(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 103995855 B
(45)授权公告日 2017.03.08

(21)申请号 201410201909.3

(22)申请日 2014.05.14

(65)同一申请的已公布的文献号
申请公布号 CN 103995855 A

(43)申请公布日 2014.08.20

(73)专利权人 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 刘昕 彭幼武

(74)专利代理机构 北京龙双利达知识产权代理有限公司 11329
代理人 王君 肖鹏

(51)Int.Cl.
G06F 17/30(2006.01)

(56)对比文件
CN 102169504 A,2011.08.31,
US 8612402 B1,2013.12.17,
US 2005102255 A1,2005.05.12,
CN 102214176 A,2011.10.12,

审查员 凌燕翔

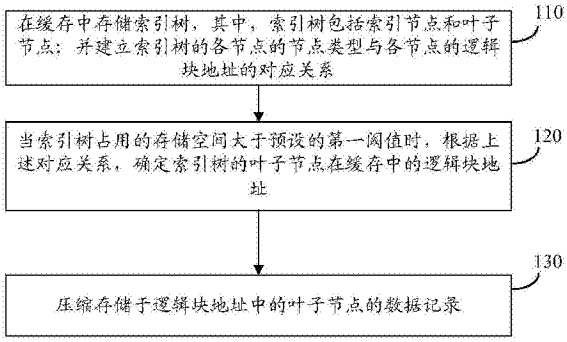
权利要求书2页 说明书10页 附图3页

(54)发明名称

存储数据的方法和装置

(57)摘要

本发明实施例提供一种存储数据的方法和装置。该方法包括：在缓存中存储索引树，其中，该索引树包括索引节点和叶子节点；并建立该索引树的各节点的节点类型与该各节点的逻辑块地址的对应关系；当该索引树占用的存储空间大于预设的第一阈值时，根据该对应关系，确定该索引树的该叶子节点在该缓存中的逻辑块地址；压缩存储于该逻辑块地址中的叶子节点的数据记录。本发明实施例中，通过对缓存中的叶子节点的数据记录进行压缩，从而能够在缓存中存储更多的节点，总体上降低了磁盘IO的次数，提高主机的IOPS。



CN 103995855 B

1.一种存储数据的方法，其特征在于，包括：

在缓存中存储索引树，其中，所述索引树包括索引节点和叶子节点；并建立所述索引树的各节点的节点类型与所述各节点的逻辑块地址的对应关系；

当所述索引树占用的存储空间大于预设的第一阈值时，根据所述对应关系，确定所述索引树的所述叶子节点在所述缓存中的逻辑块地址；

压缩存储于所述逻辑块地址中的叶子节点的数据记录。

2.如权利要求1所述的方法，其特征在于，所述索引树的每个节点对应一个逻辑块号，所述索引树的索引节点存储用于索引查询的关键字和逻辑块号的对应关系，所述索引树的叶子节点存储所述关键字与数据记录的对应关系，

所述建立所述索引树的各节点的节点类型与所述各节点的逻辑块地址的对应关系，包括：

每当所述索引树插入一个新节点时，确定所述新节点的节点类型，并为所述新节点分配逻辑块号；

在所述缓存中为所述新节点分配逻辑块地址，并建立所述新节点的节点类型、所述新节点的逻辑块号和所述新节点的逻辑块地址的对应关系。

3.如权利要求2所述的方法，其特征在于，所述方法还包括：

接收查询消息，所述查询消息携带第一关键字；

根据所述第一关键字与所述索引节点中存储的关键字的大小关系，从所述索引树的根节点开始逐层查找，直到确定第一逻辑块号，所述第一逻辑块号对应的逻辑地址中存储有所述第一关键字对应的数据记录；

根据所述索引树中各节点的逻辑块号和所述各节点的逻辑块地址的对应关系，确定所述第一逻辑块号对应的逻辑块地址；

从所述第一逻辑块号对应的逻辑块地址中获取所述第一关键字对应的数据记录。

4.如权利要求1-3中任一项所述的方法，其特征在于，所述方法还包括：

使用压缩后的所述叶子节点的逻辑块地址，更新所述叶子节点与所述叶子节点的逻辑块地址的对应关系。

5.如权利要求1-3中任一项所述的方法，其特征在于，

当所述索引树占用的存储空间大于预设的所述第一阈值时，压缩存储于所述逻辑块地址中的叶子节点的数据记录，具体包括当所述索引树占用的存储空间大于预设的所述第一阈值而小于第二阈值时，压缩存储于所述逻辑块地址中的叶子节点的数据记录；

所述方法还包括：

当所述索引树占用的存储空间大于所述第二阈值时，根据所述叶子节点在所述缓存中的逻辑块地址，仅将所述叶子节点中的数据记录淘汰至磁盘中。

6.一种存储数据的装置，其特征在于，包括：

缓存单元，用于在缓存中存储索引树，其中，所述索引树包括索引节点和叶子节点；

建立单元，用于建立所述缓存单元存储的所述索引树的各节点的节点类型与所述各节点的逻辑块地址的对应关系；

第一确定单元，用于当所述索引树占用的存储空间大于预设的第一阈值时，根据所述建立单元建立的所述对应关系，确定所述索引树的所述叶子节点在所述缓存中的逻辑块地

址；

压缩单元，用于压缩存储于所述第一确定单元确定的所述逻辑块地址中的叶子节点的数据记录。

7.如权利要求6所述的装置，其特征在于，所述索引树的每个节点对应一个逻辑块号，所述索引树的索引节点存储用于索引查询的关键字和逻辑块号的对应关系，所述索引树的叶子节点存储所述关键字与数据记录的对应关系，

所述建立单元具体用于每当所述索引树插入一个新节点时，确定所述新节点的节点类型，并为所述新节点分配逻辑块号；在所述缓存中为所述新节点分配逻辑块地址，并建立所述新节点的节点类型、所述新节点的逻辑块号和所述新节点的逻辑块地址的对应关系。

8.如权利要求7所述的装置，其特征在于，所述装置还包括：

接收单元，用于接收查询消息，所述查询消息携带第一关键字；

查询单元，用于根据所述第一关键字与所述索引节点中存储的关键字的大小关系，从所述索引树的根节点开始逐层查找，直到确定第一逻辑块号，所述第一逻辑块号对应的逻辑块地址中存储有所述第一关键字对应的数据记录；

第二确定单元，用于根据所述索引树中各节点的逻辑块号和所述各节点的逻辑块地址的对应关系，确定所述第一逻辑块号对应的逻辑块地址；

获取单元，用于从所述第一逻辑块号对应的逻辑块地址中获取所述第一关键字对应的数据记录。

9.如权利要求6-8中任一项所述的装置，其特征在于，所述装置还包括：

更新单元，用于使用压缩后的所述叶子节点的逻辑块地址，更新所述叶子节点与所述叶子节点的逻辑块地址的对应关系。

10.如权利要求6-8中任一项所述的装置，其特征在于，

所述压缩单元，具体用于当所述索引树占用的存储空间大于所述预设的第一阈值而小于第二阈值时，压缩存储于所述第一确定单元确定的所述逻辑块地址中的叶子节点的数据记录；

所述装置还包括：

淘汰单元，用于当所述索引树占用的存储空间大于所述第二阈值时，根据所述叶子节点在所述缓存中的逻辑块地址，仅将所述叶子节点中的数据记录淘汰至磁盘中。

存储数据的方法和装置

技术领域

[0001] 本发明涉及数据存储领域，并且更为具体地，涉及存储数据的方法和装置。

背景技术

[0002] 在存储系统中，通常选用索引树作为磁盘数据的索引查询的数据结构。以B+Tree为例进行说明，当B+Tree中插入的节点较多时，由于系统的缓存资源有限，会将B+Tree的一部分节点淘汰至磁盘中。B+Tree中，被淘汰到磁盘的节点越多，主机访问时，磁盘IO次数也越多，相应地，主机的IOPS(每秒输入输出的次数，Input/Output Operations Per Second)也就越少。

[0003] 目前，高性能的存储系统对主机IOPS具有较高的要求，如何提高IOPS亟待解决。

发明内容

[0004] 本发明实施例提供一种存储数据的方法和装置，以提高主机的IOPS。

[0005] 第一方面，提供一种存储数据的方法，包括：在缓存中存储索引树，其中，所述索引树包括索引节点和叶子节点；并建立所述索引树的各节点的节点类型与所述各节点的逻辑块地址的对应关系；当所述索引树占用的存储空间大于预设的第一阈值时，根据所述对应关系，确定所述索引树的所述叶子节点在所述缓存中的逻辑块地址；压缩存储于所述逻辑块地址中的叶子节点的数据记录。

[0006] 结合第一方面，在第一方面的一种实现方式中，所述索引树的每个节点对应一个逻辑块号，所述索引树的索引节点存储用于索引查询的关键字和逻辑块号的对应关系，所述索引树的叶子节点存储所述关键字与数据记录的对应关系，所述建立所述索引树的各节点的节点类型与所述各节点的逻辑块地址的对应关系，包括：每当所述索引树插入一个新节点时，确定所述新节点的节点类型，并为所述新节点分配逻辑块号；在所述缓存中为所述新节点分配逻辑块地址，并建立所述新节点的节点类型、所述新节点的逻辑块号和所述新节点的逻辑块地址的对应关系。

[0007] 结合第一方面或其上述实现方式的任一种，在第一方面的另一种实现方式中，所述方法还包括：接收查询消息，所述查询消息携带第一关键字；根据所述第一关键字与所述索引节点中存储的关键字的大小关系，从所述索引树的根节点开始逐层查找，直到确定第一逻辑块号，所述第一逻辑块号对应的逻辑块地址中存储有所述第一关键字对应的数据记录；根据所述索引树中各节点的逻辑块号和所述各节点的逻辑块地址的对应关系，确定所述第一逻辑块号对应的逻辑块地址；从所述第一逻辑块号对应的逻辑块地址中获取所述第一关键字对应的数据记录。

[0008] 结合第一方面或其上述实现方式的任一种，在第一方面的另一种实现方式中，所述方法还包括：使用压缩后的所述叶子节点的逻辑块地址，更新所述叶子节点与所述叶子节点的逻辑块地址的对应关系。

[0009] 结合第一方面或其上述实现方式的任一种，在第一方面的另一种实现方式中，当

所述索引树占用的存储空间大于预设的所述第一阈值时，压缩存储于所述逻辑块地址中的叶子节点的数据记录，具体包括当所述索引树占用的存储空间大于预设的所述第一阈值而小于第二阈值时，压缩存储于所述逻辑块地址中的叶子节点的数据记录；

[0010] 所述方法还包括：当所述索引树占用的存储空间大于所述第二阈值时，根据所述叶子节点在所述缓存中的逻辑块地址，仅将所述叶子节点中的数据记录淘汰至磁盘中。

[0011] 第二方面，提供一种存储数据的装置，包括：缓存单元，用于在缓存中存储索引树，其中，所述索引树包括索引节点和叶子节点；建立单元，用于建立所述缓存单元存储的所述索引树的各节点的节点类型与所述各节点的逻辑块地址的对应关系；第一确定单元，用于当所述索引树占用的存储空间大于预设的第一阈值时，根据所述建立单元建立的所述对应关系，确定所述索引树的所述叶子节点在所述缓存中的逻辑块地址；压缩单元，用于压缩存储于所述第一确定单元确定的所述逻辑块地址中的叶子节点的数据记录。

[0012] 结合第二方面，在第二方面的一种实现方式中，所述索引树的每个节点对应一个逻辑块号，所述索引树的索引节点存储用于索引查询的关键字和逻辑块号的对应关系，所述索引树的叶子节点存储所述关键字与数据记录的对应关系，所述建立单元具体用于每当所述索引树插入一个新节点时，确定所述新节点的节点类型，并为所述新节点分配逻辑块号；在所述缓存中为所述新节点分配逻辑块地址，并建立所述新节点的节点类型、所述新节点的逻辑块号和所述新节点的逻辑块地址的对应关系。

[0013] 结合第二方面或其上述实现方式的任一种，在第二方面的另一种实现方式中，所述装置还包括：接收单元，用于接收查询消息，所述查询消息携带第一关键字；查询单元，用于根据所述第一关键字与所述索引节点中存储的关键字的大小关系，从所述索引树的根节点开始逐层查找，直到确定第一逻辑块号，所述第一逻辑块号对应的逻辑块地址中存储有所述第一关键字对应的数据记录；第二确定单元，用于根据所述索引树中各节点的逻辑块号和所述各节点的逻辑块地址的对应关系，确定所述第一逻辑块号对应的逻辑块地址；获取单元，用于从所述第一逻辑块号对应的逻辑块地址中获取所述第一关键字对应的数据记录。

[0014] 结合第二方面或其上述实现方式的任一种，在第二方面的另一种实现方式中，所述装置还包括：更新单元，用于使用压缩后的所述叶子节点的逻辑块地址，更新所述叶子节点与所述叶子节点的逻辑块地址的对应关系。

[0015] 结合第二方面或其上述实现方式的任一种，在第二方面的另一种实现方式中，所述压缩单元，具体用于当所述索引树占用的存储空间大于所述预设的第一阈值而小于第二阈值时，压缩存储于所述第一确定单元确定的所述逻辑块地址中的叶子节点的数据记录；

[0016] 所述装置还包括：淘汰单元，用于当所述索引树占用的存储空间大于所述第二阈值时，根据所述叶子节点在所述缓存中的逻辑块地址，仅将所述叶子节点中的数据记录淘汰至磁盘中。

[0017] 本发明实施例中，通过对缓存中的叶子节点的数据记录进行压缩，从而能够在缓存中存储更多的节点，总体上降低了磁盘IO的次数，提高主机的IOPS。

附图说明

[0018] 为了更清楚地说明本发明实施例的技术方案，下面将对本发明实施例中所需要使

用的附图作简单地介绍，显而易见地，下面所描述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。

[0019] 图1是本发明一个实施例的存储数据的方法的示意性流程图。

[0020] 图2是根据本发明实施例的存储系统的系统架构图。

[0021] 图3是本发明实施例的节点映射层的示意性结构图。

[0022] 图4是本发明一个实施例的存储数据的装置的示意性框图。

[0023] 图5是本发明一个实施例的存储数据的装置的示意性框图。

具体实施方式

[0024] 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，所描述的实施例是本发明的一部分实施例，而不是全部实施例。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动的前提下所获得的所有其他实施例，都应属于本发明保护的范围。

[0025] 图1是本发明一个实施例的存储数据的方法的示意性流程图。图1的方法包括：

[0026] 110、在缓存中存储索引树，其中，索引树包括索引节点和叶子节点；并建立索引树的各节点的节点类型与各节点的逻辑块地址的对应关系。

[0027] 应理解，索引树可以是B-Tree、B+Tree、二叉树、红黑树等。

[0028] 应理解，某个节点的逻辑块地址具体可指该节点在缓存中的存储位置。

[0029] 应理解，本发明实施例对上述对应关系建立的具体方式不作限定。可选地，作为一个实施例，可以通过遍历索引树各节点的方式，确定索引树中每个节点的节点类型及其在缓存中的逻辑块地址，并建立上述对应关系。上述对应关系还可采用其他方式建立，详见后续描述。

[0030] 120、当索引树占用的存储空间大于预设的第一阈值时，根据上述对应关系，确定索引树的叶子节点在缓存中的逻辑块地址。

[0031] 举例说明，步骤110建立的对应关系为一张对应关系表，该对应关系表中记录了索引节点的逻辑块地址和叶子节点的逻辑块地址，可以通过遍历该对应关系表找到索引树的叶子节点的逻辑块地址。

[0032] 应理解，本发明实施例对上述第一阈值的大小不作具体限定，可以根据整个缓存空间大小确定。例如，可以设置为整个缓存空间的60％。

[0033] 应理解，上述索引树占用的缓存空间的大小可以通过多种方式确定。例如，可以在缓存中设置一个变量，该变量用于统计索引树占用缓存空间。当索引树在缓存中增加一个新节点时，该变量的值随之变大；当索引树删去位于缓存中的一个节点时，该变量的值随之减小。通过查看该变量的值即可确定索引树占用缓存空间的大小。

[0034] 130、压缩存储于逻辑块地址中的叶子节点的数据记录。

[0035] 具体地，当索引树占用的缓存空间达到第一阈值时，可以将叶子节点中的数据记录进行压缩，使其小于第一阈值；当索引树占用的缓存空间再次达到第一阈值时，可以对未压缩的叶子节点再次进行压缩直到缓存中全部叶子节点均被压缩为止。或者，当索引树占用的缓存空间达到第一阈值时，可以将叶子节点中的数据记录进行压缩；当索引树占用的

缓存空间再次达到第一阈值时，不再对叶子节点进行压缩。

[0036] 本发明实施例中，通过对缓存中的叶子节点的数据记录进行压缩，从而能够在缓存中存储更多的节点，总体上降低了磁盘IO的次数，提高主机的IOPS。

[0037] 可选地，作为一个实施例，图1的方法还可包括：使用压缩后的叶子节点的逻辑块地址，更新叶子节点与叶子节点的逻辑块地址的对应关系。

[0038] 也就是说，叶子节点的数据记录被压缩后，该叶子节点占用的逻辑块地址也就相应地改变，用该改变后的逻辑块地址替换之前存储的该叶子节点对应的逻辑块地址，以保证后续使用时该对应关系的正确性。

[0039] 可选地，作为另一个实施例，索引树的每个节点对应一个逻辑块号，索引树的索引节点存储用于索引查询的关键字和逻辑块号的对应关系，索引树的叶子节点存储关键字与数据记录的对应关系，上述建立索引树的各节点的节点类型与各节点的逻辑块地址的对应关系，可包括：每当索引树插入一个新节点时，确定新节点的节点类型，并为新节点分配逻辑块号；在缓存中为新节点分配逻辑块地址，并建立新节点的节点类型、新节点的逻辑块号和新节点的逻辑块地址的对应关系。

[0040] 现有技术中，索引树的索引节点存储的是关键字和直接或间接地指向叶子节点存储位置的地址指针的对应关系。这样，当叶子节点中的数据记录被压缩时，叶子节点的逻辑块地址会发生改变，要想再次准确地索引到叶子节点，不但需要更新叶子节点与其逻辑块地址的对应关系，还需要更新索引节点的存储内容，操作起来非常复杂，会大大消耗计算资源。在上述实现方式中，通过为每个节点分配一个逻辑块号，在索引节点中存储关键字和逻辑块号的对应关系，并建立各节点的节点类型、逻辑块号和逻辑块地址的对应关系，当叶子节点的逻辑块地址发生变化时，无需对索引节点的内容进行调整，仅需更新叶子节点与其逻辑块地址的对应关系即可，从而降低了后续更新操作复杂性，能够节省计算资源。

[0041] 由于索引树的索引节点存储的内容发生变化，在该实现方式的基础上，基于索引树的插入、查询和删除等操作均需相应地做出调整，以索引树的查询为例，具体可包括如下步骤：接收查询消息，查询消息携带第一关键字；根据第一关键字与索引节点中存储的关键字的大小关系，从索引树的根节点开始逐层查找，直到确定第一逻辑块号，第一逻辑块号对应的逻辑块地址中存储有第一关键字对应的数据记录；根据索引树中各节点的逻辑块号和各节点的逻辑块地址的对应关系，确定第一逻辑块号对应的逻辑块地址；从第一逻辑块号对应的逻辑块地址中获取第一关键字对应的数据记录。

[0042] 目前，高性能的存储系统对主机IO的访问时延提出了明确的要求，即要求不同主机IO的访问时延稳定在一定范围内。但是，在现有技术中，索引树(如B+Tree)的每条路径(从根节点到叶子节点的路径)上缓存的节点数目不同，导致沿不同路径索引查询时，所需的磁盘IO次数不同，从而导致不同主机IO的访问时延不稳定。

[0043] 可选地，作为一个实施例，当所述索引树占用的存储空间大于预设的所述第一阈值时，压缩存储于所述逻辑块地址中的叶子节点的数据记录，具体包括当所述索引树占用的存储空间大于预设的所述第一阈值而小于第二阈值时，压缩存储于所述逻辑块地址中的叶子节点的数据记录；图1的方法还可包括：当索引树占用的存储空间大于第二阈值时，根据叶子节点在缓存中的逻辑块地址，仅将叶子节点中的数据记录淘汰至磁盘中。也就是说，将索引节点保留在缓存中，而将叶子节点的数据记录下盘。

[0044] 当索引树占用的缓存空间达到第二阈值时，通过识别索引树的叶子节点，并将索引树的叶子节点的数据记录下盘，而将索引节点均保留在缓存中，使得访问索引树的每条路径时最多只有一次磁盘IO，从而稳定主机IO的访问时延。

[0045] 可选地，作为一个实施例，上述将叶子节点的数据记录淘汰至磁盘中，可包括：从叶子节点的存储位置中获取叶子节点的数据记录；将该数据记录进行压缩；将压缩后的数据记录淘汰至磁盘中。通过将准备下盘的数据记录进行压缩，能够将多个节点的数据记录压缩到一个IO空间的范围，从而总体上减少了利用索引树访问磁盘时，对磁盘的IO次数。

[0046] 应理解，无论是对叶子节点的数据记录进行压缩，还是将叶子节点的数据记录下盘，均会改变节点的存储位置，此时可以通过更新上述索引树各节点与其逻辑块地址的对应关系来维持该对应关系的准确性，以便后续对该对应关系的利用。

[0047] 下面结合具体例子，更加详细地描述本发明实施例。应注意，以下例子仅仅是为了帮助本领域技术人员理解本发明实施例，而非要将本发明实施例限于所例示的具体数值或具体场景。本领域技术人员根据以下例子，显然可以进行各种等价的修改或变化，这样的修改或变化也落入本发明实施例的范围内。

[0048] 现有技术中，主机IO与磁盘之间通过B+Tree直接建立映射关系。具体地，B+Tree的索引节点存储的是关键字(如逻辑单元号(Logical Unit Number，LUN)地址)和地址指针的对应关系，在B+Tree中进行关键字搜索时，通过索引节点中的地址指针的指引即可找到该关键字对应的数据记录在缓存或磁盘中的位置。本发明实施例为了提高主机的IOPS，需要对B+Tree的叶子节点进行压缩。在这种情况下，如果仍沿用在B+Tree的索引节点中存储关键字和地址指针的对应关系，当叶子节点被压缩后，就需要对整个索引节点中的地址指针进行调整，操作复杂，调整时间长。

[0049] 图2是根据本发明实施例的存储系统的系统架构图。在图2中，索引树210(以B+Tree 210为例)与磁盘230之间加入节点映射层(NML，Node Mapping Layer)220。B+Tree中的索引节点存储关键字和逻辑块号的对应关系，B+Tree中的叶子节点存储关键字和数据记录的对应关系，节点映射层220存储节点类型、逻辑块号和逻辑块地址的对应关系。当叶子节点被压缩后，由于索引节点存储的是关键字和逻辑块号的对应关系，无需作出调整，仅需要更新压缩后的叶子节点的逻辑块号与逻辑块地址的对应关系，从而节省了存储系统的计算资源，降低了操作的复杂度。下面对节点映射层220的结构、功能以及节点映射层220与B+Tree的交互进行详细描述。

[0050] 如图3所示，节点映射层220可包含三个模块：分类的压缩和缓存策略模块，数据压缩模块以及节点缓存模块。

[0051] (1)、分类的压缩和缓存策略模块

[0052] 每当B+Tree的某个节点释放逻辑块的引用时，节点映射层220的分类的压缩和缓存策略模块可以根据该节点是索引节点还是叶子节点，采用不同的压缩和缓存策略。分类的压缩和缓存策略模块的压缩缓存策略可以如下：

[0053] 如果缓存空间足够，索引节点和叶子节点都可采用不压缩的方式保存在缓存中。

[0054] 如果缓存空间达到了需要压缩叶子节点的水位(对应于上述第一阈值)，节点映射层220会启动对叶子节点数据记录的压缩，这时压缩后的叶子节点将占用更少的缓存空间，同时释放出来的缓存空间可以给其他节点继续使用，这样可以存储更多的节点。

[0055] 如果缓存空间达到了需要淘汰节点下盘的水位(对应于上述第二阈值),则节点映射层220优先将叶子节点的数据记录下盘,当缓存空间足够保存全部索引节点时,索引节点将全部被保存在缓存中。采用这样的下盘方式,B+tree的每条访问路径最多只有一次磁盘IO,从而将主机IO沿各访问路径的时延稳定在一定范围内。

[0056] (2)、数据压缩模块

[0057] 数据压缩模块,主要负责对节点的数据记录进行压缩。压缩后的数据记录包含压缩头部信息和数据信息。压缩头部信息用于记录压缩算法类型和压缩数据的长度。

[0058] (3)、节点缓存模块

[0059] 节点缓存模块,主要负责管理节点的缓存空间,负责节点缓存的分配和释放,节点缓存的淘汰刷盘。

[0060] 下面结合表一,具体说明B+Tree在完成插入、查询、删除等操作时与节点映射层220直接的交互过程。

[0061] 表一:节点映射层与B+tree之间接口的定义

[0062]

接口名称	接口定义	参数描述
AllocBlock	AllocBlock(node_type)	在节点映射层220分配一个逻辑块号,分配时需要指定节点类型是索引节点还是叶子节点,分配成功会返回一个逻辑块号给B+Tree。

[0063]

GetBlock	GetBlock(block_id)	B+Tree需要访问某个节点时，通过这个接口获取某个逻辑块号对应的逻辑块地址中的内容。
PutBlock	PutBlock(block_id)	B+Tree对某个节点访问完成后，通过这个接口释放对该节点对应的逻辑块的引用。
FreeBlock	FreeBlock(block_id)	B+Tree删除某个节点时，调用该接口释放这个逻辑块号。节点映射层220也将释放这个块相应的存储空间。

[0064] 一、对B+Tree的插入操作

[0065] 1、对B+Tree执行插入操作时，如果需要新分配一个节点，B+Tree调用AllocBlock接口分配一个逻辑块号。分配时需要指定节点类型是叶子节点还是索引节点。

[0066] 2、分配完成后，需要对这个块进行插入记录的操作，调用GetBlock操作获取这个逻辑块的内容，因为是新分配的节点，节点映射层220将直接在缓存空间中分配逻辑块地址，并返回给B+Tree，然后将这个缓存的索引和逻辑块号关联起来。

[0067] 3、B+Tree插入完成后就会调用PutBlock接口释放对这个逻辑块的引用，如果缓存空间充足，没有达到预设的水位，节点映射层220将该逻辑块对应的缓存空间留在缓存中，下层访问时可以直接从缓存中获取。如果缓存空间达到了预设的启动压缩的水位(对应于上述第一阈值)，则优先开始启动对叶子节点数据记录的压缩，并将压缩后的数据记录存储在缓存中。此时，压缩后释放的缓存空间就可以继续给其他逻辑块使用。如果缓存空间达到了需要淘汰下盘的水位(对应于上述第二阈值)，节点映射层220将启动叶子节点数据记录的下盘。这时，节点映射层220将根据节点类型把叶子节点的数据记录优先淘汰下盘。

[0068] 4、如果B+Tree的插入操作不需要新分配节点，则根据获取的节点的逻辑块号，调用GetBlock接口获取该逻辑块对应的数据记录。获取之后直接对该数据记录进行修改，修改完成后调用PutBlock释放对这个块的引用。释放后节点映射层220的操作和3中描述一致，此处不再详述。在进行GetBlock操作时，节点映射层220会根据当前逻辑块的状态标识(标识该逻辑块是否在缓存中)，如果节点位于缓存中，则直接获取，如果不在缓存中，则从磁盘中获取对应块的内容，并返回给B+Tree。

[0069] 二、对B+Tree的查询操作

[0070] 1、对B+Tree的查询操做主要是B+Tree节点根据索引节点存储的关键字和逻辑块

号的对应关系，调用节点映射层220的GetBlock操作的过程。GetBlock的操作流程和上述对B+Tree的插入操作中的1-4描述的方式类似，此处不再赘述。

[0071] 2、在调用GetBlock访问完对应的节点后，就调用PutBlock接口释放对这个块的引用。PutBlock的处理和上述对B+Tree的插入操作中3描述的方式类似，此处不再赘述。

[0072] 三、对B+Tree的删除操作

[0073] 1、对B+Tree节点进行删除操作时，需要调用到节点映射层220的FreeBlock接口。

[0074] 2、调用FreeBlock接口时，节点映射层220会释放这个块关联的磁盘空间和缓存节点空间，并将逻辑块号回收。

[0075] 上文中结合图1至图3，详细描述了根据本发明实施例的存储数据的方法，下面将结合图4至图5，详细描述根据本发明实施例的存储数据的装置。

[0076] 图4是本发明一个实施例的存储数据的装置的示意性框图。应理解，图4的装置400能够完成图1至图3中的各个步骤，为避免重复，此处不再详述。装置400包括：

[0077] 缓存单元410，用于在缓存中存储索引树，其中，索引树包括索引节点和叶子节点；

[0078] 建立单元420，用于建立缓存单元410存储的索引树的各节点的节点类型与各节点的逻辑块地址的对应关系；

[0079] 第一确定单元430，用于当索引树占用的存储空间大于预设的第一阈值时，根据建立单元420建立的对应关系，确定索引树的叶子节点在缓存中的逻辑块地址；

[0080] 压缩单元440，用于压缩存储于第一确定单元430确定的逻辑块地址中的叶子节点的数据记录。

[0081] 本发明实施例中，通过对缓存中的叶子节点的数据记录进行压缩，从而能够在缓存中存储更多的节点，总体上降低了磁盘IO的次数，提高主机的IOPS。

[0082] 可选地，作为一个实施例，索引树的每个节点对应一个逻辑块号，索引树的索引节点存储用于索引查询的关键字和逻辑块号的对应关系，索引树的叶子节点存储关键字与数据记录的对应关系，建立单元420具体用于每当索引树插入一个新节点时，确定新节点的节点类型，并为新节点分配逻辑块号；在缓存中为新节点分配逻辑块地址，并建立新节点的节点类型、新节点的逻辑块号和新节点的逻辑块地址的对应关系。

[0083] 可选地，作为一个实施例，装置400还可包括：接收单元，用于接收查询消息，查询消息携带第一关键字；查询单元，用于根据第一关键字与索引节点中存储的关键字的大小关系，从索引树的根节点开始逐层查找，直到确定第一逻辑块号，第一逻辑块号对应的逻辑块地址中存储有第一关键字对应的数据记录；第二确定单元，用于根据索引树中各节点的逻辑块号和各节点的逻辑块地址的对应关系，确定第一逻辑块号对应的逻辑块地址；获取单元，用于从第一逻辑块号对应的逻辑块地址中获取第一关键字对应的数据记录。

[0084] 可选地，作为一个实施例，装置400还可包括：更新单元，用于使用压缩后的叶子节点的逻辑块地址，更新叶子节点与叶子节点的逻辑块地址的对应关系。

[0085] 可选地，作为一个实施例，所述压缩单元，具体用于当所述索引树占用的存储空间大于所述预设的第一阈值而小于第二阈值时，压缩存储于所述第一确定单元确定的所述逻辑块地址中的叶子节点的数据记录；

[0086] 装置400还可包括：淘汰单元，用于当索引树占用的存储空间大于第二阈值时，根据叶子节点在缓存中的逻辑块地址，仅将叶子节点中的数据记录淘汰至磁盘中。

[0087] 图5是本发明一个实施例的存储数据的装置的示意性框图。应理解，图5的装置500能够完成图1至图3中的各个步骤，为避免重复，此处不再详述。装置500包括：

[0088] 存储器510，用于在缓存中存储索引树，其中，索引树包括索引节点和叶子节点；

[0089] 处理器520，用于建立存储器510存储的索引树的各节点的节点类型与各节点的逻辑块地址的对应关系；当索引树占用的存储空间大于预设的第一阈值时，根据该对应关系，确定索引树的叶子节点在缓存中的逻辑块地址；压缩存储于该逻辑块地址中的叶子节点的数据记录。

[0090] 本发明实施例中，通过对缓存中的叶子节点的数据记录进行压缩，从而能够在缓存中存储更多的节点，总体上降低了磁盘IO的次数，提高主机的IOPS。

[0091] 可选地，作为一个实施例，索引树的每个节点对应一个逻辑块号，索引树的索引节点存储用于索引查询的关键字和逻辑块号的对应关系，索引树的叶子节点存储关键字与数据记录的对应关系，处理器520具体用于每当索引树插入一个新节点时，确定新节点的节点类型，并为新节点分配逻辑块号；在缓存中为新节点分配逻辑块地址，并建立新节点的节点类型、新节点的逻辑块号和新节点的逻辑块地址的对应关系。

[0092] 可选地，作为一个实施例，处理器520还用于接收查询消息，查询消息携带第一关键字；根据第一关键字与索引节点中存储的关键字的大小关系，从索引树的根节点开始逐层查找，直到确定第一逻辑块号，第一逻辑块号对应的逻辑块地址中存储有第一关键字对应的数据记录；根据索引树中各节点的逻辑块号和各节点的逻辑块地址的对应关系，确定第一逻辑块号对应的逻辑块地址；获取单元，用于从第一逻辑块号对应的逻辑块地址中获取第一关键字对应的数据记录。

[0093] 可选地，作为一个实施例，处理器520还用于使用压缩后的叶子节点的逻辑块地址，更新叶子节点与叶子节点的逻辑块地址的对应关系。

[0094] 可选地，作为一个实施例，当所述索引树占用的存储空间大于预设的所述第一阈值时，处理器520用于压缩存储于所述逻辑块地址中的叶子节点的数据记录，具体包括当所述索引树占用的存储空间大于预设的所述第一阈值而小于第二阈值时，压缩存储于所述逻辑块地址中的叶子节点的数据记录；

[0095] 处理器520还用于当索引树占用的存储空间大于第二阈值时，根据叶子节点在缓存中的逻辑块地址，仅将叶子节点中的数据记录淘汰至磁盘中。

[0096] 本领域普通技术人员可以意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，能够以电子硬件、或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行，取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出本发明的范围。

[0097] 所属领域的技术人员可以清楚地了解到，为描述的方便和简洁，上述描述的系统、装置和单元的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

[0098] 在本申请所提供的几个实施例中，应该理解到，所揭露的系统、装置和方法，可以通过其它的方式实现。例如，以上所描述的装置实施例仅仅是示意性的，例如，所述单元的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个单元或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另一点，所显示或

讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口，装置或单元的间接耦合或通信连接，可以是电性，机械或其它的形式。

[0099] 所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

[0100] 另外，在本发明各个实施例中的各功能单元可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个单元中。

[0101] 所述功能如果以软件功能单元的形式实现并作为独立的产品销售或使用时，可以存储在一个计算机可读取存储介质中。基于这样的理解，本发明的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的部分可以以软件产品的形式体现出来，该计算机软件产品存储在一个存储介质中，包括若干指令用以使得一台计算机设备(可以是个人计算机，服务器，或者网络设备等)执行本发明各个实施例所述方法的全部或部分步骤。而前述的存储介质包括：U盘、移动硬盘、只读存储器(ROM，Read-Only Memory)、随机存取存储器(RAM，Random Access Memory)、磁碟或者光盘等各种可以存储程序代码的介质。

[0102] 以上所述，仅为本发明的具体实施方式，但本发明的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本发明揭露的技术范围内，可轻易想到变化或替换，都应涵盖在本发明的保护范围之内。因此，本发明的保护范围应所述以权利要求的保护范围为准。

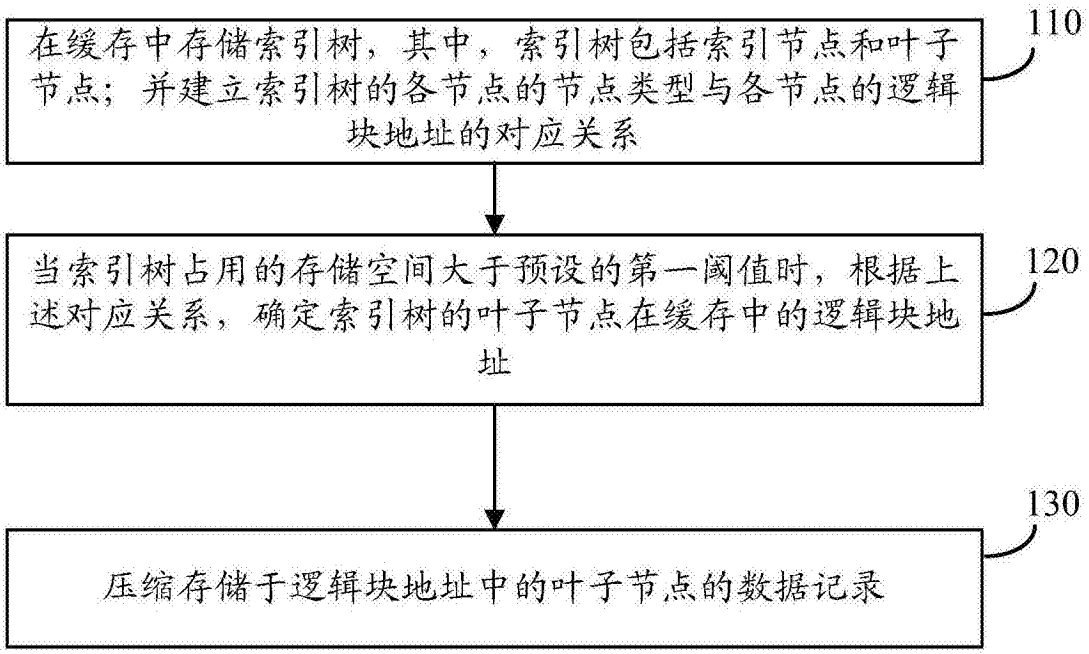

在缓存中存储索引树，其中，索引树包括索引节点和叶子节点；并建立索引树的各节点的节点类型与各节点的逻辑块地址的对应关系
110
当索引树占用的存储空间大于预设的第一阈值时，根据上述对应关系，确定索引树的叶子节点在缓存中的逻辑块地址
120
压缩存储于逻辑块地址中的叶子节点的数据记录
130

图1

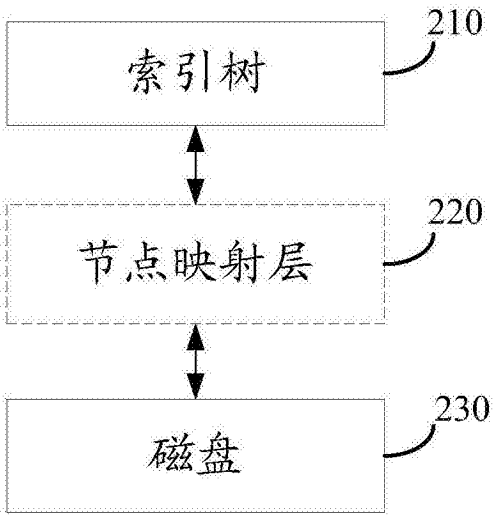

索引树
210
节点映射层
220
磁盘
230

图2

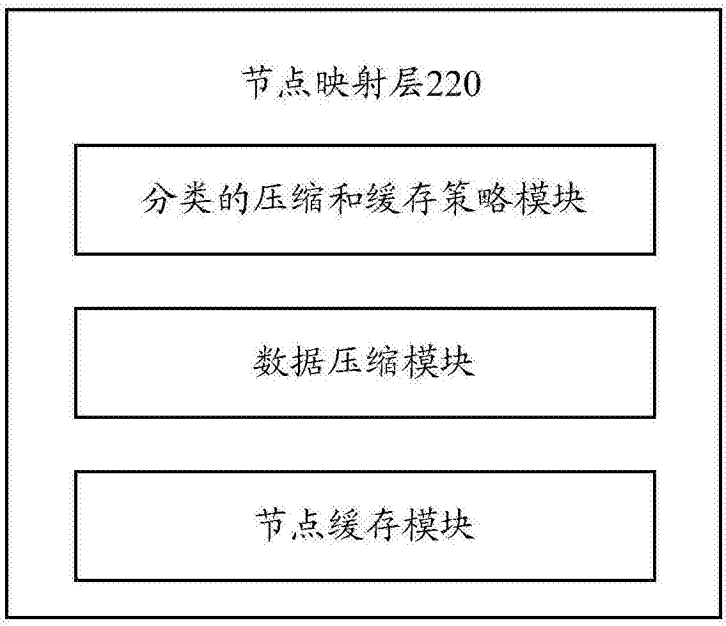

节点映射层220
分类的压缩和缓存策略模块
数据压缩模块
节点缓存模块

图3

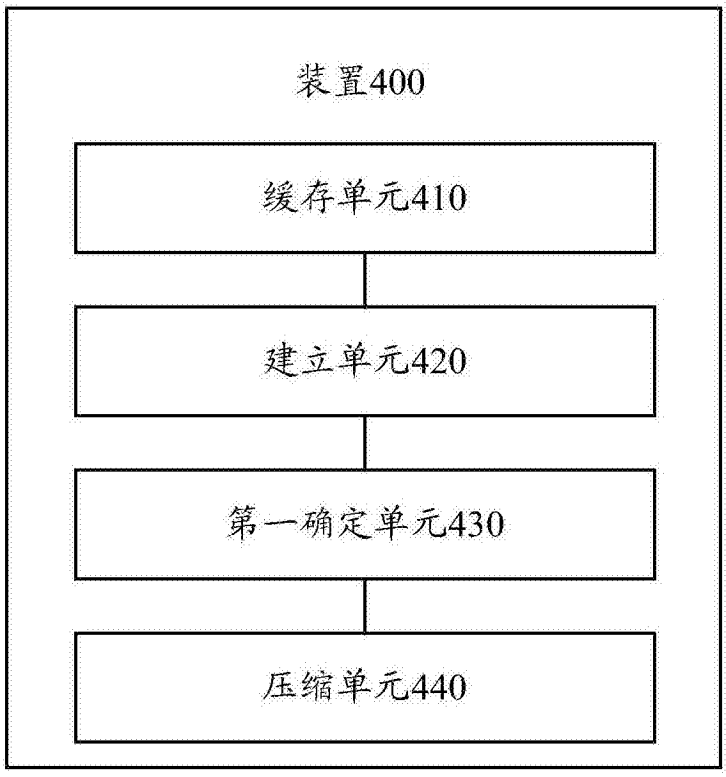

装置400
缓存单元410
建立单元420
第一确定单元430
压缩单元440

图4

图5