



(12)发明专利

(10)授权公告号 CN 103365742 B

(45)授权公告日 2017.06.23

(21)申请号 201210103131.3

(56)对比文件

(22)申请日 2012.03.30

CN 1567262 A, 2005.01.19,

(65)同一申请的已公布的文献号

CN 101576833 A, 2009.11.11,

申请公布号 CN 103365742 A

CN 101140536 A, 2008.03.12,

(43)申请公布日 2013.10.23

US 8010495 B1, 2011.08.30,

(73)专利权人 伊姆西公司

US 2004/0133602 A1, 2004.07.08,

地址 美国马萨诸塞州

JP 2010122761 A, 2010.06.03,

审查员 李妮

(72)发明人 杨加林 赵军平 谢纲 齐巍
胡风华

(74)专利代理机构 北京市金杜律师事务所
11256
代理人 王茂华 姜彦

(51)Int.Cl.

G06F 11/14(2006.01)

权利要求书1页 说明书7页 附图7页

G06F 9/44(2006.01)

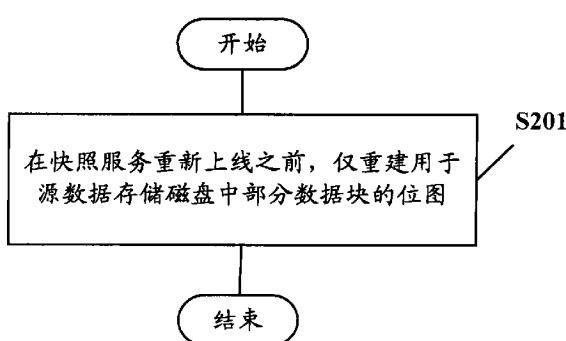
(54)发明名称

加快快照服务重新上线速度的方法、设备和
计算机程序

(57)摘要

本发明提供了一种用于在系统重启时，加快快照服务重新上线速度的方法、设备和计算机程序产品。具体地，该方法包括在快照服务重新上线之前，仅重建用于源数据存储磁盘中部分数据块的位图。通过在快照服务重新上线之前仅恢复用于部分数据块的位图，可以加快快照服务重新上线速度，以及由此缩短了主文件系统和快照文件系统恢复使用所需的时间，为用户带来了更好的体验。

200



1. 一种位图重建方法,所述方法包括:

在快照服务重新上线之前,重建仅针对源数据存储磁盘中的所有数据块中的部分数据块的位图,

其中所述部分数据块中的原始数据已被复制到快照专用磁盘并且用于所述原始数据的地址映射未同步至存储在磁盘中的地址映射表。

2. 根据权利要求1所述的方法,进一步包括:

在所述快照服务重新上线之后,异步地重建用于所述源数据存储磁盘中除所述部分数据块以外的剩余数据块的位图。

3. 根据权利要求2所述的方法,其中所述剩余数据块中的原始数据已被复制到所述快照专用磁盘并且用于所述原始数据的地址映射已同步至存储在所述磁盘中的地址映射表;或者

所述剩余数据块中的原始数据未被复制到所述快照专用磁盘。

4. 根据权利要求2所述的方法,其中:

异步地重建用于所述源数据存储磁盘中除所述部分数据块以外的剩余数据块的位图是由并发执行的线程实施的。

5. 一种位图重建设备,所述设备包括:

位图重建装置,配置用于:在快照服务重新上线之前,重建仅针对源数据存储磁盘中的所有数据块中的部分数据块的位图,

其中所述部分数据块中的原始数据已被复制到快照专用磁盘并且用于所述原始数据的地址映射未同步至存储在磁盘中的地址映射表。

6. 根据权利要求5所述的设备,其中所述位图重建装置进一步配置用于:

在所述快照服务重新上线之后,异步地重建用于所述源数据存储磁盘中除所述部分数据块以外的剩余数据块的位图。

7. 根据权利要求6所述的设备,其中:

其中所述剩余数据块中的原始数据已被复制到所述快照专用磁盘并且用于所述原始数据的地址映射已同步至存储在所述磁盘中的地址映射表;或者

所述剩余数据块中的原始数据未被复制到所述快照专用磁盘。

8. 根据权利要求6所述的设备,其中:

所述位图重建装置是由并发执行的线程实施的。

加快快照服务重新上线速度的方法、设备和计算机程序

技术领域

[0001] 本发明涉及快照服务领域,更具体地,本发明涉及一种用于在系统重启时,加快快照服务重新上线速度的方法、设备和计算机程序产品。

背景技术

[0002] 在本部分中描述的内容旨在促进对本发明的理解,但并不因为将其包括在本部分中就承认是对于本发明而言的现有技术。

[0003] 在文件系统中,为了预防数据可能由于病毒感染、配置信息崩溃、系统崩溃等原因被“污染”或者在逻辑上被损毁,可以拍下一些历史时刻的数据内容进行留底,这被形象地称为“快照(snap)”技术。流行的快照技术包括第一次写时复制(CoFW)技术。

[0004] 根据CoFW技术,在启用了快照服务之后,如果系统上层有针对主文件系统的源数据存储磁盘中自从快照生成之后从未被更新过的某个或者某些数据块(逻辑上的磁盘存储单位)的写入请求,则在更新这些数据块之前,先将数据块中存储的原始数据(即,需要获取其“快照”的数据)复制出来,放入快照文件系统的快照专用磁盘中(即,进行“快照”),然后再将新数据写入源数据存储磁盘的数据块中。换句话说,原始数据先占着位置,什么时候新数据来了,原始数据再让位。这样,源数据存储磁盘中的数据随时都是当前最新的数据,同时又能通过将源数据存储磁盘中的原始数据复制出来而使其不致丢失,以供日后使用。

[0005] 下面参考图1A-图1D对CoFW技术做详细描述。

[0006] 图1A示意性地示出了在最初未针对原始数据创建快照时的主文件系统和快照文件系统示意图。

[0007] 如图1A所示,源数据存储磁盘中的数据块(属于主文件系统)地址0、1、2...中存储有原始数据A、B、C...。在此初始状态,由于未针对原始数据执行过更新操作,所以快照专用磁盘中的数据块(属于快照文件系统)地址10、11、12...中未存储复制出来的原始数据。

[0008] 如图1A进一步所示,存在用于被复制至快照专用磁盘的原始数据的地址映射表。该地址映射表最初存在于实现主文件系统和快照文件系统的计算设备的内存中,用于存储进行了“快照”(即,从源数据存储磁盘的数据块复制到了快照专用磁盘的数据块中)的原始数据的原始存储位置与现在存储位置之间的地址映射(例如,如图1B的130所示),以使得在原始数据被移动了之后还能找到该原始数据。从图1A可见,该地址映射表存储于内存和磁盘二者中,也即,在内存中创建的地址映射表会在一段时间后被同步到磁盘,以进行持久化存储。

[0009] 图1A中还示出了位图,用于快速指示源数据存储磁盘的数据块中的原始数据是否进行了“快照”。该位图的地址空间与源数据存储磁盘中的数据块一一对应。位图的位存储有为0或1的值。若某个位存储0,则表示该位所对应的数据块中的原始数据未进行“快照”。若某个位为1,则表示原始数据进行了“快照”。可以理解,在图1A所示未针对原始数据创建快照的情况下,位图中存储的所有数值初始为0。在CoFW技术中进行访问操作时,通过查找位图(而非查找地址映射表),能够快速得知在源数据存储磁盘中某个数据块处的原始数据

是否已进行了“快照”，即，简单地通过0和1就能辨识。需要注意的是，位图仅存在于计算设备的内存中，而不会在磁盘上进行持久化。换言之，在实现主文件系统和快照文件系统的计算设备重启之后，需要在内存中重建位图（针对图1D详述）。

[0010] 进一步，参考图1B，其中示意性地示出了针对源数据存储磁盘中第一数据块地址0进行了首次更新、并且已将用于该地址0处原始数据A的地址映射(0,10)同步至磁盘时的主文件系统和快照文件系统示意图。在此示例中涉及的具体更新操作如下：

[0011] (1) 系统查看位图，发现其中对应原始数据A的地址0的位为0（如之前的图1A所示），说明原始数据A未进行“快照”；

[0012] (2) 从地址0读取原始数据A，并将该原始数据A和部分元数据（例如，原始数据A的原始地址0和校验和checksum）写入快照专用磁盘中的地址10，如110所示；

[0013] (3) 将原始数据A的原始地址0和被复制至的地址10之间的地址映射(0,10)插入到内存中的地址映射表，如130所示；

[0014] (4) 将位图中与原始数据A的原始地址0对应的位设置为1，如120所示；

[0015] (5) 将更新数据E写入至原始数据A的原始地址0，如100所示；以及

[0016] (6) 在一段时间后，将内存中的地址映射表同步至磁盘，如130所示。由此，在快照专用磁盘中创建了同步点1201。

[0017] 需要注意，同步点对应于快照专用磁盘上的一个地址，会定期地创建。位于同步点以内（地址小于同步点）的原始数据（例如，图1B中的原始数据A）所对应的地址映射都被同步到了磁盘中，这部分数据区称为稳定数据区。位于同步点以外（地址大于同步点）的原始数据所对应的地址映射还没有同步到磁盘中，在系统重启或宕机启动时需要重构修复对应的地址映射，这部分数据区称为不稳定数据区。

[0018] 与图1B类似地，图1C示意性地示出了针对第二数据块地址1进行了首次更新后的、源数据存储磁盘中的第二数据块地址1、快照专用磁盘中的数据块地址11、位图以及内存中的地址映射表的状态。与图1B中更新操作的区别在于，此时未将用于原始数据B的地址映射同步至磁盘（通过比较附图标记131与131'可知）。从上文对同步点的描述容易得知，此时的同步点仍然为1201。假设，此时计算设备例如因掉电而关闭。随后，在计算设备重启后，快照服务所涉及的内存（易失性存储器）中的地址映射表为空，位图的所有位被重新初始化为0，以及磁盘中的地址映射表尚未存储用于已被复制出来的原始数据B的地址映射（如图1D所示）。为了使快照服务重新上线，即恢复使用，需要重建内存中的位图和地址映射表，以及相应地更新磁盘中的地址映射表。

发明内容

[0019] 本发明人发现，按照已知的方法，在使快照服务重新上线之前，需要恢复用于源数据存储磁盘中所有数据块的整个位图。这样做的弊端在于：快照服务重新上线需要很长的时间，因为恢复时间与快照所涉及数据块的多少成线性比例关系。尤其在所涉及数据块非常多的情况下，这样做会严重影响各种应用的操作和用户的体验。

[0020] 为了克服上述缺陷中的一个或多个，本发明期望提供一种方法、设备和计算机程序产品，用于在系统重启时，加快快照服务重新上线速度。

[0021] 在本发明的第一方面中，提供了一种方法，例如可以包括：在快照服务重新上线之

前,仅重建用于源数据存储磁盘中部分数据块的位图。

[0022] 显然,相比于在快照服务重新上线之前需要恢复用于源数据存储磁盘中所有数据块的位图的方式,根据本发明实施方式的、通过在快照服务重新上线之前仅恢复用于部分数据块的位图,可以加快快照服务重新上线速度,以及由此缩短了主文件系统和快照文件系统恢复使用所需的时间。

[0023] 根据本发明的某些实施方式,优选地,上述部分数据块中的原始数据已被复制到快照专用磁盘并且用于该原始数据的地址映射未同步至存储在磁盘中的地址映射表,即,这些原始数据位于快照专用磁盘中的不稳定数据区(如上文所述)。

[0024] 可以理解,这种数据块中的数据是刚刚进行了快照的数据,由此,这部分数据块的活跃度较高,被进一步访问的可能性较大。可见,重建用于这部分数据块的位图,能够提高针对这部分数据块的访问访问速度,由此进一步提高系统性能。

[0025] 根据本发明的某些实施方式,优选地,在快照服务重新上线之后,异步地重建用于源数据存储磁盘中除上述部分数据块以外的剩余数据块的位图。

[0026] 根据本发明的某些实施方式,优选地,剩余数据块中的原始数据已被复制到快照专用磁盘并且用于该原始数据的地址映射已同步至存储在磁盘中的地址映射表,即,这些原始数据位于快照专用磁盘的稳定数据区(如上文所述);或者剩余数据块中的原始数据未被复制到快照专用磁盘,即,还未进行更新。

[0027] 可以理解,通常,相对于不稳定数据区中原始数据的原始数据块地址而言,这部分数据块被进一步访问的概率较小。所以,在快照服务重新上线之后再重建用于相对稳定的数据块的位图将会较小地影响、乃至不会影响文件系统用户的体验。并且,上述剩余数据块占据数据块总量的大部分。在快照服务重新上线之后、异步地(即,在后台)重建用于剩余数据块的位图,会将重建用于这一大部分数据块的位图所需的大量时间排除在快照服务重新上线所需的时间之外,由此极大地加快了快照服务重新上线速度。

[0028] 根据本发明的某些实施方式,优选地,异步地重建用于源数据存储磁盘中除上述部分数据块以外的剩余数据块的位图是由并发执行的线程实施的。

[0029] 可以理解,这样做的优势在于可以进一步加快重建用于这部分数据块的位图的速度。

[0030] 在本发明的第二方面中,提供了一种设备,例如可以包括位图重建装置,配置用于:在快照服务重新上线之前,仅重建用于源数据存储磁盘中部分数据块的位图。

[0031] 在本发明的第三方面中,提供了一种包括软件代码的计算机程序产品,当所述软件代码在计算设备上运行时,使得所述计算设备执行根据上述第一方面所述的方法。

[0032] 根据本发明的用于在系统重启时,加快快照服务重新上线速度的方法、设备和计算机程序产品,可以显著地加快快照服务重新上线速度,为文件系统的用户提供了良好的体验。

附图说明

[0033] 通过参考附图阅读下文的详细描述,本发明示例性实施方式的上述以及其他目的、特征和优点将变得易于理解。在附图中,以示例性而非限制性的方式示出了本发明的若干实施方式,其中:

[0034] 图1A示意性地示出了最初未针对原始数据创建快照时的主文件系统和快照文件系统示意图；

[0035] 图1B示意性地示出了针对源数据存储磁盘中第一数据块进行首次更新、并且已将用于第一数据块中原始数据的地址映射同步至磁盘时的主文件系统和快照文件系统示意图；

[0036] 图1C示意性地示出了针对源数据存储磁盘中第二数据块进行首次更新、但未将用于该第二数据块中原始数据的地址映射同步至磁盘时的主文件系统和快照文件系统示意图；

[0037] 图1D示意性地示出了在系统重新启动后、重建位图和地址映射之前的主文件系统和快照文件系统示意图；

[0038] 图2示意性地示出了根据本发明实施方式的一种方法，用于在系统重启时，加快快照服务重新上线速度；

[0039] 图3示意性地示出了在系统重新启动后进行位图重建的主文件系统和快照文件系统示意图；

[0040] 图4示意性地示出了根据本发明实施方式的一种设备，用于在系统重启时，加快快照服务重新上线速度；以及

[0041] 图5示意性地可以实现本发明实施方式的计算设备。

[0042] 在附图中，相同或对应的标号表示相同或对应的部分。附图中的任何元素数量均用于示例而非限制，以及任何命名都仅用于区分，而不具有任何限制含义。

具体实施方式

[0043] 下面将参考若干示例性实施方式来描述本发明的原理和精神。应当理解，给出这些实施方式仅仅是为了使本领域技术人员能够更好地理解进而实现本发明，而并非以任何方式限制本发明的范围。

[0044] 根据本发明的实施方式，提出了一种方法、设备和计算机程序产品，用于在系统重启时，加快快照服务重新上线速度。

[0045] 下面以CoFW技术为例、在文件系统中参考本发明的若干代表性实施方式，详细阐释本发明的原理和精神。需要注意的是，具体的快照技术和实现快照技术的具体系统并不构成对本发明实施方式的限制。

【0046】发明概述

[0047] 根据本发明的实施方式，通过在快照服务重新上线之前，仅重建用于源数据存储磁盘中部分数据块的位图，可以极大地加快快照服务重新上线速度。

[0048] 需要注意的是，之所以可以这样做，是因为本发明人注意到，在未重建用于源数据存储磁盘中某部分数据块的位图时，针对这部分数据块的访问可以例如通过查看地址映射表中是否存在与该数据块地址对应的地址映射来同样地实现CoFW技术。具体地，若在地址映射表中存在与源数据存储磁盘中某个数据块地址对应的地址映射，则说明该数据块地址处存储的原始数据已进行了“快照”（即，已被复制到快照专用磁盘），由此，针对这样数据块的更新可以直接进行而无需再次复制数据；而例如读取操作可以依据该地址映射被重定向至原始数据被复制到的新地址。

[0049] 在介绍了本发明的基本原理之后，下面具体介绍本发明的各种非限制性实施方式。

[0050] 示例性方法

[0051] 下面结合图1D和图3的应用场景，参考图2来描述根据本发明示例性实施方式的、用于在系统重启时，加快快照服务重新上线速度。

[0052] 具体地，如图2所示，方法200例如可以包括：在快照服务重新上线之前，仅重建用于源数据存储磁盘中部分数据块的位图，步骤S201。

[0053] 本领域技术人员可以理解，此处的“部分”数据块可以代表源数据存储磁盘中除“所有”数据块这一情况之外的任意一个或多个数据块。

[0054] 以图1D为例，例如可以重建用于源数据存储磁盘中的数据块地址0的位图，或者也可以重建用于数据块地址1的位图，等等。

[0055] 根据本发明的实施方式，优选地，上述部分数据块中的原始数据已被复制到快照专用磁盘并且用于该原始数据的地址映射未同步至存储在磁盘中的地址映射表。具体到图1D，上述部分数据块优选地是数据块地址1，因为该数据块中的原始数据B已被复制到快照专用磁盘，但用于原始数据B的地址映射未同步至存储在磁盘中的地址映射表。从上文针对快照专用磁盘的不稳定数据区的描述可知，原始数据B目前处于不稳定数据区中。通过查看快照专用磁盘中的原始数据并参考同步点，能够很容易地获知这样的原始数据。换言之，优选地，根据本发明的实施方式在快照服务重新上线之前，仅恢复用于快照专用磁盘的不稳定数据区中数据的原始地址的位图。

[0056] 接下来参考图3，以原始数据B为例详细描述重建位图的过程。

[0057] a) 系统读取快照专用磁盘的数据块地址11中的原始数据B和它的元数据(即，原始数据B的原始地址1和校验和)；

[0058] b) 将用于原始数据B的地址映射(1,11)插入内存中的地址映射表，如131所示；

[0059] c) 将位图中用于原始数据B的原始地址1的位设置为1，以表明该数据块地址1中的原始数据B已进行了“快照”，如121所示；以及

[0060] d) 在一段时间后，将内存中的地址映射(1,11)同步至磁盘中的地址映射表，如131'所示。并且，容易得知，此时快照专用磁盘上的同步点被更新为数据块地址11。

[0061] 需要注意的是，此处的步骤b) 和c) 无顺序限制，并且步骤b) 不是本发明加快快照服务重新上线速度这一技术方案必须具备的。

[0062] 进一步地，根据本发明的实施方式，优选地，在快照服务重新上线之后，异步地重建用于源数据存储磁盘中除部分数据块以外的剩余数据块的位图。

[0063] 本领域技术人员可以理解，“异步”表示在后台执行操作，这样做的优势在于：在后台重建用于剩余数据块的位图将不会影响快照文件系统和主文件系统二者的前台操作，由此不会影响应用的执行和用户的体验。

[0064] 根据本发明的实施方式，优选地，上述剩余数据块中的原始数据已被复制到快照专用磁盘并且用于该原始数据的地址映射已同步至存储在磁盘中的地址映射表。从上文针对快照专用磁盘的稳定数据区的描述可知，这部分原始数据位于稳定数据区中。同样，通过查看快照专用磁盘中的原始数据并参考同步点，能够很容易地获知这样的原始数据。在图3的示例中，这样的原始数据例如是原始数据A。

[0065] 或者,剩余数据块中的原始数据未被复制到快照专用磁盘,即,从未进行过更新。通过排除快照专用磁盘中存储的原始数据,能够很容易地获知这样的原始数据。在图3的示例中,这样的原始数据例如是原始数据C。

[0066] 根据本发明的实施方式,优选地,异步地重建用于源数据存储磁盘中除部分数据块以外的剩余数据块的位图是由并发执行的线程实施的。

[0067] 示例性设备

[0068] 在介绍了本发明示例性实施方式的方法之后,接下来,参考图4对本发明示例性实施方式的设备进行描述。

[0069] 如图4所示,设备40例如可以包括:位图重建装置401,配置用于:在快照服务重新上线之前,仅重建用于源数据存储磁盘中部分数据块的位图。

[0070] 根据本发明的实施方式,上述部分数据块中的原始数据已被复制到快照专用磁盘并且用于该原始数据的地址映射未同步至存储在磁盘中的地址映射表。

[0071] 根据本发明的实施方式,位图重建装置401进一步配置用于:在快照服务重新上线之后,异步地重建用于源数据存储磁盘中除上述部分数据块以外的剩余数据块的位图。

[0072] 根据本发明的实施方式,剩余数据块中的原始数据已被复制到快照专用磁盘并且用于该原始数据的地址映射已同步至存储在磁盘中的地址映射表;或者剩余数据块中的原始数据未被复制到快照专用磁盘。

[0073] 根据本发明的实施方式,位图重建装置401是由并发执行的线程实施的。

[0074] 计算机程序产品

[0075] 根据本发明的实施方式,提供了一种包括软件代码的计算机程序产品,当软件代码在计算设备上运行时,使得计算设备执行根据上文所述的方法。

[0076] 示例性计算设备

[0077] 下面,将参考图5来描述可以实现本发明的计算设备。图5示意性示出了可以实现根据本发明的实施例的计算设备的结构方框图。

[0078] 图5中所示的计算机系统包括CPU(中央处理单元)501、RAM(随机存取存储器)502、ROM(只读存储器)503、系统总线504、硬盘控制器505、键盘控制器506、串行接口控制器507、并行接口控制器508、显示器控制器509、硬盘510、键盘511、串行外部设备512、并行外部设备513和显示器514。在这些部件中,与系统总线504相连的有CPU 501、RAM 502、ROM 503、硬盘控制器505、键盘控制器506、串行接口控制器507、并行接口控制器508和显示器控制器509。硬盘510与硬盘控制器505相连,键盘511与键盘控制器506相连,串行外部设备512与串行接口控制器507相连,并行外部设备513与并行接口控制器508相连,以及显示器514与显示器控制器509相连。

[0079] 图5所述的结构方框图仅仅为了示例的目的而示出的,并非是对本发明的限制。在一些情况下,可以根据需要添加或者减少其中的一些设备。

[0080] 本发明的方法和设备的实施方式可以通过能够实现所需功能的硬件、软件或者软件和硬件的结合来实现,而不受任何特定软件或硬件结构的限制。硬件部分可以利用专用逻辑来实现;软件部分可以存储在存储器中,由适当的指令执行系统,例如微处理器或者专用设计硬件来执行。本领域的普通技术人员可以理解上述的设备和方法可以使用计算机可执行指令和/或包含在处理器控制代码中来实现,例如在诸如磁盘、CD或DVD-ROM的载体介

质、诸如只读存储器(固件)的可编程的存储器或者诸如光学或电子信号载体的数据载体上提供了这样的代码。本发明的设备及其模块可以由诸如超大规模集成电路或门阵列、诸如逻辑芯片、晶体管等的半导体、或者诸如现场可编程门阵列、可编程逻辑设备等的可编程硬件设备的硬件电路实现,也可以用由各种类型的处理器执行的软件实现,也可以由上述硬件电路和软件的结合例如固件来实现。

[0081] 应当注意,尽管在上文详细描述中提及了设备40的若干装置或子装置,但是这种划分仅仅并非强制性的。实际上,根据本发明的实施方式,上文描述的两个或更多装置的特征和功能可以在一个装置中具体化。反之,上文描述的一个装置的特征和功能可以进一步划分为由多个装置来具体化。

[0082] 此外,尽管在各个方面中对本发明实施方式的方法、设备和计算机程序产品进行了描述,但这并不表明在不同方面中描述的技术特征不能结合。而且,尽管在附图中以特定顺序描述了本发明方法的操作,但是,这并非要求或者暗示必须按照该特定顺序来执行这些操作,或是必须执行全部所示的操作才能实现期望的结果。相反,流程图中描绘的步骤可以改变执行顺序。附加地或备选地,可以省略某些步骤,将多个步骤合并为一个步骤执行,和/或将一个步骤分解为多个步骤执行。

[0083] 申请文件中提及的动词“包括”、“包含”及其词形变化的使用不排除除了申请文件中记载的那些元素或步骤之外的元素或步骤的存在。元素前的冠词“一”或“一个”不排除多个这种元素的存在。

[0084] 虽然已经参考若干具体实施方式描述了本发明,但是应该理解,本发明并不限于所公开的具体实施方式。本发明旨在涵盖所附权利要求的精神和范围内所包括的各种修改和等同布置。所附权利要求的范围符合最宽泛的解释,从而包含所有这样的修改及等同结构和功能。

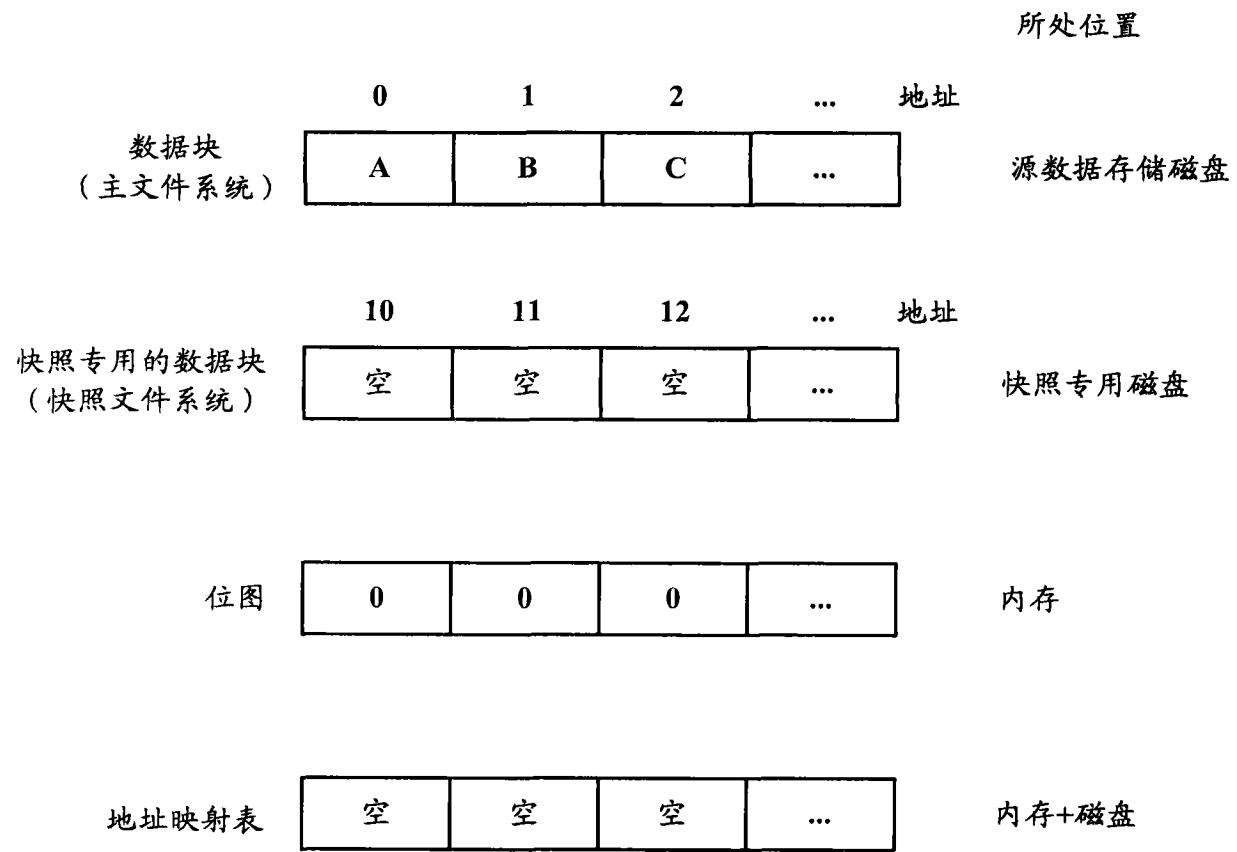


图1A

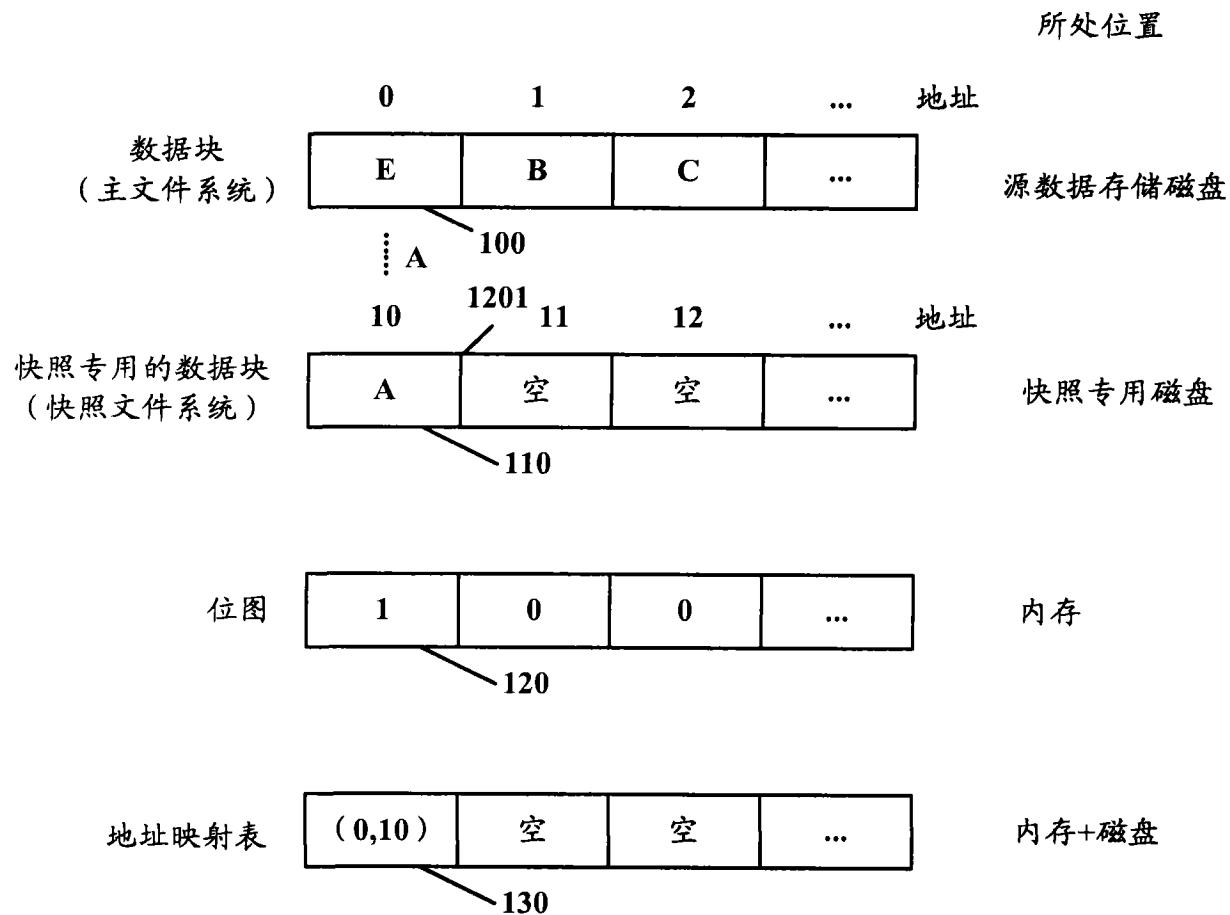


图1B

所处位置

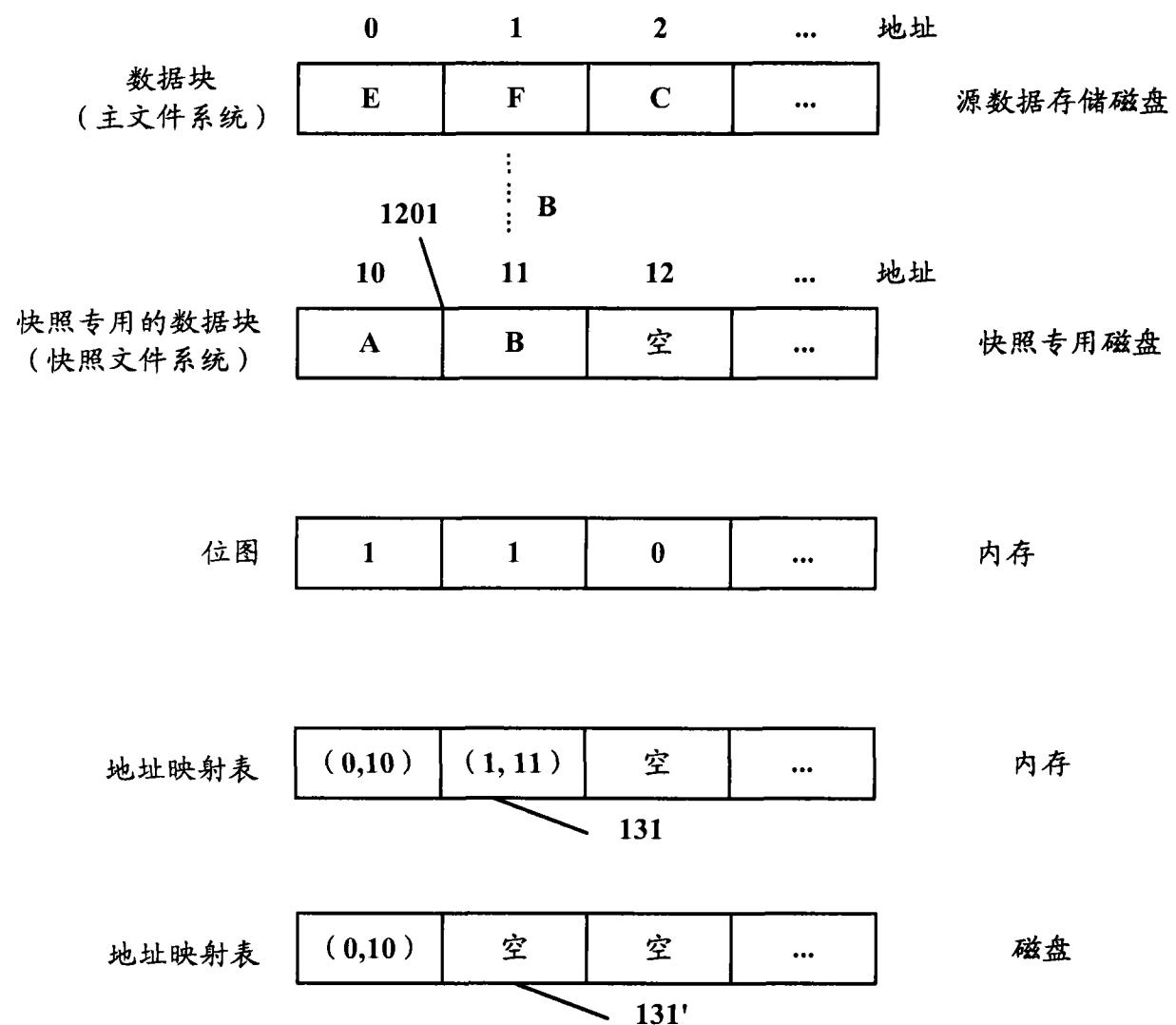


图1C

所处位置

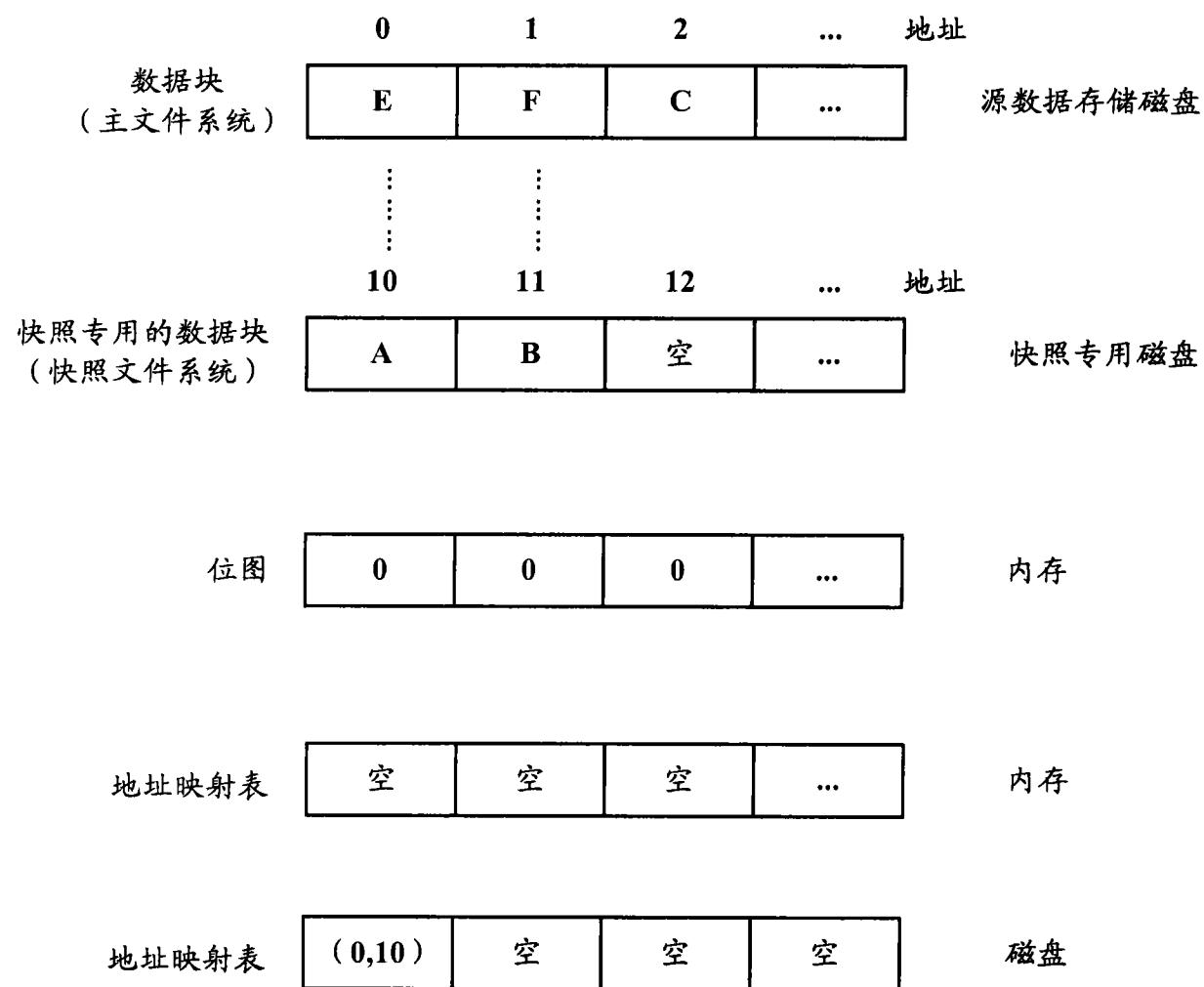


图1D

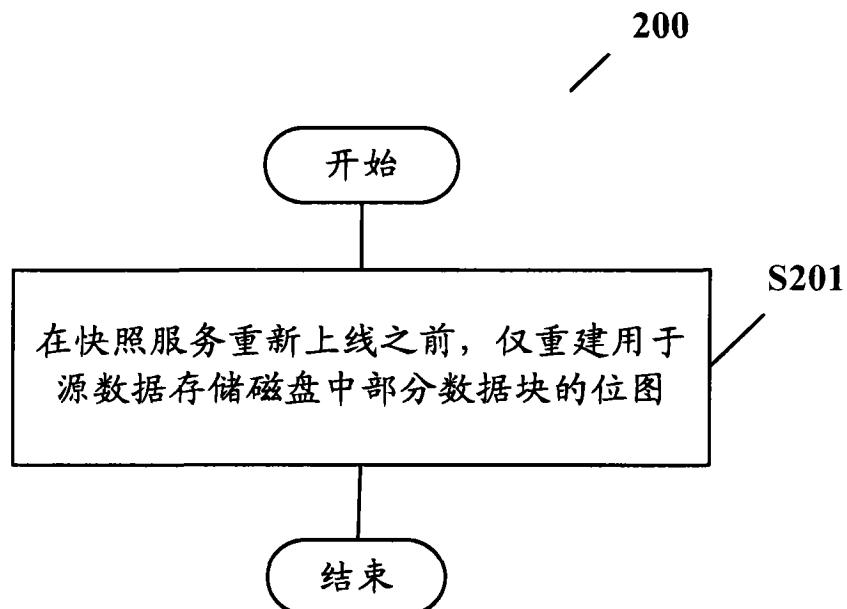


图2

所处位置

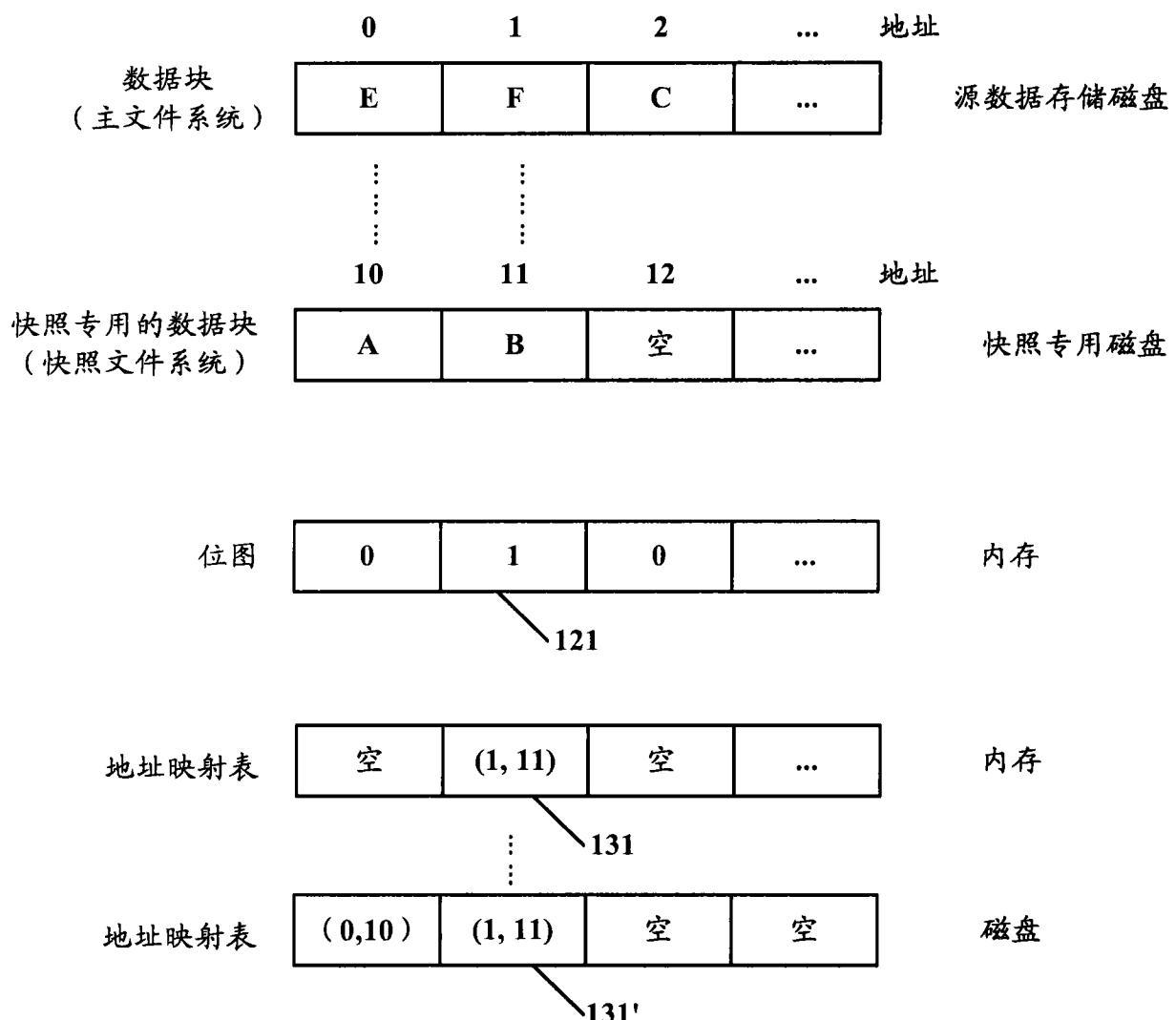


图3

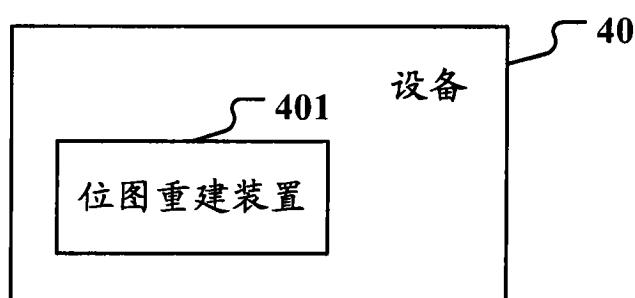


图4

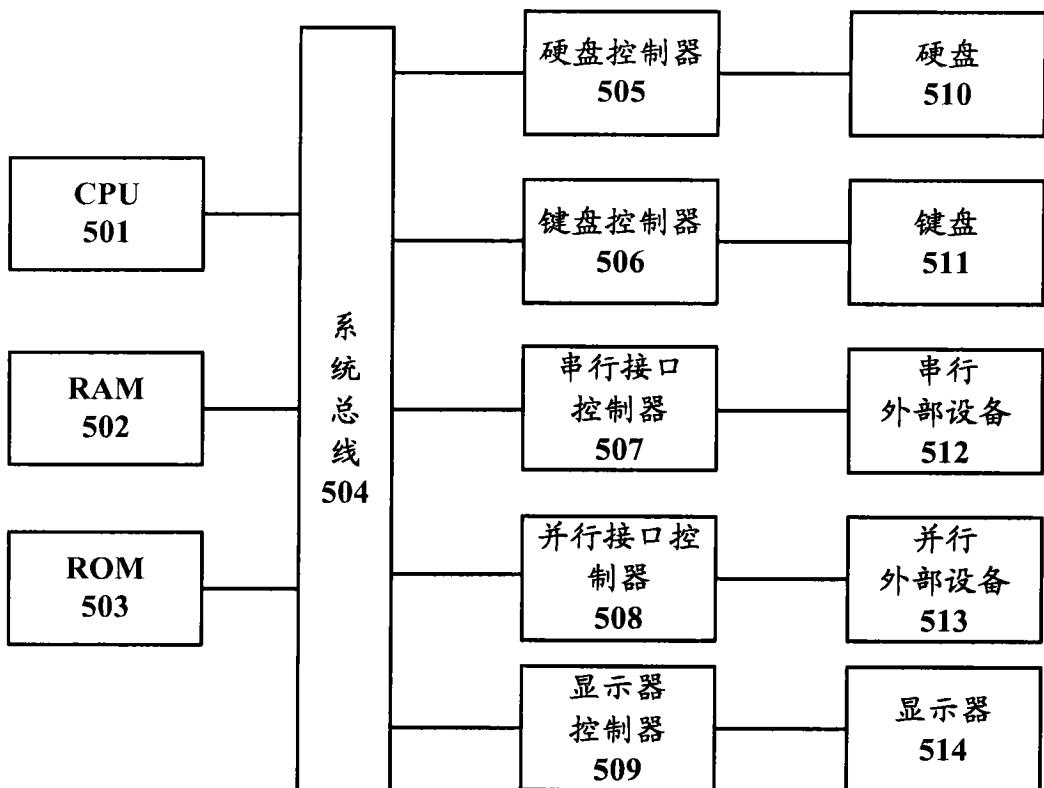


图5