低I/O操作的次数,进而能够提高数据库的性能。

[0151] 与上述方法实施例相对应,本说明书还提供了索引生成装置实施例,图7示出了本说明书一实施例提供的一种索引生成装置的结构示意图。

[0152] 参照图6所示,该索引生成装置600包括:

[0153] 排序模块610,用于获取数据表的预定字段对应的第一批数据,根据目标维度对所述第一批数据进行排序,所述数据为至少一个维度的数值类型数据,所述目标维度为从所述至少一个维度中选取的维度;

[0154] 叶节点生成模块620,用于根据排序后的顺序从所述第一批数据中获取预定数量个数据,为所述预定数量个数据配置索引树的叶节点;

[0155] 索引写入模块630,用于将所述叶节点对应的数据写入到索引文件,返回与所述叶节点对应的文件指针,所述文件指针包含所述叶节点对应的数据在所述索引文件中的位置信息;

[0156] 非叶节点生成模块640,用于在完成所述第一批数据的写入的情况下,根据至少一个叶节点中每个叶节点对应的数据以及文件指针,生成所述索引树的非叶节点的索引,所述非叶节点的索引包含所述非叶节点对应的叶节点的数据在所述目标维度上的取值范围以及对应的叶节点的所述文件指针。

[0157] 根据图6的示例实施例中的技术方案,一方面,由于对数据表的预定字段的第一批数据排序后进行存储,预定字段的数据类型相同,从而能够对第一批数据进行高效地压缩,降低数据库索引占用的存储空间;另一方面,由于对数据表的预定字段的第一批数据排序后进行存储,第一批数据在磁盘中按照顺序存储,不仅提高了数据写入速度,而且能够在查询该预定字段时顺序读取数据,从而能够降低I/O操作的次数;再一方面,索引树的非叶节点存储对应的叶节点的指针以及索引范围,能够通过索引范围以及指针快速定位到待查询的数据,从而能够进一步地降低I/O操作的次数,进而能够提高数据库的性能。

[0158] 图7是本说明书一实施例提供的一种非叶节点生成模块的结构示意图。参照图7所示,在本说明书的一些示例实施例中,所述非叶节点生成模块640具体包括:

- [0159] 节点确定单元710,用于确定所述索引树的当前非叶节点对应的叶节点;
- [0160] 值确定单元720,用于确定所述当前非叶节点对应的叶节点的数据在所述目标维度上的最大值与最小值;
- [0161] 索引生成单元730,用于基于所述当前非叶节点对应的叶节点的数据在所述目标维度上的最大值与最小值以及所述文件指针,生成所述当前非叶节点的索引。
- [0162] 在本说明书的一些示例实施例中,在所述根据目标维度对所述第一批数据进行排序之前,所述装置还包括:
- [0163] 维度值确定模块,用于确定所述第一批数据的各个维度的最大值与最小值;
- [0164] 维度确定模块,用于确定各个所述维度的最大值与最小值中差值最大的维度,将所述差值最大的维度作为所述目标维度。
- [0165] 在本说明书的一些示例实施例中,所述取值范围为所述索引树的当前节点的索引区间,所述装置还包括:
- [0166] 区间比较模块,用于若接收到对所述目标维度的待查询区间的查询请求,则将所述待查询区间与所述索引树的当前节点的索引区间进行比较;
- [0167] 区间内处理模块,用于若所述当前节点的索引区间在所述待查询区间内,则递归获取所述当前节点的各个子节点的数据,直到获取到所述当前节点对应的叶节点的数据;
- [0168] 区间外处理模块,用于若所述当前节点的索引区间在所述待查询区间之外,则返回;
- [0169] 区间相交处理模块,用于若所述当前节点的索引区间与所述待查询区间相交,则基于所述当前节点的节点类型进行查询,所述节点类型包括叶节点和非叶节点。
- [0170] 在本说明书的一些示例实施例中,所述索引树为平衡二叉树,所述非叶节点的索引包括所述非叶节点的左子节点的指针与右子节点的指针,所述基区间相交处理模块具体还用于:
- [0171] 若所述当前节点为叶节点,则将所述当前节点的每个数据的值与所述待查询区间比较,获取满足所述待查询区间的数据;
- [0172] 若所述当前节点为非叶节点,则确定所述当前节点的左子节点和右子节点,并分别以所述左子节点和所述右子节点作为所述当前节点,基于所述当前节点的节点类型进行查询。
- [0173] 在本说明书的一些示例实施例中,所述区间比较模块具体还用于:
- [0174] 从所述索引树的根节点开始,递归获取所述索引树的下一节点;
- [0175] 将所述待查询区间与递归到的当前节点的索引区间进行比较。
- [0176] 在本说明书的一些示例实施例中,所述装置还包括:
- [0177] 值比较模块,用于若接收到对目标维度的待查询数据的查询请求,则将所述待查询数据与所述索引树的当前节点的索引区间进行比较;
- [0178] 等值查询模块,用于若所述待查询数据在所述索引区间,则从所述当前节点对应的叶节点查询所述待查询数据;若所述待查询数据不在所述索引区间,则返回。
- [0179] 在本说明书的一些示例实施例中,所述装置还包括:
- [0180] 合并处理模块,用于获取所述数据表的所述预定字段对应的第二批数据,根据所述目标维度对所述第二批数据进行排序;在生成所述第二批数据的索引的过程中,若所述

第二批数据中存在与所述第一批数据中的数据具有相同大小的数据;则对所述第二批数据与所述第一批数据中具有相同大小的数据进行合并处理。

[0181] 在本说明书的一些示例实施例中,所述索引树为平衡二叉树,所述非叶节点生成模块包括:

[0182] 旋转单元,用于根据所述平衡二叉树的生成过程,对所述索引树的非叶节点进行左右旋转或者双向旋转,以使所述索引树保持平衡。

[0183] 在本说明书的一些示例实施例中,所述数据为键值对数据,所述键值对包括字段值和对应的行号。

[0184] 在本说明书的一些示例实施例中,所述索引树为一维BKD树。

[0185] 上述为本实施例的一种索引生成装置的示意性方案。需要说明的是,该索引生成装置的技术方案与上述的索引生成方法的技术方案属于同一构思,索引生成装置的技术方案未详细描述的细节内容,均可以参见上述索引生成方法的技术方案的描述。

[0186] 图8示出了根据本说明书一实施例提供的一种计算设备800的结构框图。该计算设备800的部件包括但不限于存储器810和处理器820。处理器820与存储器810通过总线830相连接,数据库850用于保存数据。

[0187] 计算设备800还包括接入设备840,接入设备840使得计算设备800能够经由一个或多个网络860通信。这些网络的示例包括公用交换电话网(PSTN)、局域网(LAN)、广域网(WAN)、个域网(PAN)或诸如因特网的通信网络的组合。接入设备840可以包括有线或无线的任何类型的网络接口(例如,网络接口卡(NIC))中的一个或多个,诸如IEEE802.11无线局域网(WLAN)无线接口、全球微波互联接入(Wi-MAX)接口、以太网接口、通用串行总线(USB)接口、蜂窝网络接口、蓝牙接口、近场通信(NFC)接口,等等。

[0188] 在本说明书的一个实施例中,计算设备800的上述部件以及图8中未示出的其他部件也可以彼此相连接,例如通过总线。应当理解,图8所示的计算设备结构框图仅仅是出于示例的目的,而不是对本说明书范围的限制。本领域技术人员可以根据需要,增添或替换其他部件。

[0189] 计算设备800可以是任何类型的静止或移动计算设备,包括移动计算机或移动计算设备(例如,平板计算机、个人数字助理、膝上型计算机、笔记本计算机、上网本等)、移动电话(例如,智能手机)、可佩戴的计算设备(例如,智能手表、智能眼镜等)或其他类型的移动设备,或者诸如台式计算机或PC的静止计算设备。计算设备800还可以是移动式或静止式的服务器。

[0190] 处理器820用于执行如下计算机可执行指令:获取数据表的预定字段对应的第一批数据,根据目标维度对所述第一批数据进行排序,所述数据为至少一个维度的数值类型数据,所述目标维度为从所述至少一个维度中选取的维度;根据排序后的顺序从所述第一批数据中获取预定数量个数据,为所述预定数量个数据配置索引树的叶节点;将所述叶节点对应的数据写入到索引文件,返回与所述叶节点对应的文件指针,所述文件指针包含所述叶节点对应的数据在所述索引文件中的位置信息;在完成所述第一批数据的写入的情况下,根据至少一个叶节点中每个叶节点对应的数据以及文件指针,生成所述索引树的非叶节点的索引,所述非叶节点的索引包含所述非叶节点对应的叶节点的数据在所述目标维度上的取值范围以及对应的叶节点的所述文件指针。

[0191] 上述为本实施例的一种计算设备的示意性方案。需要说明的是,该计算设备的技术方案与上述的索引生成方法的技术方案属于同一构思,计算设备的技术方案未详细描述的细节内容,均可以参见上述索引生成方法的技术方案的描述。

[0192] 本说明书一实施例还提供一种计算机可读存储介质,其存储有计算机指令,该指令被处理器执行时以用于:获取数据表的预定字段对应的第一批数据,根据目标维度对所述第一批数据进行排序,所述数据为至少一个维度的数值类型数据,所述目标维度为从所述至少一个维度中选取的维度;根据排序后的顺序从所述第一批数据中获取预定数量个数据,为所述预定数量个数据配置索引树的叶节点;将所述叶节点对应的数据写入到索引文件,返回与所述叶节点对应的文件指针,所述文件指针包含所述叶节点对应的数据在所述索引文件中的位置信息;在完成所述第一批数据的写入的情况下,根据至少一个叶节点中每个叶节点对应的数据以及文件指针,生成所述索引树的非叶节点的索引,所述非叶节点的索引包含所述非叶节点对应的叶节点的数据在所述目标维度上的取值范围以及对应的叶节点的所述文件指针。

[0193] 上述为本实施例的一种计算机可读存储介质的示意性方案。需要说明的是,该存储介质的技术方案与上述的索引生成方法的技术方案属于同一构思,存储介质的技术方案未详细描述的细节内容,均可以参见上述索引生成方法的技术方案的描述。

[0194] 上述对本说明书特定实施例进行了描述。其它实施例在所附权利要求书的范围内。在一些情况下,在权利要求书中记载的动作或步骤可以按照不同于实施例中的顺序来执行并且仍然可以实现期望的结果。另外,在附图中描绘的过程不一定要求示出的特定顺序或者连续顺序才能实现期望的结果。在某些实施方式中,多任务处理和并行处理也是可以的或者可能是有利的。

[0195] 所述计算机指令包括计算机程序代码,所述计算机程序代码可以为源代码形式、对象代码形式、可执行文件或某些中间形式等。所述计算机可读介质可以包括:能够携带所述计算机程序代码的任何实体或装置、记录介质、U盘、移动硬盘、磁碟、光盘、计算机存储器、只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、电载波信号、电信信号以及软件分发介质等。需要说明的是,所述计算机可读介质包含的内容可以根据司法管辖区内立法和专利实践的要求进行适当的增减,例如在某些司法管辖区,根据立法和专利实践,计算机可读介质不包括电载波信号和电信信号。

[0196] 需要说明的是,对于前述的各方法实施例,为了简便描述,故将其都表述为一系列的动作组合,但是本领域技术人员应该知悉,本说明书并不受所描述的动作顺序的限制,因为依据本说明书,某些步骤可以采用其它顺序或者同时进行。其次,本领域技术人员也应该知悉,说明书中所描述的实施例均属于优选实施例,所涉及的动作和模块并不一定是本说明书所必须的。

[0197] 在上述实施例中,对各个实施例的描述都各有侧重,某个实施例中未详述的部分,可以参见其它实施例的相关描述。

[0198] 以上公开的本说明书优选实施例只是用于帮助阐述本说明书。可选实施例并没有详尽叙述所有的细节,也不限制该发明仅为所述的具体实施方式。显然,根据本说明书的内容,可作很多的修改和变化。本说明书选取并具体描述这些实施例,是为了更好地解释本说明书的原理和实际应用,从而使所属技术领域技术人员能很好地理解和利用本说明书。本

说明书仅受权利要求书及其全部范围和等效物的限制。

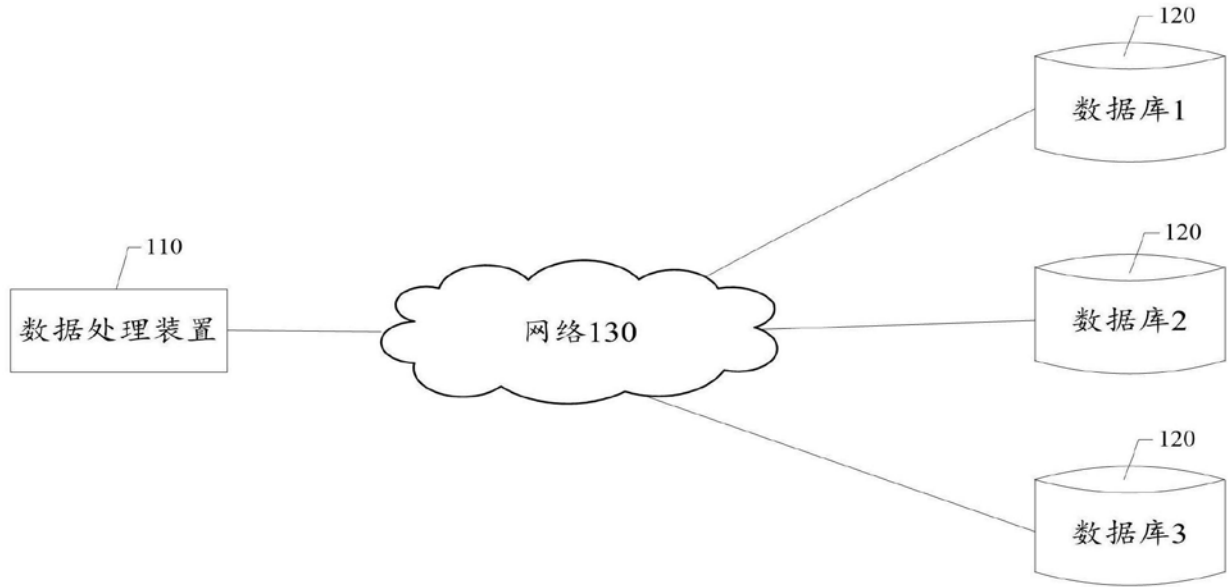


图1

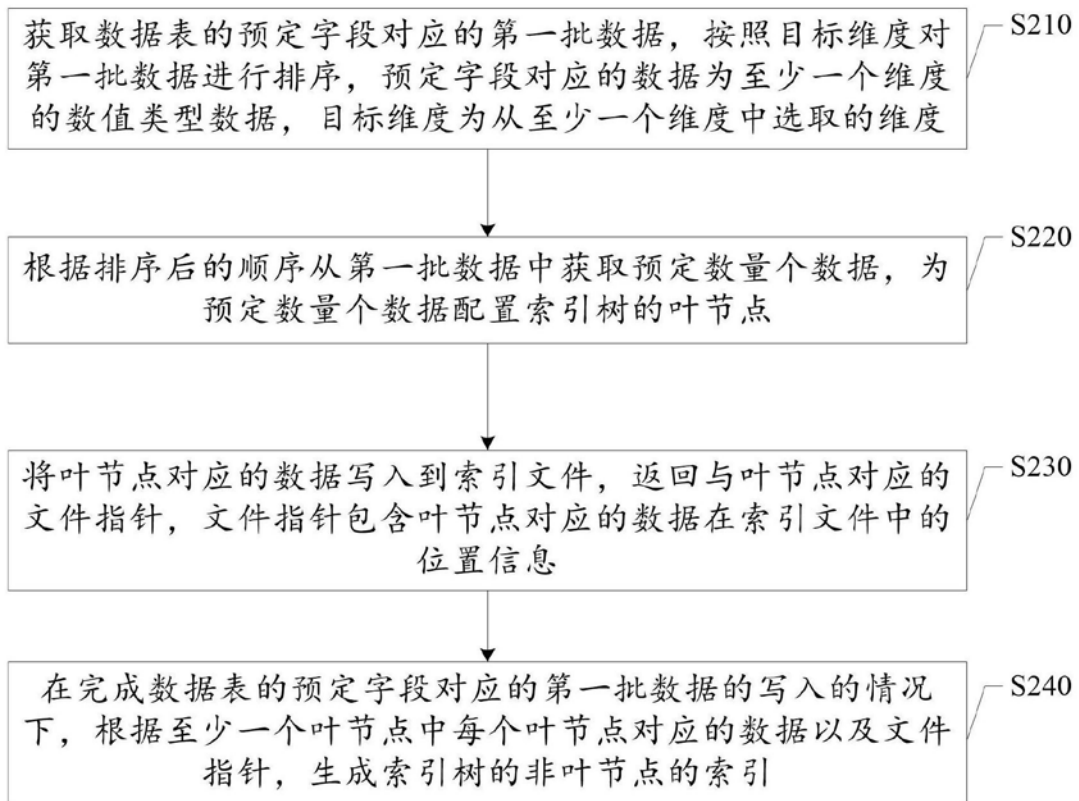


图2

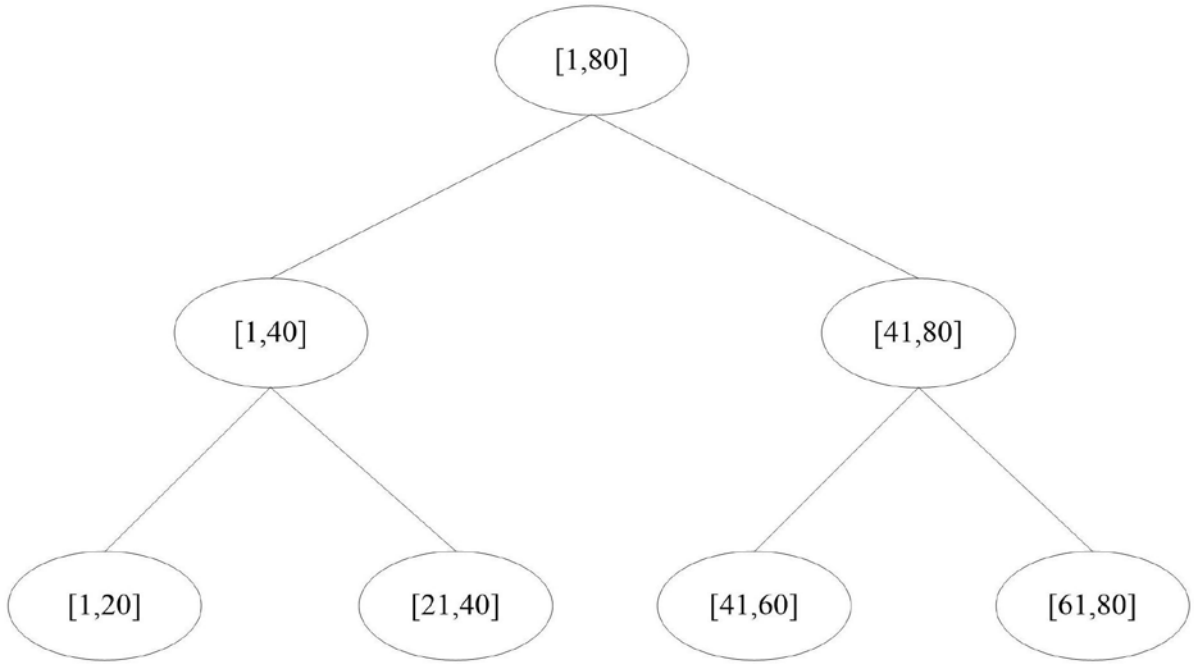


图3

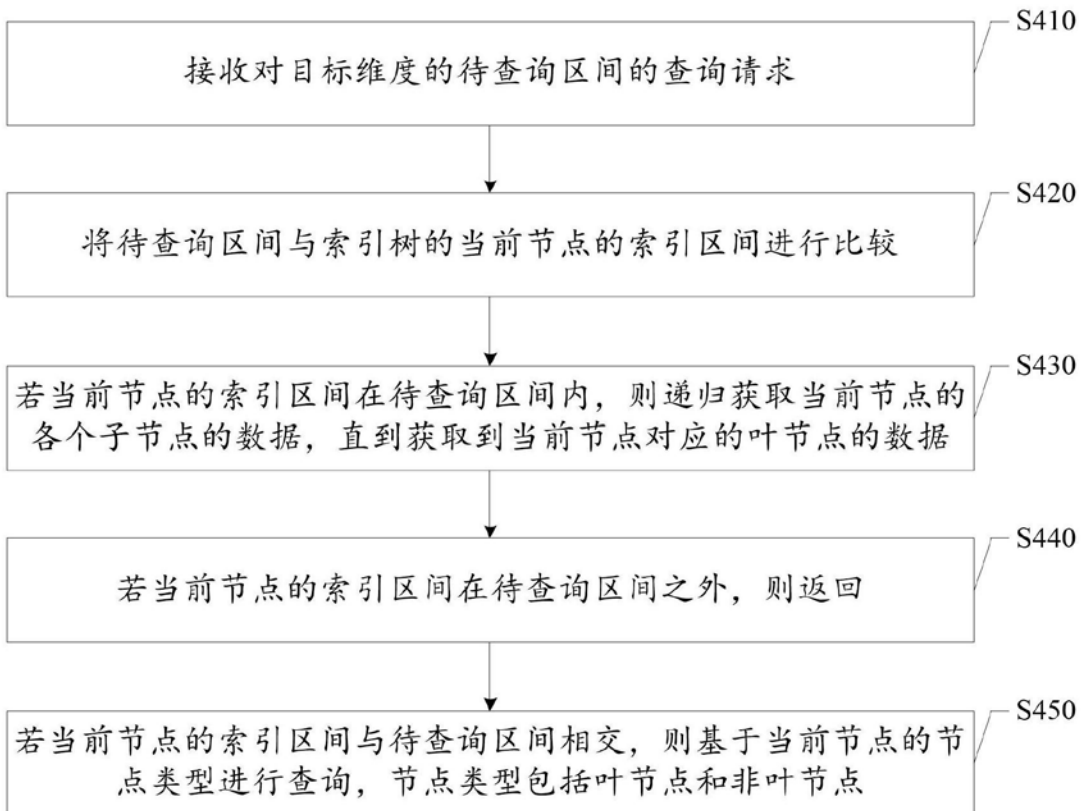


图4

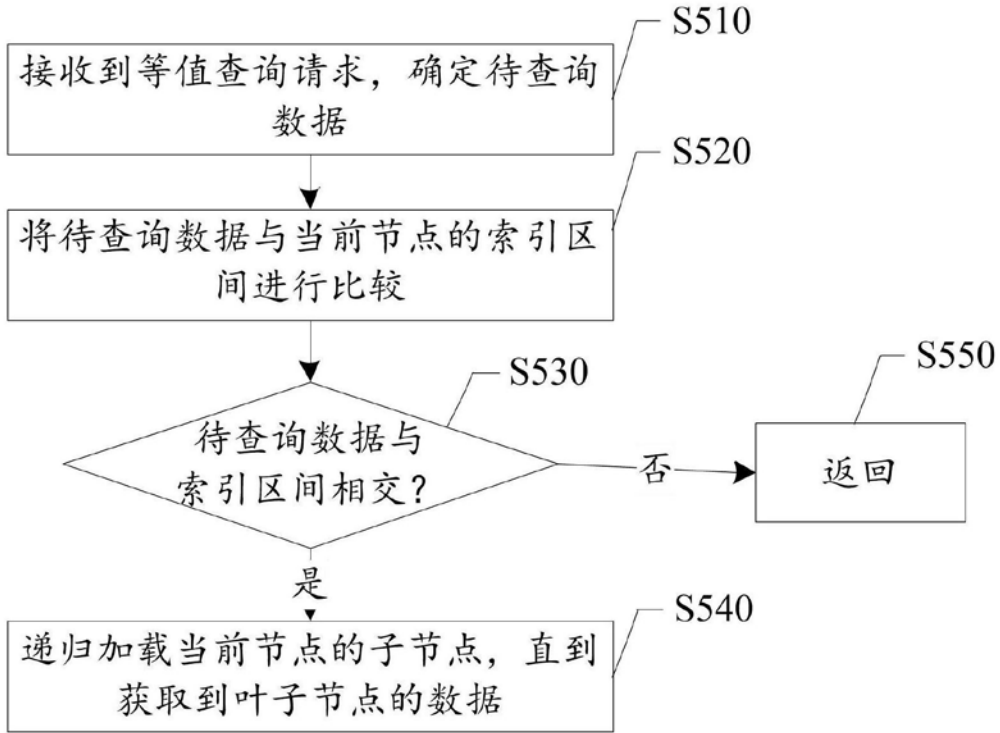


图5

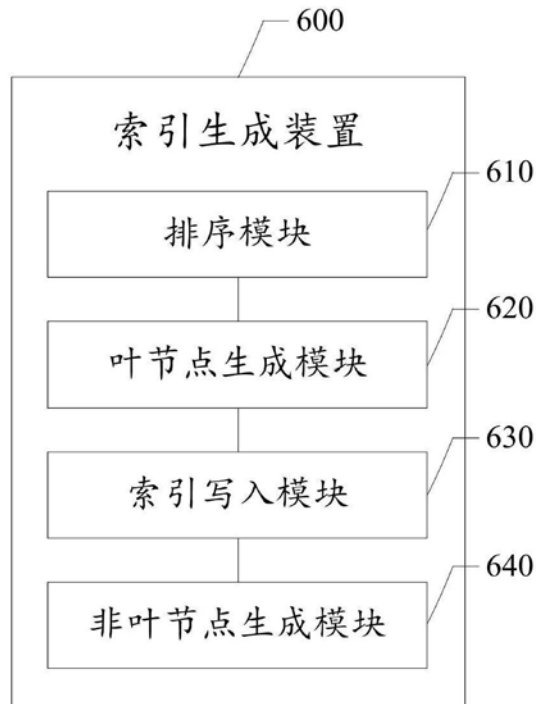


图6

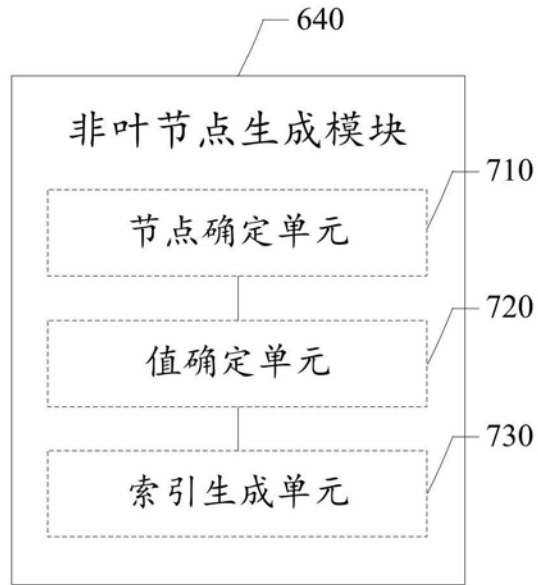


图7

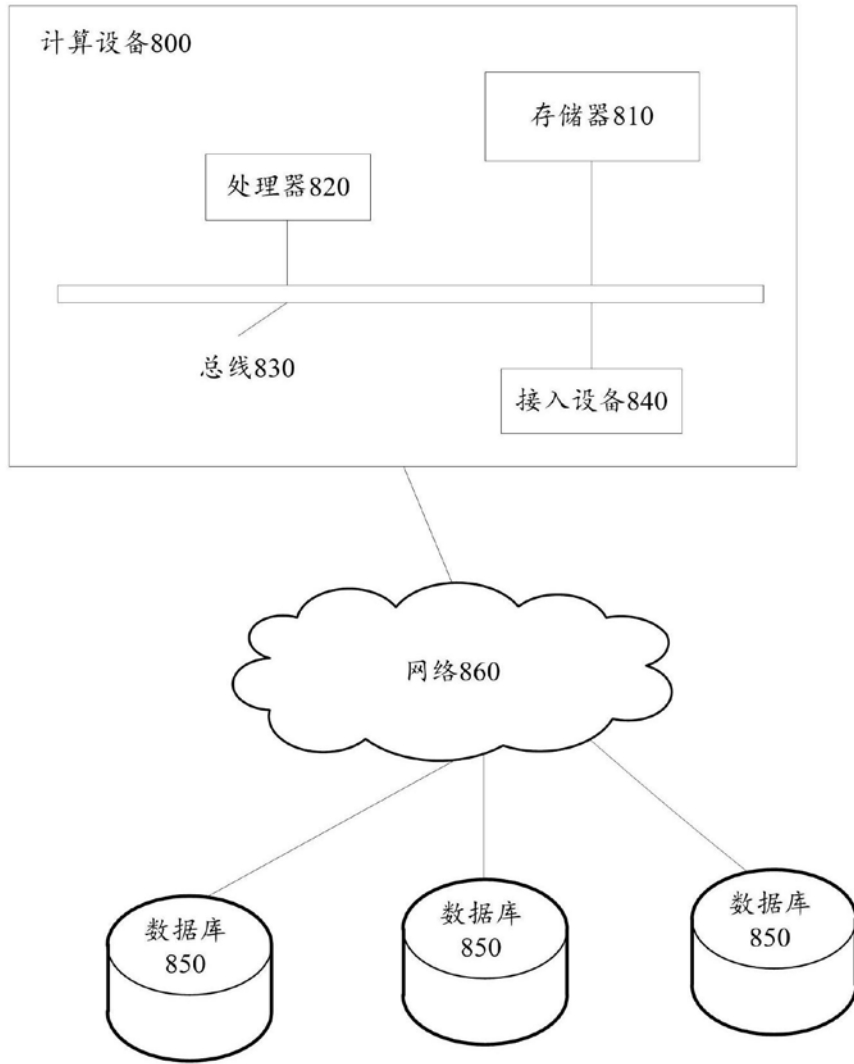


图8