(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国 际 局

(43) 国际公布日
2021 年 8 月 19 日 (19.08.2021)

WIPO | PCT

(10) 国际公布号
WO 2021/159711 A1

(51) 国际专利分类号:
G06F 3/06 (2006.01) *G06F 16/13* (2019.01)

(21) 国际申请号: PCT/CN2020/117331

(22) 国际申请日: 2020 年 9 月 24 日 (24.09.2020)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202010093601.7 2020年2月14日 (14.02.2020) CN

(71) 申请人: 苏州浪潮智能科技有限公司 (SUZHOU INSPUR INTELLIGENT TECHNOLOGY CO., LTD) [CN/CN]; 中国江苏省苏州市吴中区郭巷街道官浦路1号9幢, Jiangsu 215000 (CN)。

(72) 发明人: 来炜国 (LAI, Weiguo); 中国江苏省苏州市吴中区郭巷街道官浦路1号9幢, Jiangsu 215000 (CN)。 刘志勇 (LIU, Zhiyong); 中国江苏省苏州市吴中区郭巷街道官浦路1号9幢, Jiangsu 215000 (CN)。

(74) 代理人: 北京集佳知识产权代理有限公司 (UNITALEN ATTORNEYS AT LAW); 中国北京市朝阳区建国门外大街22号赛特广场7层, Beijing 100004 (CN)。

(81) 指定国 (除另有指明，要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国 (除另有指明，要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU,

(54) **Title:** B+ TREE ACCESS METHOD AND APPARATUS, AND COMPUTER-READABLE STORAGE MEDIUM

(54) 发明名称: 一种B+树的存取方法、装置和计算机可读存储介质

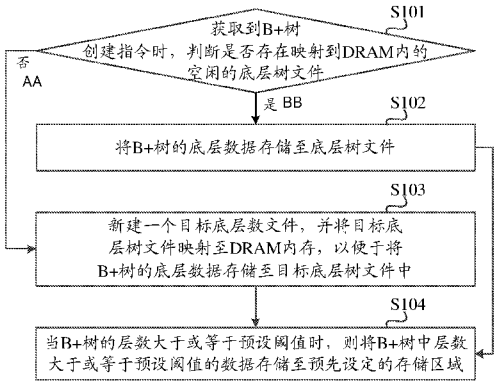

图 1

S101 When a B+ tree creation instruction is acquired, determine whether there is an idle underlying tree file mapped to a DRAM
S102 Store underlying data of a B+ tree in the underlying tree file
S103 Create a new target underlying tree file, and map the target underlying tree file to the DRAM, so as to store the underlying data of the B+ tree in the target underlying tree file
S104 When the number of layers of the B+ tree is greater than or equal to a preset threshold, store data, the number of layers of which, in the B+ tree, is greater than or equal to the preset threshold, in a preset storage area
AA No
BB Yes

(57) **Abstract:** Provided are a B+ tree access method and apparatus, and a medium. The method comprises: when a B+ tree creation instruction is acquired, determining whether there is an idle underlying tree file mapped to a DRAM; if so, storing underlying data of a B+ tree in the underlying tree file; if not, creating a new target underlying tree file, and mapping the target underlying tree file to the DRAM, so as to store the underlying data of the B+ tree in the target underlying tree file; and when the number of layers of the B+ tree is greater than or equal to a preset threshold, storing data, the number of layers of which, in the B+ tree, is greater than or equal to the preset threshold, in a preset storage area. On the basis of the data structure of the B+ tree, for reading data each time, the underlying data needs to be accessed. The access efficiency of the underlying data is effectively improved by means of storing underlying files in the idle underlying tree file mapped to the DRAM; and the utilization rate of DRAM resources is increased by means of storing other data in a storage space other than the DRAM.

WO 2021/159711 A1

[见续页]

IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

(57) 摘要: 一种B+树的存取方法、装置和介质，获取到B+树创建指令时，判断是否存在映射到DRAM内存的空闲的底层树文件；若是，则将B+树的底层数据存储至底层树文件。若否，则新建一个目标底层数文件，并将目标底层树文件映射至DRAM内存，以便于将B+树的底层数据存储至目标底层树文件中。当B+树的层数大于或等于预设阈值时，则将B+树中层数大于或等于预设阈值的数据存储至预先设定的存储区域。基于B+树的数据结构，每次数据的读取都需要从底层数据访问，通过将底层文件存储在映射到DRAM内存的空闲的底层树文件，有效的提升了底层数据的访问效率。将其它数据存储在除DRAM内存外的存储空间中，提升了DRAM内存资源的利用率。

一种 B+树的存取方法、装置和计算机可读存储介质

本申请要求于 2020 年 02 月 14 日提交中国专利局、申请号为 202010093601.7、发明名称为“一种 B+树的存取方法、装置和计算机可读存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本发明涉及数据存储技术领域，特别是涉及一种 B+树的存取方法、装置和计算机可读存储介质。

背景技术

在全闪存阵列存储系统中，由于其结构的特殊性，大量采用 B+树结构。例如，全闪存阵列普遍采用自动精简的容量分配方式（thin provisioning），其卷的逻辑地址与卷在磁盘阵列(Redundant Arrays of Independent Disks，RAID)上的物理地址不再是线性对应关系，而变成了近似随机映射的关系。为了管理这种映射关系，采用 B+树来保存从卷逻辑地址到物理地址的映射，并且采用 B+树来保存物理地址到逻辑地址的逆映射。全闪存阵列的重删功能，采用 B+树来保存数据块 HASH 值到物理地址的映射。

动态随机存取存储器(Dynamic Random Access Memory，DRAM)是一种较为常见的系统内存。由于 DRAM 内存是较为昂贵的部件，因此存储系统中通常只配置较少数量的 DRAM 内存以降低成本。现有技术中，B+树的数据通常保存在固态硬盘(Solid State Drives，SSD)中。SSD 盘中的数据必须读到内存中才能进行读写访问，而 SSD 读写的 IO 路径比较长，速度慢。B+树结构的频繁的读入及换出内存，会带来很大的 CPU 开销。

可见，如何提升 B+树的读写效率，是本领域技术人员需要解决的问题。

发明内容

本发明实施例的目的是提供一种 B+树的存取方法、装置和计算机可读

存储介质，可以提升 B+树的读写效率。

为解决上述技术问题，本发明实施例提供一种 B+树的存取方法，包括：

获取到 B+树创建指令时，判断是否存在映射到 DRAM 内存的空闲的底层树文件；

若是，则将所述 B+树的底层数据存储至所述底层树文件；

若否，则新建一个目标底层数文件，并将所述目标底层树文件映射至 DRAM 内存，以便于将所述 B+树的底层数据存储至所述目标底层树文件中；

当所述 B+树的层数大于或等于预设阈值时，则将所述 B+树中层数大于或等于所述预设阈值的数据存储至预先设定的存储区域。

可选地，所述 B+树对应有第一层级、第二层级和第三层级；其中，第一层级的数据作为底层数据；

相应的，所述当所述 B+树的层数大于或等于预设阈值时，则将所述 B+树中大于或等于所述预设阈值的数据存储至预先设定的存储区域包括：

将所述 B+树的第二层级的数据存储至 DCPMM 内存中；

将所述 B+树的第三层级的数据存储至预设的硬盘中。

可选地，在所述获取到 B+树创建指令时，判断是否存在映射到内存的底层树文件之前还包括：

将所述 DCPMM 内存的最小读写粒度作为所述 B+树的节点容量。

可选地，所述将所述 B+树的底层数据存储至所述底层树文件包括：

将所述 B+树的底层数据按照所述节点容量向所述底层树文件中存储各节点数据；其中，每个节点数据的键值对中存储有下一层级节点的偏移地址。

可选地，还包括：

当获取到数据查询指令时，依据所述数据查询指令中携带的逻辑地址，确定出根节点；

根据所述根节点中包含的偏移地址，确定出叶子节点；并读取所述叶子节点对应的数据。

可选地，还包括：

当获取到数据修改指令时，判断所述 DRAM 内存中是否存在与所述数据修改指令中携带的节点标识相匹配的节点数据；

若是，则依据所述数据修改指令中携带的待替换数据，对所述节点数据进行修改，并对修改后的节点数据设置脏标志；

若否，则判断所述 DCPMM 内存中是否存在与所述数据修改指令中携带的节点标识相匹配的节点数据；

当所述 DCPMM 内存中存在与所述数据修改指令中携带的节点标识相匹配的节点数据时，则依据所述数据修改指令中携带的待替换数据，对所述节点数据进行修改；

当所述 DCPMM 内存中不存在与所述数据修改指令中携带的节点标识相匹配的节点数据时，则将所述硬盘中与所述数据修改指令中携带的节点标识相匹配的节点数据读取至所述 DRAM 内存中；依据所述数据修改指令中携带的待替换数据，在所述 DRAM 内存中完成对所述节点数据的修改，并对修改后的节点数据设置脏标志。

可选地，在所述将所述 B+树的底层数据存储至所述底层树文件之后还包括：

按照预设的周期时间将映射至 DRAM 内存中设置有脏标志的数据迁移至所述 DCPMM 内存中。

本发明实施例还提供了一种 B+树的存取装置，包括第一判断单元、第一存储单元、创建单元和第二存储单元；

所述第一判断单元，用于获取到 B+树创建指令时，判断是否存在映射到 DRAM 内存的空闲的底层树文件；若是，则触发所述第一存储单元；若否，则触发所述创建单元；

所述第一存储单元，用于将所述 B+树的底层数据存储至所述底层树文件；

所述创建单元，用于新建一个目标底层数文件，并将所述目标底层树文件映射至 DRAM 内存，以便于将所述 B+树的底层数据存储至所述目标底层树文件中；

所述第二存储单元，用于当所述 B+树的层数大于或等于预设阈值时，

则将所述 B+树中层数大于或等于所述预设阈值的数据存储至预先设定的存储区域。

可选地，所述 B+树对应有第一层级、第二层级和第三层级；其中，第一层级的数据作为底层数据；

相应的，所述第二存储单元具体用于将所述 B+树的第二层级的数据存储至 DCPMM 内存中；将所述 B+树的第三层级的数据存储至预设的硬盘中。

可选地，还包括作为单元；

所述作为单元，用于将所述 DCPMM 内存的最小读写粒度作为所述 B+树的节点容量。

可选地，所述第一存储单元具体用于将所述 B+树的底层数据按照所述节点容量向所述底层树文件中存储各节点数据；其中，每个节点数据的键值对中存储有下一层级节点的偏移地址。

可选地，还包括查询单元、确定单元和读取单元；

所述查询单元，用于当获取到数据查询指令时，依据所述数据查询指令中携带的逻辑地址，确定出根节点；

所述确定单元，用于根据所述根节点中包含的偏移地址，确定出叶子节点；

所述读取单元，用于读取所述叶子节点对应的数据。

可选地，还包括第二判断单元、修改单元、设置单元、第三判断单元和读取单元；

所述第二判断单元，用于当获取到数据修改指令时，判断所述 DRAM 内存中是否存在与所述数据修改指令中携带的节点标识相匹配的节点数据；若是，则触发所述修改单元；若否，则触发所述第三判断单元；

所述修改单元，用于依据所述数据修改指令中携带的待替换数据，对所述节点数据进行修改，

所述设置单元，用于对修改后的节点数据设置脏标志；

所述第三判断单元判断所述 DCPMM 内存中是否存在与所述数据修改指令中携带的节点标识相匹配的节点数据；

所述修改单元还用于当所述DCPMM内存中存在与所述数据修改指令中携带的节点标识相匹配的节点数据时，则依据所述数据修改指令中携带的待替换数据，对所述节点数据进行修改；

所述读取单元，用于当所述DCPMM内存中不存在与所述数据修改指令中携带的节点标识相匹配的节点数据时，则将所述硬盘中与所述数据修改指令中携带的节点标识相匹配的节点数据读取至所述DRAM内存中；

所述修改单元还用于依据所述数据修改指令中携带的待替换数据，在所述DRAM内存中完成对所述节点数据的修改，并触发所述设置单元对修改后的节点数据设置脏标志。

可选地，还包括迁移单元；

所述迁移单元，用于按照预设的周期时间将映射至DRAM内存中设置有脏标志的数据转移至所述DCPMM内存中。

本发明实施例还提供了一种B+树的存取装置，包括：

存储器，用于存储计算机程序；

处理器，用于执行所述计算机程序以实现如上述任意一项所述B+树的存取方法的步骤。

本发明实施例还提供了一种计算机可读存储介质，所述计算机可读存储介质上存储有计算机程序，所述计算机程序被处理器执行时实现如上述任意一项所述B+树的存取方法的步骤。

由上述技术方案可以看出，获取到B+树创建指令时，判断是否存在映射到DRAM内存的空闲的底层树文件；当存在映射到DRAM内存的空闲的底层树文件时，则可以直接将B+树的底层数据存储至底层树文件。当不存在映射到DRAM内存的空闲的底层树文件时，则需要新建一个目标底层数文件，并将目标底层树文件映射至DRAM内存，以便于将B+树的底层数据存储至目标底层树文件中。当B+树的层数大于或等于预设阈值时，则将B+树中层数大于或等于预设阈值的数据存储至预先设定的存储区域。基于B+树的数据结构，每次数据的读取都需要从底层数据访问，因此底层数据被访问的次数较多，通过将底层文件存储在映射到DRAM内存的空闲的底层树文件，有效的提升了底层数据的访问效率。而B+树中除底层数据外

的其它数据被访问的次数相对较少，为了降低对 DRAM 内存的占用，可以将其它数据存储在除 DRAM 内存外的存储空间中，既提升了 DRAM 内存资源的利用率，又可以保证 B+树的读写效率。

附图说明

为了更清楚地说明本发明实施例，下面将对实施例中所需要使用的附图做简单的介绍，显而易见地，下面描述中的附图仅仅是本发明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图 1 为本发明实施例提供的一种 B+树的存取方法的流程图；

图 2 为本发明实施例提供的一种 B+树的数据修改方法的流程图；

图 3 为本发明实施例提供的一种 B+树的存取装置的结构示意图；

图 4 为本发明实施例提供的一种 B+树的存取装置的硬件结构示意图。

具体实施方式

下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本发明一部分实施例，而不是全部实施例。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前提下，所获得的所有其他实施例，都属于本发明保护范围。

为了使本技术领域的人员更好地理解本发明方案，下面结合附图和具体实施方式对本发明作进一步的详细说明。

接下来，详细介绍本发明实施例所提供的一种 B+树的存取方法。图 1 为本发明实施例提供的一种 B+树的存取方法的流程图，该方法包括：

S101：获取到 B+树创建指令时，判断是否存在映射到 DRAM 内存的空闲的底层树文件。

当存在映射到 DRAM 内存的空闲的底层树文件时，则说明 DRAM 内存中存在可以用于存储 B+树的文件，此时可以执行 S102。

当不存在映射到 DRAM 内存的空闲的底层树文件时，则说明当前的DRAM 内存中不存在用于存储 B+树的文件，此时可以执行 S103。

S102: 将 B+树的底层数据存储至底层树文件。

S103: 新建一个目标底层数文件，并将目标底层树文件映射至 DRAM 内存，以便于将 B+树的底层数据存储至目标底层树文件中。

基于 B+树的数据结构，每次数据的读取都需要从底层数据访问，因此底层数据被访问的次数较频繁，通过将底层文件存储在映射到 DRAM 内存的空闲的底层树文件，有效的提升了底层数据的访问效率。

需要说明的是，为了便于区分，在本发明实施例中，将新建的底层树文件称作目标底层树文件，对于存储系统而言，新建的目标底层树文件与底层树文件均是用于存储底层数据的文件，两者并不存在实质性区别。

S104: 当 B+树的层数大于或等于预设阈值时，则将 B+树中层数大于或等于预设阈值的数据存储至预先设定的存储区域。

B+树包含有多层数据，在本发明实施例中，可以将 B+树的数据进行不同层级的划分，不同层级的数据被访问的频率和次数有所差异，因此，可以将不同层级的数据存储在不同的位置。

在具体实现中，可以将 B+树划分为第一层级、第二层级和第三层级;其中，第一层级的数据作为底层数据。

在 B+树创建初期，底层数据被访问的频率较高，因此，在本发明实施例中，将底层数据存储在映射到 DRAM 内存的空闲的底层树文件中。

考虑到数据中心级持久性内存模块 (DC Persistent Memory Module，DCPMM)在 Device DAX 模式下可以实现数据的直接存取，无需依赖于DRAM 内存，Device DAX 模式下的 DCPMM 器件可以看作是 DCPMM 内存。在本发明实施例中，可以将 B+树的第二层级的数据存储至 DCPMM 内存中。

为了精确的控制 DCPMM 的使用，避免 DCPMM 资源的浪费，可以将B+树的第三层级的数据存储至预设的硬盘中。

在实际应用中，可以将 B+树的第一层文件至第三层文件作为第一层级，将第四层文件和第五层文件作为第二层级，将大于第五层文件的数据

作为第三层级。

由上述技术方案可以看出，获取到B+树创建指令时，判断是否存在映射到DRAM内存的空闲的底层树文件；当存在映射到DRAM内存的空闲的底层树文件时，则可以直接将B+树的底层数据存储至底层树文件。当不存在映射到DRAM内存的空闲的底层树文件时，则需要新建一个目标底层数文件，并将目标底层树文件映射至DRAM内存，以便于将B+树的底层数据存储至目标底层树文件中。当B+树的层数大于或等于预设阈值时，则将B+树中层数大于或等于预设阈值的数据存储至预先设定的存储区域。基于B+树的数据结构，每次数据的读取都需要从底层数据访问，因此底层数据被访问的次数较多，通过将底层文件存储在映射到DRAM内存的空闲的底层树文件，有效的提升了底层数据的访问效率。而B+树中除底层数据外的其它数据被访问的次数相对较少，为了降低对DRAM内存的占用，可以将其它数据存储在除DRAM内存外的存储空间中，既提升了DRAM内存资源的利用率，又可以保证B+树的读写效率。

在B+树的创建初期，底层数据被访问的次数较为频繁，因此可以将底层数据存储在DRAM内存中，随着存储时间的增长，B+树的底层数据被访问的频率有所下降，为了提升DRAM内存的利用率，因此，在具体实现中，可以对存储于DRAM内存中的底层数据设置脏标识，以便于可以将存储在DRAM内存中的底层数据周期性迁移至DCPMM内存中。

为了提升数据的读取效率，在具体实现中，可以将DCPMM内存的最小读写粒度作为B+树的节点容量。DCPMM内存的内部最小读写粒度为256字节，因此，可以将一个B+树节点的容量设置为256字节，也即每256字节数据产生一组校验保护数据。

在本发明实施例中，可以将B+树的底层数据按照节点容量向底层树文件中存储各节点数据；其中，每个节点数据的键值对中存储有下一层级节点的偏移地址。

在实际应用中，每个节点可以包含一个16字节的节点头和最多15个键值对。节点头包含节点类型等信息。其中，节点类型包括叶子节点和非

叶子节点。对于叶子节点，其键值对中的值为 RAID 物理地址，依据该 RAID 物理地址，可以读取到叶子节点所对应的具体数据；对于非叶子节点，其键值对中可以存储下一层级节点的偏移地址。

当获取到数据查询指令时，可以依据数据查询指令中携带的逻辑地址，确定出根节点；根据根节点中包含的偏移地址，确定出叶子节点；并读取叶子节点对应的数据。

举例说明，当需要查询 B+树的数据时，可以在数据查询指令中携带数据卷的逻辑地址。根据该逻辑地址，可以找到这个数据卷的元数据，如果元数据中保护根节点内存地址，则该树的底层树文件在内存中，依据底层树文件中相应节点中存储的偏移地址，可以查找叶子节点，如果走到第 3 层节点仍然不是叶子节点，则从元数据读取第 3 层文件中相应节点的偏移地址，从而得到第 4 层文件的存储地址。如果该节点仍然为非叶子节点，则继续进行第 5 层节点的访问。如果 5 层节点仍然不是叶子节点，则读入第三层级的文件到内存，继续查找，直到到达叶子节点为止，根据叶子节点的键值对存储的 RAID 物理地址，便可以读取到叶子节点所对应的具体数据。

B+树的数据按照划分的层级存储在不同的位置，当需要修改 B+树的数据时，可以按照存储位置查询所需修改的数据所在的存储位置，以实现对 B+树中数据的修改，如图 2 所示为本发明实施例提供的一种 B+树的数据修改方法的流程图，包括:

S201: 当获取到数据修改指令时，判断 DRAM 内存中是否存在与数据修改指令中携带的节点标识相匹配的节点数据。

当 DRAM 内存中存在与数据修改指令中携带的节点标识相匹配的节点数据，则执行 S202。

当 DRAM 内存中不存在与数据修改指令中携带的节点标识相匹配的节点数据，则需要进一步确定所需修改的节点数据所在的位置，此时可以执行 S203。

S202: 依据数据修改指令中携带的待替换数据，对节点数据进行修改，

并对修改后的节点数据设置脏标志。

节点数据存储在DRAM内存中，会占用DRAM内存的资源，为了提升DRAM资源的利用率，可以对修改后的节点数据设置脏标志，以便于后续可以按照预设的周期时间将映射至 DRAM 内存中设置有脏标志的数据迁移至DCPMM内存中。

S203: 判断DCPMM内存中是否存在与数据修改指令中携带的节点标识相匹配的节点数据。

当DCPMM内存中存在与数据修改指令中携带的节点标识相匹配的节点数据，则执行S204。

在本发明实施例中，B+树的数据按照不同的层级会分别存储在DRAM内存、DCPMM内存以及预设的硬盘中，当DRAM内存中不存在与数据修改指令中携带的节点标识相匹配的节点数据，并且DCPMM内存中也不存在与数据修改指令中携带的节点标识相匹配的节点数据，则说明所需修改的节点数据存储在预设的硬盘中，此时可以执行S205。

S204: 依据数据修改指令中携带的待替换数据，对节点数据进行修改。

S205: 将硬盘中与数据修改指令中携带的节点标识相匹配的节点数据读取至DRAM内存中。

由于硬盘中的数据无法直接进行修改，因此，当所需修改的节点数据存储在预设的硬盘中时，需要将硬盘中与数据修改指令中携带的节点标识相匹配的节点数据读取至DRAM内存中。

S206: 依据数据修改指令中携带的待替换数据，在DRAM内存中完成对节点数据的修改，并对修改后的节点数据设置脏标志。

当所需修改的节点数据存储在DRAM内存中或者硬盘中时，修改后的节点数据会存储在DRAM内存中，由于DRAM内存资源有限，为了避免修改后的节点数据长时间占用 DRAM 内存的资源，因此在本发明实施例中，可以将存储在DRAM内存中的修改后的节点数据设置脏标识，以便于后续可以按照预设的周期时间将映射至 DRAM 内存中设置有脏标志的数据迁移至DCPMM内存中。

通过定期对存储在DRAM内存中数据迁移，避免了B+树的数据长时

间占用 DRAM 内存的资源，有效的提升了存储系统的处理性能。

图 3 为本发明实施例提供的一种 B+树的存取装置的结构示意图，包括第一判断单元 31、第一存储单元 32、创建单元 33 和第二存储单元 34;

第一判断单元 31，用于获取到 B+树创建指令时，判断是否存在映射到 DRAM 内存的空闲的底层树文件；若是，则触发第一存储单元 32；若否，则触发创建单元 33;

第一存储单元 32，用于将 B+树的底层数据存储至底层树文件;

创建单元 33，用于新建一个目标底层数文件，并将目标底层树文件映射至 DRAM 内存，以便于将 B+树的底层数据存储至目标底层树文件中;

第二存储单元 34，用于当 B+树的层数大于或等于预设阈值时，则将 B+树中层数大于或等于预设阈值的数据存储至预先设定的存储区域。

可选地，B+树对应有第一层级、第二层级和第三层级；其中，第一层级的数据作为底层数据;

相应的，第二存储单元具体用于将 B+树的第二层级的数据存储至 DCPMM 内存中；将 B+树的第三层级的数据存储至预设的硬盘中。

可选地，还包括作为单元;

作为单元，用于将 DCPMM 内存的最小读写粒度作为 B+树的节点容量。

可选地，第一存储单元具体用于将 B+树的底层数据按照节点容量向底层树文件中存储各节点数据；其中，每个节点数据的键值对中存储有下一层级节点的偏移地址。

可选地，还包括查询单元、确定单元和读取单元;

查询单元，用于当获取到数据查询指令时，依据数据查询指令中携带的逻辑地址，确定出根节点;

确定单元，用于根据根节点中包含的偏移地址，确定出叶子节点;

读取单元，用于读取叶子节点对应的数据。

可选地，还包括第二判断单元、修改单元、设置单元、第三判断单元和读取单元;

第二判断单元，用于当获取到数据修改指令时，判断 DRAM 内存中是否存在与数据修改指令中携带的节点标识相匹配的节点数据；若是，则触发修改单元；若否，则触发第三判断单元；

修改单元，用于依据数据修改指令中携带的待替换数据，对节点数据进行修改，

设置单元，用于对修改后的节点数据设置脏标志；

第三判断单元判断 DCPMM 内存中是否存在与数据修改指令中携带的节点标识相匹配的节点数据；

修改单元还用于当 DCPMM 内存中存在与数据修改指令中携带的节点标识相匹配的节点数据时，则依据数据修改指令中携带的待替换数据，对节点数据进行修改；

读取单元，用于当 DCPMM 内存中不存在与数据修改指令中携带的节点标识相匹配的节点数据时，则将硬盘中与数据修改指令中携带的节点标识相匹配的节点数据读取至 DRAM 内存中；

修改单元还用于依据数据修改指令中携带的待替换数据，在 DRAM 内存中完成对节点数据的修改，并触发设置单元对修改后的节点数据设置脏标志。

可选地，还包括迁移单元；

迁移单元，用于按照预设的周期时间将映射至 DRAM 内存中设置有脏标志的数据转移至 DCPMM 内存中。

图 3 所对应实施例中特征的说明可以参见图 1 和图 2 所对应实施例的相关说明，这里不再一一赘述。

由上述技术方案可以看出，获取到 B+树创建指令时，判断是否存在映射到 DRAM 内存的空闲的底层树文件；当存在映射到 DRAM 内存的空闲的底层树文件时，则可以直接将 B+树的底层数据存储至底层树文件。当不存在映射到 DRAM 内存的空闲的底层树文件时，则需要新建一个目标底层数文件，并将目标底层树文件映射至 DRAM 内存，以便于将 B+树的底层数据存储至目标底层树文件中。当 B+树的层数大于或等于预设阈值时，则将 B+树中层数大于或等于预设阈值的数据存储至预先设定的存储区域。基

于 B+树的数据结构，每次数据的读取都需要从底层数据访问，因此底层数据被访问的次数较多，通过将底层文件存储在映射到 DRAM 内存的空闲的底层树文件，有效的提升了底层数据的访问效率。而 B+树中除底层数据外的其它数据被访问的次数相对较少，为了降低对 DRAM 内存的占用，可以将其它数据存储在除 DRAM 内存外的存储空间中，既提升了 DRAM 内存资源的利用率，又可以保证 B+树的读写效率。

图 4 为本发明实施例提供的一种 B+树的存取装置 40 的硬件结构示意图，包括:

存储器 41，用于存储计算机程序;

处理器 42，用于执行所述计算机程序以实现如上述任意实施例所述的 B+树的存取方法的步骤。

本发明实施例还提供了一种计算机可读存储介质，所述计算机可读存储介质上存储有计算机程序，所述计算机程序被处理器执行时实现如上述任意实施例所述的 B+树的存取方法的步骤。

以上对本发明实施例所提供的一种 B+树的存取方法、装置和计算机可读存储介质进行了详细介绍。说明书中各个实施例采用递进的方式描述，每个实施例重点说明的都是与其他实施例的不同之处，各个实施例之间相同相似部分互相参见即可。对于实施例公开的装置而言，由于其与实施例公开的方法相对应，所以描述的比较简单，相关之处参见方法部分说明即可。应当指出，对于本技术领域的普通技术人员来说，在不脱离本发明原理的前提下，还可以对本发明进行若干改进和修饰，这些改进和修饰也落入本发明权利要求的保护范围内。

专业人员还可以进一步意识到，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、计算机软件或者二者的结合来实现，为了清楚地说明硬件和软件的可互换性，在上述说明中已经按照功能一般性地描述了各示例的组成及步骤。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。

结合本文中所公开的实施例描述的方法或算法的步骤可以直接用硬件、处理器执行的软件模块，或者二者的结合来实施。软件模块可以置于随机存储器（RAM）、内存、只读存储器（ROM）、电可编程ROM、电可擦除可编程ROM、寄存器、硬盘、可移动磁盘、CD-ROM、或技术领域内所公知的任意其它形式的存储介质中。

权 利 要 求

1、一种 B+树的存取方法，其特征在于，包括:

获取到 B+树创建指令时，判断是否存在映射到 DRAM 内存的空闲的底层树文件;

若是，则将所述 B+树的底层数据存储至所述底层树文件;

若否，则新建一个目标底层数文件，并将所述目标底层树文件映射至 DRAM 内存，以便于将所述 B+树的底层数据存储至所述目标底层树文件中;

当所述 B+树的层数大于或等于预设阈值时，则将所述 B+树中层数大于或等于所述预设阈值的数据存储至预先设定的存储区域。

2、根据权利要求 1 所述的方法，其特征在于，所述 B+树对应有第一层级、第二层级和第三层级; 其中，第一层级的数据作为底层数据;

相应的，所述当所述 B+树的层数大于或等于预设阈值时，则将所述 B+树中大于或等于所述预设阈值的数据存储至预先设定的存储区域包括:

将所述 B+树的第二层级的数据存储至 DCPMM 内存中;

将所述 B+树的第三层级的数据存储至预设的硬盘中。

3、根据权利要求 2 所述的方法，其特征在于，在所述获取到 B+树创建指令时，判断是否存在映射到内存的底层树文件之前还包括:

将所述 DCPMM 内存的最小读写粒度作为所述 B+树的节点容量。

4、根据权利要求 3 所述的方法，其特征在于，所述将所述 B+树的底层数据存储至所述底层树文件包括:

将所述 B+树的底层数据按照所述节点容量向所述底层树文件中存储各节点数据; 其中，每个节点数据的键值对中存储有下一层级节点的偏移地址。

5、根据权利要求 4 所述的方法，其特征在于，还包括:

当获取到数据查询指令时，依据所述数据查询指令中携带的逻辑地址，确定出根节点;

根据所述根节点中包含的偏移地址，确定出叶子节点; 并读取所述叶子节点对应的数据。

6、根据权利要求 2 所述的方法，其特征在于，还包括:

当获取到数据修改指令时，判断所述 DRAM 内存中是否存在与所述数据修改指令中携带的节点标识相匹配的节点数据;

若是，则依据所述数据修改指令中携带的待替换数据，对所述节点数据进行修改，并对修改后的节点数据设置脏标志;

若否，则判断所述 DCPMM 内存中是否存在与所述数据修改指令中携带的节点标识相匹配的节点数据;

当所述 DCPMM 内存中存在与所述数据修改指令中携带的节点标识相匹配的节点数据时，则依据所述数据修改指令中携带的待替换数据，对所述节点数据进行修改;

当所述 DCPMM 内存中不存在与所述数据修改指令中携带的节点标识相匹配的节点数据时，则将所述硬盘中与所述数据修改指令中携带的节点标识相匹配的节点数据读取至所述 DRAM 内存中;依据所述数据修改指令中携带的待替换数据，在所述 DRAM 内存中完成对所述节点数据的修改，并对修改后的节点数据设置脏标志。

7、根据权利要求 6 所述的方法，其特征在于，在所述将所述 B+树的底层数据存储至所述底层树文件之后还包括:

按照预设的周期时间将映射至 DRAM 内存中设置有脏标志的数据迁移至所述 DCPMM 内存中。

8、一种 B+树的存取装置，其特征在于，包括第一判断单元、第一存储单元、创建单元和第二存储单元;

所述第一判断单元，用于获取到 B+树创建指令时，判断是否存在映射到 DRAM 内存的空闲的底层树文件;若是，则触发所述第一存储单元;若否，则触发所述创建单元;

所述第一存储单元，用于将所述 B+树的底层数据存储至所述底层树文件;

所述创建单元，用于新建一个目标底层数文件，并将所述目标底层树文件映射至 DRAM 内存，以便于将所述 B+树的底层数据存储至所述目标底层树文件中;

所述第二存储单元，用于当所述 B+树的层数大于或等于预设阈值时，则将所述 B+树中层数大于或等于所述预设阈值的数据存储至预先设定的存储区域。

9、一种 B+树的存取装置，其特征在于，包括:

存储器，用于存储计算机程序;

处理器，用于执行所述计算机程序以实现如权利要求 1 至 7 任意一项所述 B+树的存取方法的步骤。

10、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上存储有计算机程序，所述计算机程序被处理器执行时实现如权利要求 1 至 7 任意一项所述 B+树的存取方法的步骤。

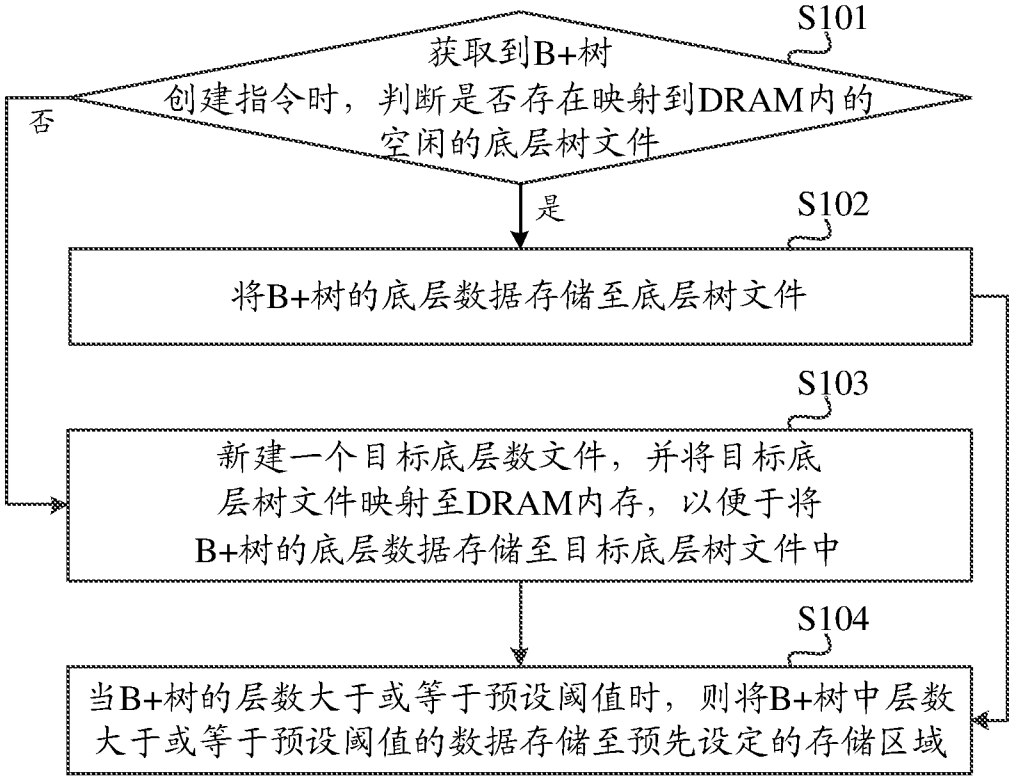
S101
获取到B+树
创建指令时，判断是否存在映射到DRAM内的
空闲的底层树文件
否
是
S102
将B+树的底层数据存储至底层树文件
S103
新建一个目标底层数文件，并将目标底
层树文件映射至DRAM内存，以便于将
B+树的底层数据存储至目标底层树文件中
S104
当B+树的层数大于或等于预设阈值时，则将B+树中层数
大于或等于预设阈值的数据存储至预先设定的存储区域

图 1

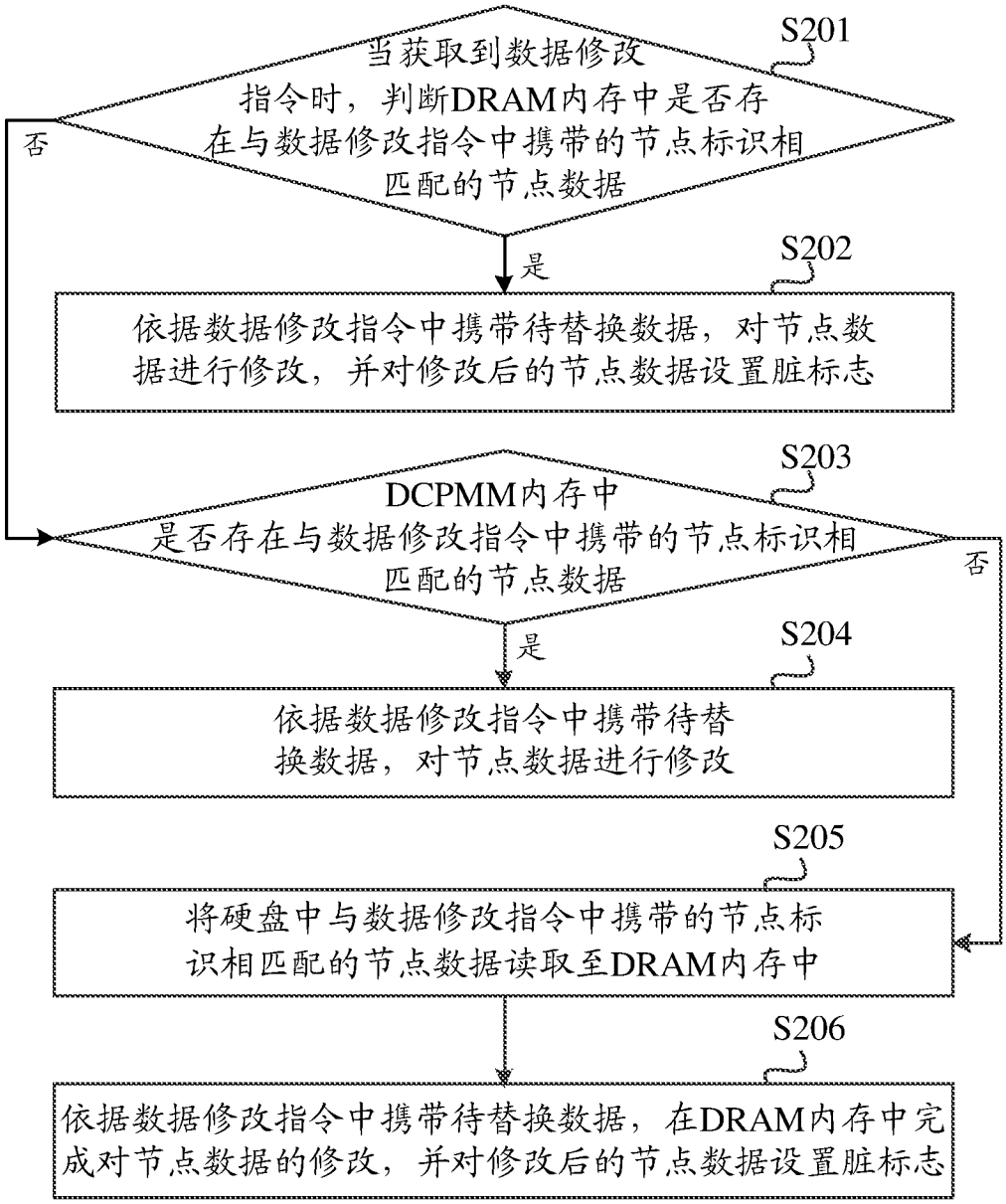
S201
当获取到数据修改指令时，判断DRAM内存中是否存在与数据修改指令中携带的节点标识相匹配的节点数据
否
是
S202
依据数据修改指令中携带待替换数据，对节点数据进行修改，并对修改后的节点数据设置脏标志
S203
DCPMM内存中是否存在与数据修改指令中携带的节点标识相匹配的节点数据
否
是
S204
依据数据修改指令中携带待替换数据，对节点数据进行修改
S205
将硬盘中与数据修改指令中携带的节点标识相匹配的节点数据读取至DRAM内存中
S206
依据数据修改指令中携带待替换数据，在DRAM内存中完成对节点数据的修改，并对修改后的节点数据设置脏标志

图 2

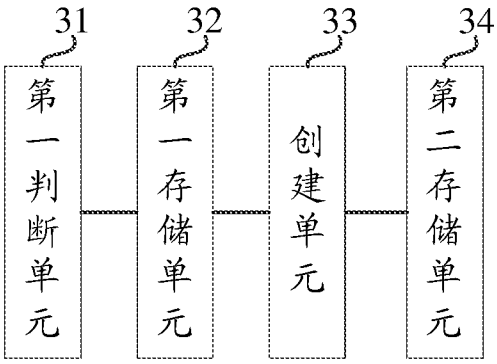

图 3

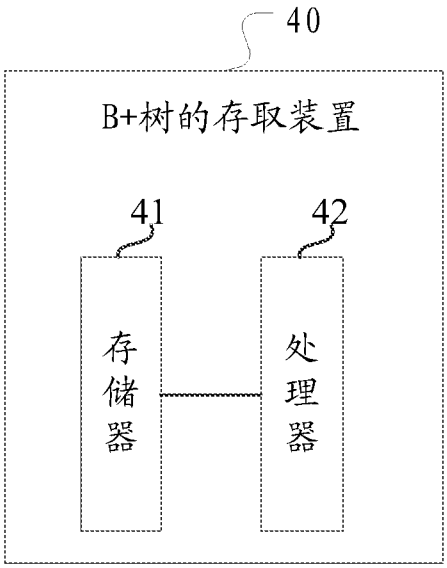

图 4

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2020/117331

A. CLASSIFICATION OF SUBJECT MATTER

G06F 3/06(2006.01)i; G06F 16/13(2019.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, CNKI, WPI, EPODOC, GOOGLE: 树, 索引, 层, 根目录, 根节点, 内存, 闪存, 快闪, 频繁, 热点, 最常, 经常, 访问, 效率, 时间, 映射, tree, b+, bottom, layer+, ram, flash, offen, frequen+, visit+, time, map+

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 111309258 A (SUZHOU INSPUR INTELLIGENT TECHNOLOGY CO., LTD.) 19 June 2020 (2020-06-19) claims 1-10	1-10
X	CN 109766312 A (SHENZHEN UNIVERSITY) 17 May 2019 (2019-05-17) description paragraphs 0029-0046, 0081-0084, figures 1, 6	1-5, 8-10
A	CN 104899297 A (NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS) 09 September 2015 (2015-09-09) entire document	1-10
A	CN 108733678 A (HUAWEI TECHNOLOGIES CO., LTD.) 02 November 2018 (2018-11-02) entire document	1-10
A	JP 2001350635 A (NEC CORPORATION) 21 December 2001 (2001-12-21) entire document	1-10

☐ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
13 December 2020	**21 December 2020**
Name and mailing address of the ISA/CN **China National Intellectual Property Administration (ISA/CN)** **No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088** **China** Facsimile No. **(86-10)62019451**	Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2020/117331

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	111309258	A	19 June 2020	None			
CN	109766312	A	17 May 2019	None			
CN	104899297	A	09 September 2015	CN	104899297	B	26 February 2019
				CN	109299113	A	01 February 2019
				CN	109376156	A	22 February 2019
				CN	109284299	A	29 January 2019
CN	108733678	A	02 November 2018	WO	2018188416	A1	18 October 2018
JP	2001350635	A	21 December 2001	None			

国际检索报告

国际申请号
PCT/CN2020/117331

A. 主题的分类

G06F 3/06(2006.01)i; G06F 16/13(2019.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06F

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称，和使用的检索词（如使用））

CNPAT, CNKI, WPI, EPODOC, GOOGLE: 树, 索引, 层, 根目录, 根节点, 内存, 闪存, 快闪, 频繁, 热点, 最常, 经常, 访问, 效率, 时间, 映射, tree, b+, bottom, layer+, ram, flash, offen, frequen+, visit+, time, map+

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
PX	CN 111309258 A (苏州浪潮智能科技有限公司) 2020年 6月 19日 (2020-06-19) 权利要求1-10	1-10
X	CN 109766312 A (深圳大学) 2019年 5月 17日 (2019-05-17) 说明书第0029-0046、0081-0084段，附图1、6	1-5, 8-10
A	CN 104899297 A (南京航空航天大学) 2015年 9月 9日 (2015-09-09) 全文	1-10
A	CN 108733678 A (华为技术有限公司) 2018年 11月 2日 (2018-11-02) 全文	1-10
A	JP 2001350635 A (NEC CORP.) 2001年 12月 21日 (2001-12-21) 全文	1-10

☐ 其余文件在C栏的续页中列出。

☑ 见同族专利附件。

* 引用文件的具体类型：

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件，或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件（如具体说明的）

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布，与申请不相抵触，但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件，单独考虑该文件，认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件，当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时，要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期	国际检索报告邮寄日期
2020年 12月 13日	2020年 12月 21日
ISA/CN的名称和邮寄地址 中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451	受权官员 王荣 电话号码 86-(10)-53961199

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2020/117331

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	111309258	A	2020年 6月 19日	无			
CN	109766312	A	2019年 5月 17日	无			
CN	104899297	A	2015年 9月 9日	CN	104899297	B	2019年 2月 26日
				CN	109299113	A	2019年 2月 1日
				CN	109376156	A	2019年 2月 22日
				CN	109284299	A	2019年 1月 29日
CN	108733678	A	2018年 11月 2日	WO	2018188416	A1	2018年 10月 18日
JP	2001350635	A	2001年 12月 21日	无			