(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 107025070 B
(45)授权公告日 2019.07.16

(21)申请号 201710026009.3

(22)申请日 2017.01.13

(65)同一申请的已公布的文献号
申请公布号 CN 107025070 A

(43)申请公布日 2017.08.08

(30)优先权数据
62/279,655 2016.01.15 US
15/086,020 2016.03.30 US

(73)专利权人 三星电子株式会社
地址 韩国京畿道

(72)发明人 A.雷扎伊 T.苏瑞 R.布伦南

(74)专利代理机构 北京市柳沈律师事务所 11105
代理人 邵亚丽

(51)Int.Cl.
G06F 3/06(2006.01)

(56)对比文件
US 2010241790 A1,2010.09.23,
US 2013145089 A1,2013.06.06,
CN 104081739 A,2014.10.01,

审查员 白雪涛

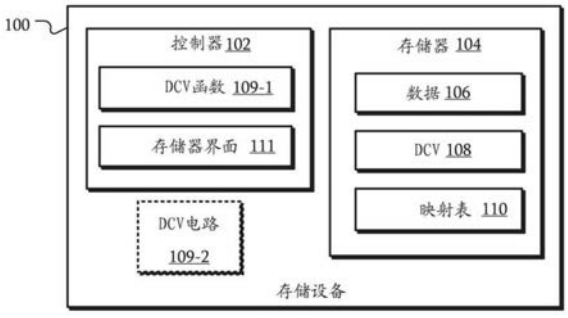
权利要求书3页 说明书13页 附图11页

(54)发明名称

版本化存储设备和方法

(57)摘要

实施例包括一种存储设备,包含:存储器;控制器,包括耦合到所述存储器的存储器接口,所述控制器被配置为:接收写数据以写到与存储在存储器中的第一数据和存储在存储器中的第一差分压缩值相关的地址;基于所述写数据和所述第一数据计算第二差分压缩值;将所述第二差分压缩值存储在存储器中;并改变地址的关联以引用第二差分压缩值而不是第一差分压缩值。



CN 107025070 B

1.一种存储设备,包括:

存储器;以及

控制器,包括耦合到所述存储器的存储器接口,所述控制器被配置为:

接收写数据以写到与存储在存储器中的第一数据和存储在存储器中的第一差分压缩值相关联的地址,所述第一差分压缩值是基于关于所述第一数据的散列函数和压缩函数确定的;

基于所述写数据和所述第一数据计算第二差分压缩值,所述第二差分压缩值是基于关于所述写数据和所述第一数据的散列函数和压缩函数确定的;

将所述第二差分压缩值存储在存储器中;以及

改变地址的关联以引用所述第二差分压缩值而不是所述第一差分压缩值。

2.如权利要求1所述的存储设备,其中,所述控制器还被配置为:

接收与地址相关联的读请求;

读取第一数据;

读取第二差分压缩值;

将所述第一数据和所述第二差分压缩值组合以形成第二数据;以及

以所述第二数据响应读请求。

3.如权利要求1所述的存储设备,其中,所述控制器还被配置为:

接收与地址相关联的读请求;

读取第一数据;以及

以第一数据响应所述读请求。

4.如权利要求1所述的存储设备,其中,所述控制器还被配置为:

保持所述地址和第一差分压缩值的关联;

接收与所述地址相关联的读请求;

读取第一数据;

读取所述第一差分压缩值;

将所述第一数据和所述第一差分压缩值组合以形成第三数据;以及

以所述第三数据响应所述读请求。

5.如权利要求1所述的存储设备,其中:

所述地址被称为逻辑地址;以及

所述控制器还被配置为:

读取第一数据;

读取第二差分压缩值;

将所述第一数据和所述第二差分压缩值组合以形成第二数据;

将所述第二数据存储在存储器中的物理地址中;以及

更新逻辑地址的关联以引用存储所述第二数据的物理地址。

6.如权利要求1所述的存储设备,其中,所述控制器还被配置为在所述存储器中保持包括多个条目的映射表,每个条目包括逻辑地址、物理地址以及差分压缩值的指示。

7.如权利要求1所述的存储设备,其中:

所述存储器包括非易失性存储器和易失性存储器;以及

所述控制器还被配置为将第一数据存储在非易失性存储器中并将第二差分压缩值存储在易失性存储器中。

8.如权利要求7所述的存储设备，其中，所述控制器还被配置为将第二差分压缩值从易失性存储器传送到非易失性存储器。

9.如权利要求1所述的存储设备，其中，所述第一数据被存储在无法执行原地写的存储器的至少一部分中。

10.如权利要求1所述的存储设备，其中，所述控制器还被配置为：

接收具有未与存储在存储器中的数据关联的新地址的写请求；

将写请求的数据写到存储器；以及

创建写到存储器的数据与所述新地址的关联。

11.一种方法，包括：

接收写数据以写到与存储在存储器中的第一数据和存储在存储器中的第一差分压缩值相关联的地址，所述第一差分压缩值是基于关于所述第一数据的散列函数和压缩函数确定的；

基于所述写数据和所述第一数据计算第二差分压缩值，所述第二差分压缩值是基于关于所述写数据和所述第一数据的散列函数和压缩函数确定的；

将所述第二差分压缩值存储在所述存储器中；以及

改变地址的关联以引用所述第二差分压缩值而不是所述第一差分压缩值。

12.如权利要求11所述的方法，还包括：

接收与所述地址相关联的读请求；

读取所述第一数据；

读取所述第二差分压缩值；

将所述第一数据和所述第二差分压缩值组合以形成第二数据；以及

以所述第二数据响应所述读请求。

13.如权利要求11所述的方法，还包括：

接收与所述地址相关联的读请求；

读取所述第一数据；以及

以所述第一数据响应读请求。

14.如权利要求11所述的方法，其中：

所述地址被称为逻辑地址；以及

所述方法还包括：

读取所述第一数据；

读取所述第二差分压缩值；

将所述第一数据和所述第二差分压缩值组合以形成第二数据；

将所述第二数据存储在存储器中的物理地址中；以及

更新逻辑地址的关联以引用物理地址。

15.如权利要求11所述的方法，还包括在存储器中保持包括多个条目的映射表，每个条目包括逻辑地址、物理地址以及差分压缩值的指示。

16.如权利要求11所述的方法，还包括：

将所述第一数据存储在非易失性存储器中;以及

将所述第二差分压缩值存储在易失性存储器中。

17.如权利要求16所述的方法,还包括将所述第二差分压缩值从易失性存储器传送到非易失性存储器。

18.一种系统,包括:

通信接口;以及

处理器,通过通信接口耦合到存储器,所述处理器被配置为:

接收写数据以写到与存储在耦合到所述处理器的存储设备中的第一数据和第一差分压缩值相关联的地址,所述第一差分压缩值是基于关于所述第一数据的散列函数和压缩函数确定的;

基于所述写数据和所述第一数据计算第二差分压缩值,所述第二差分压缩值是基于关于所述写数据和所述第一数据的散列函数和压缩函数确定的;以及

改变地址的关联以引用所述第二差分压缩值而不是第一差分压缩值。

19.如权利要求18所述的系统,还包括:

耦合到所述处理器的存储器;

其中所述处理器还被配置为在存储器中存储所述第二差分压缩值。

20.如权利要求18所述的系统,其中,所述处理器还被配置为在所述存储设备中存储第二差分压缩值。

版本化存储设备和方法

[0001] 相关申请的交叉引用

[0002] 本申请要求于2016年1月15日提交的美国临时专利申请62/279,655的权益,其内容用于所有目的通过引用全部并入本文。

技术领域

[0003] 本公开涉及存储设备,尤其是,版本化存储设备和方法。

背景技术

[0004] 存储设备可能以影响延迟的方式运行。例如,数据以页为单位写到固态存储设备/驱动器(SSD)。块从多个页创建。闪速存储器(flash memory)仅能以块为单位擦除。如果不再需要块中的一些页,则在该块中的其他有效页被读取并写到另一个块以释放旧块。然后旧块可以被擦除。这个过程被称为垃圾回收。

[0005] 垃圾回收可能增加存储设备的延迟时间。特别是,当执行垃圾回收时SSD可能无法处理读和/或写请求。其结果是,进入的读/写请求可能被延迟直到垃圾回收已完成。

[0006] 某些硬盘驱动器使用叠瓦式磁记录(shingled magnetic recording)。使用叠瓦式磁记录,磁道在存储介质上重叠。当存储在磁道数据被修改,并且磁道被重写时,重叠的磁道还必须能读和重写。当被执行时,这些额外的操作增加延迟。

发明内容

[0007] 实施例包括存储设备,包括:存储器;控制器,包括耦合到存储器的存储器接口,控制器被配置为:接收写数据以写到与存储在存储器的第一数据和存储在存储器中的第一差分压缩值相关联的地址;基于写数据和第一数据计算第二差分压缩值;将第二差分压缩值存储在存储器中;及以改变地址的关联以引用第二差分压缩值而不是第一差分压缩值。

[0008] 实施例包括一种方法,包括:接收写数据以写到与存储在存储器的第一数据和存储在存储器中的第一差分压缩值相关联的地址;基于写数据和第一数据计算第二差分压缩值;将第二差分压缩值存储在存储器中;以及改变地址的关联以引用第二差分压缩值而不是第一差分压缩值。

[0009] 实施例包括一种系统,包括:通信接口;以及通过通信接口耦合到存储器的处理器,该处理器被配置为:接收写数据以写到与存储在存储器的第一数据和存储在存储器中的第一差分压缩值相关联的地址;基于写数据和第一数据计算第二差分压缩值;将第二差分压缩值存储在存储器中;以及改变地址的关联以引用第二差分压缩值而不是第一差分压缩值。

附图说明

[0010] 图1是根据一些实施例的存储设备的示意图。

[0011] 图2A至图2E是根据一些实施例的写到存储设备的示意图。

[0012] 图3A至图3E是根据一些实施例的从存储设备读取的示意图。

[0013] 图4A至图4B是根据一些实施例的存储设备的示意图。

[0014] 图5是根据一些实施例的存储设备中的页大小的示意图。

[0015] 图6是根据一些实施例的固态存储设备的示意图。

[0016] 图7是根据一些实施例的具有版本化存储系统的系统的示意图。

[0017] 图8是根据实施例的服务器的示意图。

[0018] 图9是根据实施例的服务器系统的示意图。

[0019] 图10是根据实施例的数据中心的示意图。

具体实施方式

[0020] 该实施例涉及版本化存储设备和方法。给出以下描述以使本领域的普通技术人员能够制造和使用本实施例，并且以下描述是在专利申请及其要求的上下文中提供的。对实施例的各种修改以及本文所述的一般原理和特征将是显而易见的。实施例主要是依据在特定的实现中提供的具体方法、设备和系统进行说明。

[0021] 然而，所述方法、设备和系统将在其它的实现中有效地操作。诸如"实施例"、"一个实施例"和"另一实施例"可能指相同或不同的实施例以及多个实施例。将相对于具有特定组件的系统和/或设备描述实施例。然而，所述系统和/或设备可包括比示出的那些更多或更少的部件，并且可以对组件的布置和类型做出变化而不脱离本公开范围。实施例也将在具有某些操作的特定方法的上下文中描述。然而，所述方法和系统可以根据与实施例不一致的具有不同和/或额外的操作以及不同顺序和/或并行的操作的其他方法操作。因此，实施例并不旨在限于示出的具体实施例，而是将被赋予与本文所描述的原理和特征一致的最广范围。

[0022] 实施例在具有某些部件的特定系统或设备的上下文中描述。本领域的普通技术人员将容易认识到，实施例与具有其它和/或额外组件和/或其它特征的系统或设备的使用相一致。也可在单个元件的上下文中描述方法、设备和系统。然而，本领域的普通技术人员将容易认识到，方法和系统与具有多个元件的体系结构的使用相一致。

[0023] 本领域技术人员将理解，在一般情况下，本文，特别是在所附权利要求(例如，所附权利要求的主体)所用的术语一般旨在作为"开放式"术语(例如，术语"包括了"应被理解为"包括了但不限于"，术语"具有"应该被理解为"具有至少"，术语"包括"应被理解为"包括但不限于"，等等)。本领域的人员还将理解，如果旨在指定所引入的权利要求陈述的特定数量，则这样的意图将明确地陈述，并且如果不存在这样的陈述，则不存在这样的意图。例如，为帮助理解，下面的所附权利要求可以含有对引导性短语"至少一个"和"一个或多个"的使用，以引入权利要求的陈述。然而，即使当相同的权利要求包括引导短语"一个或多个"或"至少一个"和诸如"一"或"一个"的不定冠词(例如，"一"和/或"一个"应被解释为是指"至少一个"或"一个或多个")时，使用这样的短语不应被理解为暗示通过不定冠词"一"或"一个"所引入的权利要求陈述会限制含有这样引入的权利要求陈述为仅包含一个这样的陈述的例子的任何特定权利要求；对于用于引入权利要求陈述的定冠词的使用同样适用。此外，在其中使用了类似于"A、B或C等中的至少一个"的那些实例中，通常这样的造句的意图是本领域技术人员将理解该约定的某种意义上(例如，"具有A、B或C中的至少一个的系统"将包

括但不局限于具有单个A、单个B、单个C、A和B两者、A和C两者、B和C两者和/或A、B和C三者等的系统)。本领域技术人员还将理解,无论在说明书、权利要求书或附图中,事实上任何表示两个或多个可选术语的析取词语和/或短语,应该被理解为旨在包括术语中的一个、术语中的任一个或术语两者的可能性。例如,短语"A或B"将被理解为包括"A"或"B"或"A和B"的可能性。

[0024] 图1是根据一些实施例的存储设备的示意图。在一些实施例中,存储设备100包括控制器102和存储器104。控制器102是被配置为管理存储设备100的操作的电路,并且包括诸如通用处理器、数字信号处理器(DSP)、专用集成电路、微控制器、可编程逻辑器件、分立电路、这些设备的组合等的组件。控制器102可以包括诸如寄存器、高速缓冲存储器、处理内核等的内部部分,还可以包括诸如地址和数据总线接口、中断接口等的外部接口。尽管在存储设备100中只示出了一个控制器102,但可以存在多个控制器102。此外,诸如缓冲区、存储器接口电路、通信接口等的其它的接口设备,可以是存储设备100的一部分,以连接控制器102到内部和外部组件。

[0025] 在一些实施例中,控制器102包括通信接口,其包括使存储设备100能够通信的电路。例如,通信接口可以包括通用串行总线(USB)、小型计算机系统接口(SCSI)、外围组件互连快速(PCIe)、串行连接SCSI(SAS)、并行ATA(PATA)、串行ATA(SATA)、NVM快速(NVMe)、通用快闪存储(UFS)、光纤通道、以太网、远程直接存储器访问(RDMA)、无限带宽或其它接口。使用这样的通信接口,存储设备100可以被配置为通过关联媒介与外部设备和系统进行通信。在一些实施例中,控制器102被配置为通过通信接口接收读和写请求。

[0026] 存储器104是能够存储数据的任何设备。这里,示出了一个存储器104作为存储设备100;然而,在存储设备100中可以包括任何数量的存储器104,包括不同类型的存储器。存储器104的例子包括动态随机存取存储器(DRAM)、根据各种标准如DDR、DDR2、DDR3、DDR4的双倍数据速率同步动态随机存取存储器(DDR SDRAM)、静态随机存取存储器(SRAM)、诸如闪速存储器的非易失性存储器诸、自旋转移力矩磁阻随机存取存储器(STT-MRAM)、相变RAM,纳米浮栅存储器(NFGM)或聚合物随机存取存储器(PoRAM)磁或光媒介等。

[0027] 存储器104被配置为存储数据106、差分压缩值(DCV(多个))108和映射表110。如将在下面进一步详细描述的,存储器104可以包括多个存储器设备存储,并且在存储器104中的数据可以以各种方式分布在这样的设备之中;然而,为了方便起见,在这里数据将被描述为存储在单个存储器104中。

[0028] 控制器102包括耦合到存储器104的存储器接口111。控制器102被配置为通过存储器接口111访问存储器104。存储器接口111可以包括控制器102和存储器104通过其进行通信的命令、地址和/或数据总线的接口。尽管存储器104被示出为和控制器102分开,但在一些实施例中,存储器104的部分,诸如高速缓冲存储器、SRA等,是控制器102的一部分。控制器102可以包括内部通信总线,诸如处理内核、外部通信接口、高速缓冲存储器等的内部组件通过其进行通信。

[0029] 数据106代表已存储在存储设备100中的数据。如将在下面进一步详细描述的,DCV(多个)108代表当与相关联的数据106相组合时,代表存储在存储设备100中的当前数据的数据。在一些实施例中,DCV 108是具有小于相应的数据106的大小的大小的值。例如,数据106和DCV 108可以各自存储在具有不同大小的页中。在本例中,数据106的页的大小为8K字

节。相比之下，相应的DCV 108的页大小为4K字节。尽管已使用特定的大小作为例子，在其他实施例中，大小是不同的。

[0030] 控制器102被配置为接收带有写数据的写请求以写到与存储在存储器104中的数据106和存储在存储器104中的第一DCV 108相关联的地址。映射表110包括具有诸如逻辑地址、物理地址、DCV地址/值的信息的条目以创建地址/值或其他信息之间的关联。映射表110可以使用页、块或混合映射策略；然而，本文中的实施例将使用块映射作为例子。

[0031] 控制器102被配置为使用存储在映射表110中的数据以识别与逻辑地址相关联的物理地址、DCV 108等。例如，在通过通信接口接收到带有逻辑地址的读或写请求之后，控制器102被配置为访问映射表110以读取与逻辑地址相关联的条目。在一些实施例中，控制器102被配置为访问存储映射表110的内部高速缓存存储器；然而，在其他实施例中，控制器102被配置为访问例如DRAM的外部存储器。

[0032] 控制器102被配置为读取存储在与逻辑地址相关联的存储器104的物理地址上的数据106。控制器102被配置为基于包括在写请求中的写数据和从物理地址读取的数据106来计算DCV。该计算基于存储在物理地址上的数据106和进入的写数据之间的差生成DCV。此外，该计算生成具有大小比从物理地址读取的数据和/或写数据106小的DCV。

[0033] 可以以各种不同的方式来计算DCV。在一些实施例中，控制器102被配置为在软件或内部电路中计算DCV。DCV函数109-1代表控制器102的这个操作。也即，在控制器102已经从物理地址接收到写数据和读数据106之后，控制器102被配置为使用写数据和从存储器104读取的数据106来执行数学计算以生成DCV。在其它实施例中，控制器102使用外部电路来计算DCV。例如，控制器102被配置为将写数据和读数据引导到DCV电路109-2，并且作为响应，DCV电路109-2被配置为生成DCV。DCV电路109-2包括运算单元、查找表、输入/输出缓冲器等，以计算DCV并与控制器102和/或存储器104连接。DCV电路109-2以虚线示出以指示其可以作为控制器102的DCV函数109-1的替代和/或与其配合来使用。

[0034] 可以使用由控制器102直接地或通过DCV电路109-2间接地执行的各种函数生成DCV。在一些实施例中，函数是简单的减法操作。在其他实施例中，函数是更复杂的散列函数。在其它实施例中，可以创建记录指示哪些位已翻转。在一些实施例中，函数可以被称为“diff”函数。在特定的实施例中，函数被优化以减小DCV的大小。在一些实施例中，函数包括压缩函数以减小差别的大小。

[0035] 控制器102还被配置为直接地或通过DCV电路109-2间接地执行DCV函数的反函数。反DCV函数是使用源数据和从源数据以及差别数据生成的DCV作为输入，以重新生成差别数据的函数。因此，通过保持数据106和DCV 108，可以通过反DCV函数获得差别数据。

[0036] 控制器102被配置为在存储器104中存储计算出的DCV。例如，控制器102被配置为将计算出的DCV和DCV(多个)108存储在存储器104中。然而，如将在下面进一步详细描述的，在其他实施例中，在和DCV(多个)108一起存储之前，DCV可以被高速缓存在存储器104中的一个或多个其他部分中。

[0037] 控制器102被配置为改变地址的关联来引用计算出的DCV而不是较早的与地址相关联的DCV。如将在下面进一步详细描述的，当读取了与地址相关联的数据106时，现在将访问新的DCV而不是较早的DCV。换句话说，在一些实施例中，仅保持一个DCV以代表与地址相关联的数据。

[0038] 在一些实施例中，存储器104的至少一部分具有相对于写的非对称的性能。例如，基于闪速存储器的存储设备不允许原地写。必须分配新的块用于写，并且为了准备未来的写，先前的块必须被擦除。在具有叠瓦磁记录的存储设备中，向与另一磁道重叠的磁道写包括重写重叠的磁道。如将在下面进一步详细描述的，通过使用DCV(多个)，可以减小不对称性能的效果。

[0039] 图2A至图2E是根据一些实施例的向存储设备写的示意图。将使用图1的存储设备作为例子。参照图1和图2A，在一些实施例中，映射表210对应于存储在存储器104的映射表110。映射表210包括多个条目211。每个条目211包括用于逻辑块地址(LBA)、物理块地址(PBA)以及DCV的指示的字段。尽管已使用特定字段作为例子，但在其它实施例中，可以存在其他字段和/或字段可以采取不同的形式。例如，在这里，逻辑和物理地址与块相关联；然而，在其他实施例中，逻辑和物理地址可以与块中的页或存储器104的其他组织相关联。在其他实施例中，单个地址，如物理地址，可以是仅有的存在的地址。

[0040] 在映射表210中，两个条目211-1和211-2是预先存在的条目。当接收到向新的逻辑块地址的写请求202-1时，由控制器102创建新条目211-3。在本例中，新条目211-3用于逻辑块地址1。控制器102已将逻辑块地址1与物理块地址23相关联。然而，由于与物理块地址相关联的唯一数据是输入数据D，例如，数据D与新的文件相关联，所以DCV未被计算。也就是说，在数据存储器204-1中的物理地址23上不存在有效的数据。DCV字段中的"X"代表DCV不存在或者无效的指示。在一些实施例中，标志(flag)代表DCV是否存在或有效。在其他实施例中，字段的特定地址/值可被定义为DCV不存在或者无效的指示。在其他实施例中，DCV字段可以不是条目211-3的一部分。DCV不存在或者DCV是无效的指示可以采取其它形式。

[0041] 由于写请求202-1与新条目211-3相关联，所以控制器102被配置为将数据D存储在数据存储器204-1的物理块地址23中。这个数据存储器204-1代表其中存储了数据106的存储器104的一部分。因此，引用存储在数据存储器204-1中的数据，创建新的有效的条目211-3。虽然数据D被描述为被写到数据存储器204-1，但在其他实施例中，在提交数据D到数据存储器204-1之前，可以作为写的一部分执行各种缓冲区、高速缓存等。

[0042] 参照图1和图2B，在一些实施例中，接收了新的写请求202-2。再次地，写请求202-2被导向到逻辑块地址1。然而，由于条目211-3存在，所以在数据存储器204-1的相关的物理块地址中已经存储了数据，即，图2A的数据D。因此，该写请求202-2旨在以新数据D'更新存储在逻辑块地址1中的数据。

[0043] 不同于如上相对于图2A所描述的数据D的原始写，这里，数据D'不写到数据存储器204-1的物理块地址23。控制器102被配置为确定在映射表210中是否存在条目211。在一些实施例中，控制器102将写请求的逻辑块地址与现存的条目211的逻辑块地址进行比较。如果发现了匹配，则先前的写已经发生，并且在数据存储器204-1中数据存在。控制器102被配置为读取存储在相关联的物理块地址中的数据D。这里，控制器102从物理块地址23中读取。如图2A中所描述的先前写的数据D被从数据存储器204-1中读出。

[0044] 写请求202-2的新数据D'和来自数据存储器204-1的现有数据D被用作DCV函数208-1的输入。DCV函数208-1代表如上所述的由控制器102执行的计算DCV的操作。DCV函数208-1被配置为基于新数据D'和现有数据D生成DCV'。在这里，以撇号标记DCV'以指示该DCV'可以与数据D组合以生成数据D'。

[0045] 控制器102被配置为在DCV存储器204-2中存储新的DCV'。在一些实施例中，控制器102被配置为将DCV地址存储在映射表210中的条目211-3的DCV字段中。这里，以下划线示出0x12的值，以指示该值是新的或已改变。在其他实施例中，可以不使用DCV存储器204-2并且可以在条目211-3的DCV字段中存储DCV'。例如，DCV(多个)可能相对较小，因此，用于在条目211中存储DCV的额外的存储器的量可以相对小。无论如何，条目211-3现在具有逻辑块地址、物理块地址和DCV的关联。

[0046] 参照图1和图2C，在一些实施例中，控制器102接收到另一个写请求202-3。该写请求202-3是将新数据D'写到逻辑块地址1的请求。再次地，控制器102被配置为试图在映射表210中找到匹配。这里，条目211-3存在。与图2B类似，控制器102被配置为从由条目211-3所指示的物理块地址23读取数据D。数据D从数据存储器204-1输出。

[0047] 写请求202-3的新数据D″和来自数据存储器204-1的现有数据D被用作DCV函数208-1的输入。DCV函数208-1被配置为基于新的数据D″和现有数据D生成新的DCV″。尤其是，新DCV″的生成不涉及在图2B中写的数据D'的中间状态。由于数据D是存储在数据存储器204-1中的物理块地址23的数据，因此感兴趣的差别是数据D和要写的数据即数据D″之间的差别。控制器102被配置为在DCV存储器204-2中存储所得到的DCV″。控制器102被配置为使用条目211-3的DCV字段中的DCV地址以使用新DCV″去覆盖先前存在的DCV'。因此，在条目211-3中的DCV字段现在引用新DCV″。

[0048] 参考图1和图2D，操作可以类似于那些相对于图2C描述的操作。然而，在一些实施例中，将DCV″写到DCV存储器204-2可能会导致新的DCV地址。例如，如果DCV存储器204-2在闪速存储器中实现，则新的DCV″可被写到与存储DCV″页不同的DCV存储器204-2中的页。新DCV地址是在其中存储了DCV″的DCV存储器204-2的地址。在条目211-3中，新地址由0x21代表并加了下划线以指示DCV字段中的变化。

[0049] 尽管已经使用在DCV 204-2存储器中的相同地址更新存储器位置或以新地址更新条目211-3作为如何改变逻辑块地址与DCV的关联来引用新DCV的例子，但在其他实施例中，关联的变化可能不同。例如，在一些实施例中，控制器102被配置为将DCV″存储在映射表中，例如在DCV字段中。存储在条目211-3中的DCV'可以使用DCV″替代。

[0050] 参照图1和图2E，在一些实施例中，控制器102被配置为接收DCV冲刷访问(flush access)202-4。这里，DCV冲刷访问202-4指示与上述图类似的数字逻辑块地址1。映射表210在类似于图2的映射表210的初始状态210-1中。

[0051] 再次地，控制器102被配置为访问条目211-3并从数据存储器204-1中的相关联的物理块地址23并读取数据D。然而，控制器102被配置为使用DCV字段来访问存储在DCV存储器204-2的DCV″。这里，控制器102被配置为使用DCV地址来访问DCV存储器204-2。

[0052] 控制器102被配置为将DCV″和数据D提供给反DCV函数208-2。反DCV函数208-2代表如上所述的组合数据和DCV以产生数据的更新的版本的函数。这里，反DCV函数208-2使用数据D和DCV″重新创建数据D″。控制器102被配置为在数据存储器204-1中保存数据D″替代数据D。映射表210被更新为状态210-2。这里更新了条目211-3以指示由″x″代表的有效的DCV不存在。因此，后续的向逻辑块地址1的写可以由如上述图2B的控制器102处理。

[0053] 作为上述操作的结果，经常修改的数据对带有具有非对称性能的存储器104的存储设备的性能具有减小的影响。例如，被持续修改的数据可能包括不到5％的整个数据集。

使用200GB为例，只有1％或2GB可能被持续地更新。较小大小的DCV(多个)值减小了被写的数据量。在一些应用中，DCV的大多数可能在数据的整个块大小的20％的量级。其余的大多数可能仍然小于块的大小的50％。因此，在200GB的例子中，可以写400MB到1GB的DCV。减小的大小导致空间效率并且可以减小磨损。特别是，本来要使用2GB的新擦除块的更新现在可以使用400MB。对于存储设备的给定容量，减小对新擦除块的要求使得较不频繁地执行垃圾回收，并减小介质上的磨损的频率。

[0054] 图3A至图3E是根据一些实施例的从存储设备读取的示意图。将使用图1的存储设备100作为例子。在图3A至图3D中，为简洁起见，省略了与图2A至图2E的元件类似的元件的描述。参照图1和图3A，映射表310代表在如图2A所述的数据写后的状态。也即，数据D已经被存储在数据存储器304-1中，并且条目311-3已被添加到映射表310；然而，条目311-3包括DCV不存在或者DCV是无效的指示。

[0055] 控制器102被配置为接收读请求302。在此，读请求302是读取逻辑块地址1的请求。作为响应，控制器102被配置为访问映射表310并读取物理块地址23。使用该物理块地址，控制器102被配置为从数据存储器304-1读取数据D。控制器102被配置为以数据D响应读请求302。特别地，因为条目311-3包括DCV不存在或DCV无效的指示，所以控制器102被配置为以未经修改的数据D来响应。

[0056] 参照图1和图3B，在一些实施例中，当映射表310是在如图2B中所述的数据写之后的状态时，控制器102可以接收读请求302。即，数据D最初被写到数据存储器304-1并且写了更新的数据D'，导致DCV'存储在DCV存储器304-2中。

[0057] 因此，控制器102被配置为再次接收读请求302，并且作为响应，通过访问物理块地址23来从数据存储器304-1读取数据。然而，因为在条目311-3中存在有效的DCV字段，所以控制器102被配置为访问DCV存储器304-2来读取DCV'。控制器102被配置为使用数据D和DCV'作为反DCV函数308-2的输入以组合数据D和DCV'成为数据D'。控制器102被配置为以数据D'响应读请求302。

[0058] 虽然该技术涉及增加的读取量，但该增加的读出量具有小的，如果不是可忽略的影响。例如，在一些实施例中，存储器104的内部读带宽高于存储设备100的外部接口带宽。即使对读性能具有不可忽略的影响，读取也不会导致使用擦除过的块、重写相邻的磁道等。因此，能够降低读取性能的操作相比减小延迟、改善延迟一致性等相关联的操作具有较小的影响。

[0059] 参照图1和图3C，在一些实施例中，当映射表310是在如图2C中所述的数据写之后的状态时，控制器102可以接受读请求302。即，数据D最初被写到数据存储器304-1，写更新的数据D'导致DCV'被存储在DCV存储器304-2中，并再次写更新的数据D″，导致DCV″被存储在DCV存储器304-2中。

[0060] 控制器102被再次配置为访问条目311-3得到物理块地址23，使用物理块地址23来访问数据存储器304-1以访问数据D，并使用DCV地址访问DCV存储器304-2。然而，由于数据D″是最近被写到逻辑块地址1的数据，所以DCV'是可用并在DCV存储器304-2中被访问的DCV。

[0061] 控制器102被配置为使用DCV″和数据D作为反DCV函数308-2的输入以生成数据D″。控制器102被配置为以数据D″响应读请求302。因此，即使初始数据D被访问，也可以再生成

最近的数据D″。特别是，生成数据D″未使用数据D'和相关联的DCV'。

[0062] 如图3A至图3C所示，控制器102被配置为基于DCV字段是否指示DCV存在或有效来不同地操作。然而，在其他实施例中，即使只有初始数据被存储在数据存储器304-1中，控制器102也可被配置为类似于图3B和图3C来操作。特别是，无论是在DCV字段本身中或是在DCV存储器304-2中，条目311的DCV字段都可以被初始化以指示身份(identity)DCV。当和初始数据D一起被用作反DCV函数308-2的输入时，身份DCV指示，生成数据D。作为结果，不管与实际数据比较的DCV数据是否存在或有效，控制器102可以执行基本相同的操作。

[0063] 参照图1和图3D，在一些实施例中，数据D的较早初始版本是可访问的。特别是，控制器102可以被配置为接收源读请求302-1。这里，源读请求302-1参照逻辑块地址1。作为响应，控制器102被配置为类似于相对于图3A描述的访问来访问数据存储器304-1的物理块地址23中的数据D。然而，条目311-3的DCV字段是有效的，类似于图3B和图3C。也即，在逻辑块地址1中存在数据的更新版本。相比于图3B和图3C，被返回的是初始数据D而不是再生成的当前数据D'或D″。因此，可以使用类似源读请求302-1的请求来读取存储在存储设备100上的数据的早期版本。

[0064] 虽然使用了从条目311-3的DCV字段读取DCV地址作为例子，但在其他实施例中，DCV值是从条目311-3的DCV字段读取。例如，图3B的DCV'和图3C的DCV″可以分别从条目311-3读取。

[0065] 参照图1和图3E，在一些实施例中，当映射表310是在如图2C所述的数据写之后的状态时，控制器102可以接收读请求302-2。也即，数据D最初被写到数据存储器304-1，写更新数据D'导致DCV'被存储在DCV存储器304-2中，并且写进一步更新的数据D″，导致DCV″被存储在DCV存储器304-2中。然而，在这个实施例中，保持了一个或多个中间DCV(多个)。在这个示例中，DCV″是当前DCV;然而，也保持了DCV'。在条目311-4中的附加参数0x55代表DCV'的指示，即其值或地址。

[0066] 在本示例中，读请求302-2是对LBA 1'的请求。LBA 1'代表例如数据D'的数据的状态。因此，控制器102被配置为访问DCV'地址，其是其中DCV'在DCV存储器304-2中存储的位置的地址。结果是，DCV'可以被访问并在反DCV函数308-2中与数据D组合来生成数据D'。

[0067] 虽然仅使用一个中间DCV，即，DCV'，作为例子，但在其他实施例中，可以存储任何数量的中间DCV(多个)。例如，(多个)DCV、DCV'、DCV″、DCV″'和DCV″″全部可以被存储在DCV存储器304-2中。这些DCV(多个)中的每一个可分别与数据D组合来生成诸如数据D'、D″、D″'和D″″的更高版本的数据。

[0068] 图4A和图4B是根据一些实施例的存储设备的示意图。参照图4A，在一些实施例中，存储设备400包括类似于图1的存储设备100的控制器102的控制器402。然而，存储设备400包括非易失性存储器404-1和易失性存储器404-2。非易失性存储器404-1的例子包括诸如闪速存储器、STT-MRAM、相变RAM、NFGM或PoRAM、磁或光介质等的存储器。易失性存储器的404-2例子包括DRAM、根据诸如DDR、DDR2、DDR3、DDR4的各种标准等的DDR SDRAM、SRAM等。

[0069] 控制器402被配置为在非易失性存储器404-1中存储数据406并将在易失性存储器404-2中存储DCV 408。控制器402还可以被配置为将映射表410存储在易失性存储器404-2中。

[0070] 在一些实施例中，存储设备400的使用使得一致性延迟的优先级高于一致性。因

此，数据的一致性可以被放宽。例如，一些互联网规模的应用考虑放宽界限内的一致性是可以接受的。这些应用包括tweets和照片标记。然而，对于这样的应用，延迟尖峰可能是不可接受的。在这些界限内的数据，如由DCV(多个)408所表示的，被存储在易失性存储器404-2中。控制器402被配置为冲刷(flush)DCV 408的溢出(overflow)到非易失性存储器404-1。在一些实施例中，控制器402被配置为如上述图2E中描述的冲刷DCV 408到非易失性存储404-1的数据406。虽然易失性存储器404-2可以由电池、超级电容器或NVRAM备份；然而，在一些实施例中，这样的备份不是必要的，因为导致DCV 408的损失的故障将仍然在可接受的范围内。省略用于备份的电池、超级电容器或NVRAM可以降低存储设备400的成本。

[0071] 在一些实施例中，在提交数据到非易失性存储器404-1之前可以使用高速缓存412高速缓存数据。例如，在图2A至图2D中描述的写请求的数据可以被存储在高速缓存412中。没有使用图2A至图2D中描述的各种技术来执行该存储。然而，当数据被从高速缓存412中逐出、从高速缓存412中提交等时，可以使用图2A至图2D中描述的技术。在特定的示例中，存储在高速缓存412中的数据块可以由多个写请求更新。这些更新不涉及DCV的计算。当该数据块是要被逐出或提交时，基于当它被逐出或被提交时数据的状态，可以如图2A至图2D中所述生成DCV。

[0072] 参照图4B，在一些实施例中，存储设备401类似于图4A的存储设备400。然而，在存储设备401中，控制器402被配置为在易失性存储器404-2中高速缓存DCV(多个)，其由高速缓存DCV 408-1代表。具体地，控制器402被配置为保持在易失性存储器404-2中频繁访问的DCV(多个)。控制器402被配置为使用高速缓存算法、启发式算法等来确定在高速缓存中保持哪些DCV 408-1。控制器402被配置为传送其他DCV(多个)408-1到存储在非易失性存储器404-1中的DCV 408-2，反之亦然。其结果是，在一些实施例中，可以高速缓存被频繁访问或重度访问的数据。此外，由于DCV(多个)408-1的大小比相应的数据406要小，所以被保持在易失性存储器404-2中的更新的数量比如果高速缓存实际的数据更大。

[0073] 图5是根据一些实施例的存储设备中的页大小的示意图。在一些实施例中，数据页502和DCV页504被存储在相同的存储器500中。这里，数据页502-1和502-2以及DCV页504-1到504-4是存储在存储器500中的数据页502和DCV页504的示例。这里，DCV页504被示出为较小以代表相对于源数据的DCV的更小的大小。在这个例子中，DCV页504是数据页502的一半大小；然而，在其他实施例中，DCV页504的大小取决于所使用的特定的DCV函数而不同。

[0074] 在一些实施例中，数据页502和DCV页504两者可能会受到同样的限制，如缺乏原地写，其可能会增加延迟。如上所述，通常保持数据页502而DCV变化，导致DCV页的改变招致延迟影响。然而，由于DCV页504在大小上更小，所以可以减小延迟影响。例如，DCV页504可以被存储在与数据页502不同的块中。因此，随着更多DCV(多个)在DCV页504中累积，可以对那些块进行垃圾回收以恢复空闲块。因为DCV页504较小，所以垃圾回收可能需要较少的时间和/或以较低的频率执行。此外，数据页502可以比DCV页504保持更长有效。其结果是，存储数据页502的块将比存储DCV页504的块更不太可能经历垃圾回收。另外，由于DCV页504可能更容易很快失效，所以在存储DCV页504的块中更多页将是无效的，因而由于较少有效页需要被复制所以降低了垃圾回收操作的时间。

[0075] 图6是根据一些实施例的固态存储设备的示意图。在一些实施例中，SSD600包括类似于图1的存储设备100的控制器102的控制器602。然而，SSD600包括闪速存储器604-1、

DRAM 604-2和SRAM 604-3。控制器602被配置为将数据606和DCV(多个)608-1存储在闪速存储器604-1中、高速缓存DCV 608-2在DRAM 604-2中、并将映射表610存储在SRAM 604-3中。虽然已对特定的存储器的配置进行描述,但在其它实施例中,SSD 600包括不同配置和/或不同数据606、DCV(多个)608-1和608-2分布的不同存储器,并且映射表610是不同的。例如,SSD 600可以具有类似于图1、图4A和图4B的配置。

[0076] 存储在闪速存储器604-1中的数据606和DCV 608-1很容易受到其中包含数据606和DCV 608-1页被擦除的垃圾回收的影响,可能产生额外的延迟。然而,如以上描述的,DCV 608-1的大小比相应的数据606小。也就是说,闪速存储器604-1中用于存储数据606的页比用于存储DCV 608-1的页大。

[0077] 使用DCV(多个)降低了SSD 600中的写修改(异地更新(out-of-placeupdate))。这导致最小化可能需要被回收的无效块。这转而降低垃圾回收的频率。特别是,SSD 600可以以更低的垃圾回收频率和因此较低机会的更高延迟,尤其是延迟高峰,来处理诸如大数据和云应用的更新I/O密集的流量。具体而言,在数据606写更新时,数据606的第一副本在闪速存储器604-1中保持有效。即使对于写更新密集的工作量,在闪速存储器604-1中的以及高速缓存在DRAM 604-2中的作为高速缓存的DCV 608-2的DCV(多个)608-1也能增加闪速存储器604-1中的页的寿命。此外,如上所述,数据606的早期版本是可用的。除了改进延迟,该架构也可以改善闪速单元磨损。

[0078] 特别地,控制器602被配置为连同代表先前页和更新之间差别的DCV一起,在闪速存储器604-1中保持先前页为有效/活跃,而不是使前一页无效并请求更新数据的新页。在随后的读操作时,控制器602同时读取先前的页和DCV两者并组合提供最新的页。如上所述,存在在哪里存储DCV的多种配置,例如在闪速存储器604-1、DRAM 604-2、SRAM 604-3或这些存储器的组合中。此外,DCV可以被高速缓存,诸如在DRAM 604-2上高速缓存活跃或"热"页DCV(多个),并在闪速存储器604-1中保持DCV 608-1的持久副本。

[0079] 在一些实施例中,即使当读取与写并非不对称、具有相同的开销等,本文描述的存储设备仍然可能具有性能优势。尤其是,如果写次数随着写的大小扩展,并且DCV的大小小于对应的数据块的大小,则可以减少写时间。

[0080] 在一些实施例中,DRAM 604-2用于高速缓存写请求和/或写数据。当处理写请求涉及数据606中的现有数据的读取时,对应的数据606作为高速缓存数据606-1被存储在DRAM 604-2中。控制器602被配置为使用存储在DRAM 604-2中的高速缓存数据606-1作为上述DCV函数的输入。也即,控制器602被配置为从存储在DRAM 604-2中的高速缓存的数据606-1读取数据,而不是从存储在闪速存储器604-1中的数据606中读取数据。

[0081] 在一些实施例中,闪速存储器604-1被分为平面和通道。SSD 600以页为粒度读和写,以块为粒度擦除,块可以包括多个页。映射策略定义翻译的粒度,即,页级别的映射可能需要较大的印迹,但提供了较高程度的灵活性。块级别的映射可以使用更小的印迹,但可能在布置上有限制性。已经提出"混合型"策略的几个变形以利用基于页和块映射两者的组合。映射表610中的映射可以使用任何这样的映射技术。

[0082] 虽然已经描述了SSD 600作为例子,但在其他实施例中,也可以使用其它类型的存储介质,如叠瓦磁驱动器(shingled magnetic drive)。尤其是,在叠瓦磁盘上,写可能与先前写的磁道的一部分重叠。这导致写性能的降低,而且还导致覆盖相邻的磁道。叠瓦磁驱动

器通过在固件中管理这个问题来隐藏该复杂性。较低的写性能(当写到相邻磁道时)可能会导致不一致的延迟,并且是大多数云应用的顾虑。使用如本文所述的DCV减小覆盖,因此对更新密集型工作负载提供一致的处理延迟。

[0083] 图7是根据一些实施例的具有版本化存储系统的系统的示意图。在一些实施例中,系统700包括通过通信链路706耦合到存储设备704的主机702。主机702是使用存储设备的数据存储容量的系统,例如通用处理器、数字信号处理器(DSP)、专用集成电路、微控制器、可编程逻辑器件、分立电路、这些设备的组合等。在一些实施例中,主机702是计算机、服务器、工作站等。主机702被配置为执行诸如操作系统和应用程序的软件。主机702被耦合到存储器708。存储器708包括用于主机702的可操作存储器和/或高速缓存存储器,如DRAM、根据诸如DDR、DDR2、DDR3、DDR4的各种标准的DDR SDRAM、SRAM等。尽管存储器708被示为与主机702分开,但在一些实施例中,存储器708是主机702的一部分。

[0084] 通信链路706代表主机702和存储设备704被配置为通过其通信的媒介。例如,通信链路706是如USB、SCSI、PCIe、SAS、ATA、SATA、NVMe、UFS、光纤通道、以太网、RDMA、Infiniband或其他类似链接的链路。主机702和存储设备704中的每一个被配置为具有通过这种通信链路进行通信的接口。

[0085] 在一些实施例中,如图2A至图3D,存储设备704是如上所述的类似于存储设备100、400、401和600并被配置为如上所述操作的存储设备。主机702可以被配置为读和写数据到存储设备704。存储设备704被配置为如上所述的使用DCV以改进系统700的操作。

[0086] 然而,在其他实施例中,主机702被配置为相对于DCV的执行类似于如上所述的操作。例如,主机702可以包括存储设备驱动器710。存储设备驱动器710代表在主机702上的可执行的操作存储设备704的软件。

[0087] 特别地,存储设备驱动器710被配置为相对于DCV(多个)执行类似于如上所述的操作。也即,上述由控制器102、402、602等执行的操作由存储设备驱动器710执行。另外,存储器708被用于类似于存储器的104、404-2、604-2和604-3的至少一部分。也就是说,存储器708被用于,例如,存储映射表110、410或610,DCV(多个)108或408,和/或高速缓存的DCV(多个)408-1或608-2。

[0088] 在一些实施例中,在映射表中不需要存在相对于图2A至图3D所描述的逻辑块地址与物理块地址的关联。存储设备驱动器710可被配置为保持逻辑地址和DCV(多个)的指示的关联。也即,存储设备驱动器710被配置为类似于物理块地址使用逻辑块地址作为初始数据D存储位置的指示。在一些实施例中,存储设备驱动器710被配置为如相对于图2A至图3D所描述的使用存储设备704作为数据存储器204-1或304-1。

[0089] 在一些实施例中,对于逻辑块地址,存储设备驱动器710被配置为将初始数据写到存储设备704上的逻辑块地址。存储设备驱动器710被配置为将初始数据高速缓存在存储器708中。当随后数据变化时,存储设备驱动器710被配置为使用高速缓存的初始数据来计算DCV并将DCV写到存储设备。在其它实施例中,存储设备驱动器710被配置为在存储器708中保持DCV(多个)。可以在存储设备驱动器710中使用任何上述的DCV管理技术。

[0090] 图8是根据实施例的服务器的示意图。在本实施例中,服务器800可包括独立的服务器、机架式服务器、刀片服务器等。服务器800包括存储设备802和处理器804。处理器804耦合到存储设备802。虽然仅示出了一个存储设备802,但可能存在任何数量的存储设备

802。存储设备802可以是任何上述的存储设备。相应地，服务器800的性能可以得到改进。

[0091] 图9是根据实施例的服务器系统的示意图。在本实施例中，服务器系统900包括多个服务器902-1至902-N。服务器902各自耦合到管理器904。服务器902中的一个或多个可以类似于上面描述的服务器800。

[0092] 管理器904被配置为管理服务器系统900的服务器902和其它组件。在实施例中，管理器904可被配置为监视服务器902的性能。例如，服务器902中的每一个可以包括如上所述的存储设备。

[0093] 图10是根据实施例的数据中心的示意图。在本实施例中，数据中心1000包括多个服务器系统1002-1至1002-N。服务器系统1002可以类似于上述图9中的服务器系统900。服务器系统1002被耦合到诸如因特网的网络1004。因此，服务器系统1002可以通过网络1004与各个节点1006-1至1006-M通信。例如，节点1006可以是客户端计算机、其他服务器、远程数据中心、存储系统等。

[0094] 可以使用实施例到诸如云和大规模延迟敏感这样的其中一致性能是重要因素的服务中。这些服务的例子包括数据分析、机器学习、音频和视频流；然而，在其他实施例中，服务的类型是不同的。一致性延迟可以是高优先级，并且一些实施例使用采用宽松的一致性要求的分布式软件栈支持更可预测的性能。尽管有几个因素导致加载延迟的不一致，如资源共享和队列，但SSD(多个)中的垃圾回收可能占了显著的比重。

[0095] 当前的软件堆栈可使用复制，但该解决方案没有解决诸如SSD(多个)内的垃圾回收的根本问题，结果是，仅可以提供有限的增益。此外，复制固有地代价更高，导致增加的网络流量，其可能进一步影响网络延迟，并且也使用软件层的协调。

[0096] 随着大速率的数据生成，可以使用更快的数据分析。在诸如搜索引擎和社交媒体的在线服务中，一个重要的设计目标是提供可预测的性能。在这样的场景下，平均响应时间可能无法代表性能；最坏情况下的性能可能受到更多顾虑。响应时间的可变性可能引起在服务组件中的更高的尾延迟。其结果是，用户可能会体验到长响应时间。根据工作量和SSD固件策略，尾延迟高峰可能偶尔或经常发生，但在大多数情况下，可能足以危及用户体验和当前高度竞争市场中的服务提供商的声誉。

[0097] 在共享基础设施上尾延迟的不利进一步加剧，如Amazon AWS和谷歌云。几个云供应商都有这种顾虑，并在设计硬件/软件系统两者的体系结构时，这是公认的主要挑战之一。尽管有几个因素，诸如资源共享和排队，导致加载等待时间不一致，但在SSD(多个)中的垃圾回收占据了显著的比重。由于闪速存储器的特性，其不容许原地更新，所以SSD固件在异地写更新，并使以前的副本无效。为了回收空间，在可被重新写之前无效空间需要被擦除。然而，相比于读或写操作(微秒)，擦除操作(以毫秒为单位)显著较慢，并且通常以粗粒度进行。该过程构成SSD(多个)中的垃圾回收，并当垃圾回收活动时通道无法服务读/写请求。因此，垃圾回收可能严重地影响关键操作的延迟(和性能)。在一些情况下，在垃圾回收期间读取延迟可能增加100倍。此外，随着SSD老化，垃圾回收器更加频繁地运行。

[0098] 在云中几种突出的应用类型是写更新密集的。模拟真实世界的使用情况的一个这样的例子是雅虎的云服务标准(YCSB)，这是评估云计算系统的基准套件。在YCSB中提供的一些更新密集工作量中，访问比可以是50％读取、0％插入和50％更新。电子商务应用是这样的应用的例子，包括诸如存储在用户会话中最近的行动记录，电子商务用户的典型动作

的操作。在这种类别的应用中，由于SSD(多个)中的垃圾回收的尾延迟效果可以造成用户响应时间的较高的延迟。在另一个工作量例子中，访问比率是95％读取、0％插入和5％更新。一个例子是社交媒体，即使很小的更新比例也可以触发GC及违反服务水平目标。例如，对于照片标记，加入标记是更新，但大多数操作是读取标签。

[0099] 大多数大数据和云应用使可扩展性和一致性能的优先级高于更传统的方面，如事务(原子性、一致性、隔离性、耐用性或ACID)属性。强一致性不能很好地扩展，而且大多数云应用倾向于具有稳定性能的目标的较为宽松的事务一致性。因此，弱一致性模型被广泛使用于最流行的云规模分布式软件栈。这提供了在高负荷、高并发系统中显著的I/O性能的改进。

[0100] 虽然所有的云应用存储大量的数据，并且被设计为随着数据印迹的增加扩展，但只有一个子集比其他的更频繁地被访问。这种不均衡的分配代表数据中心的"热"(或#trending)数据访问模式。

[0101] 本文所述的实施例可以在各种应用中使用，并提供减小的和/或更一致的延迟的益处。上述云应用仅仅是这样应用的示例。

[0102] 虽然已经根据特定的实施例描述了结构、设备、方法和系统，本领域的普通技术人员将很容易地认识到对所公开的实施例的许多变化是可能的，因此任何变化应被认为在本公开的结构、设备和系统的精神和范围之内。相应地，可以由本领域的普通技术人员作出许多修改而不脱离所附权利要求的精神和范围。

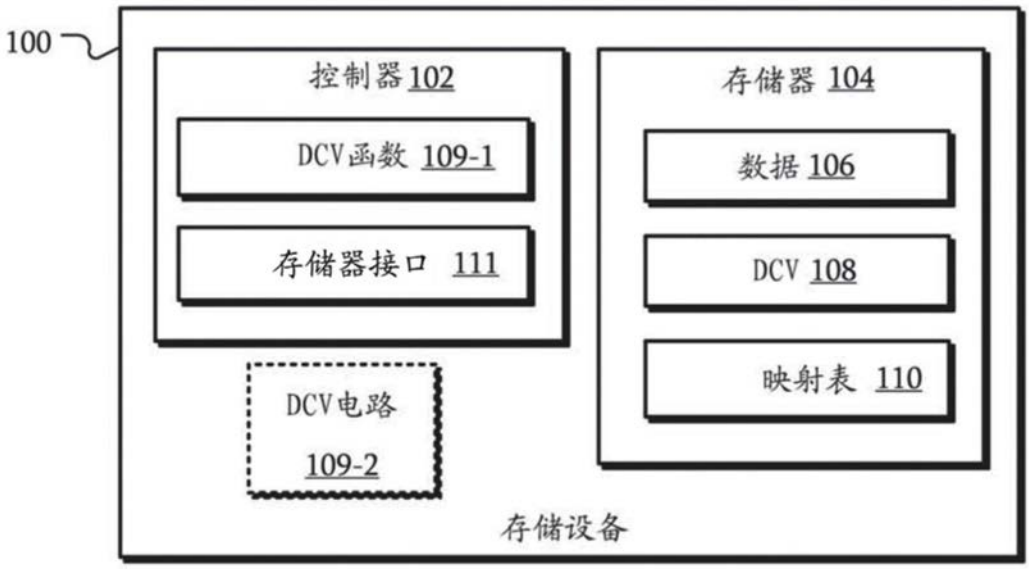

100
控制器102
DCV函数 109-1
存储器接口 111
存储器 104
数据 106
DCV 108
映射表 110
DCV电路
109-2
存储设备

图1

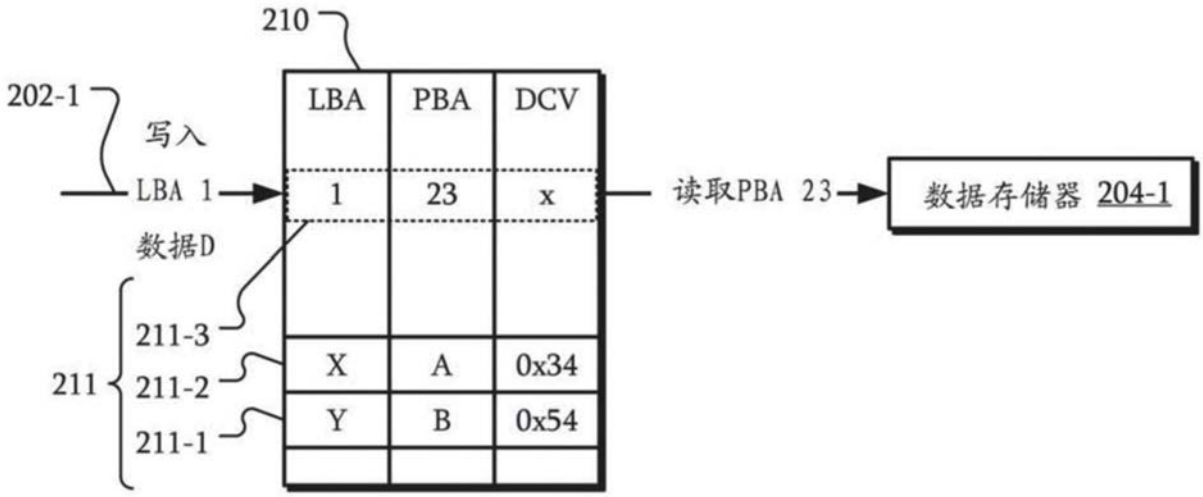

210
202-1
写入
LBA 1
数据D
LBA
PBA
DCV
1
23
x
读取PBA 23
数据存储器 204-1
211
211-3
211-2
211-1
X
A
0x34
Y
B
0x54

图2A

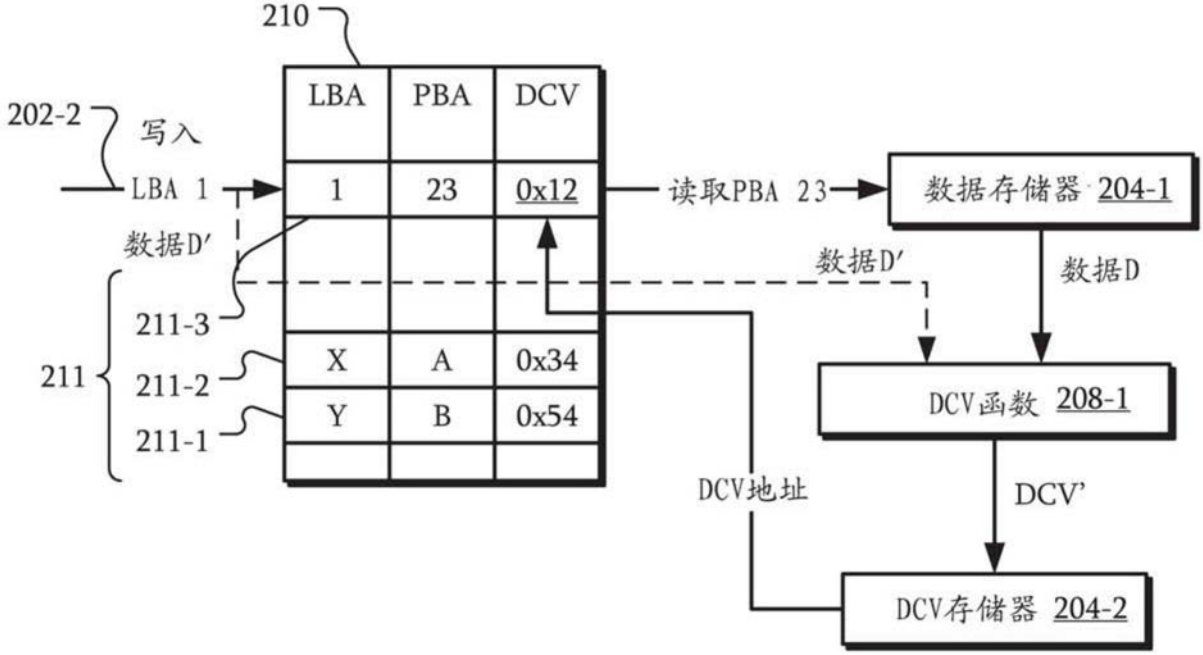

210
LBA
PBA
DCV
202-2
写入
LBA 1
1
23
0x12
读取PBA 23
数据存储器 204-1
数据D′
数据D′
数据D
211-3
211
X
A
0x34
211-2
DCV函数 208-1
Y
B
0x54
211-1
DCV地址
DCV'
DCV存储器 204-2

图2B

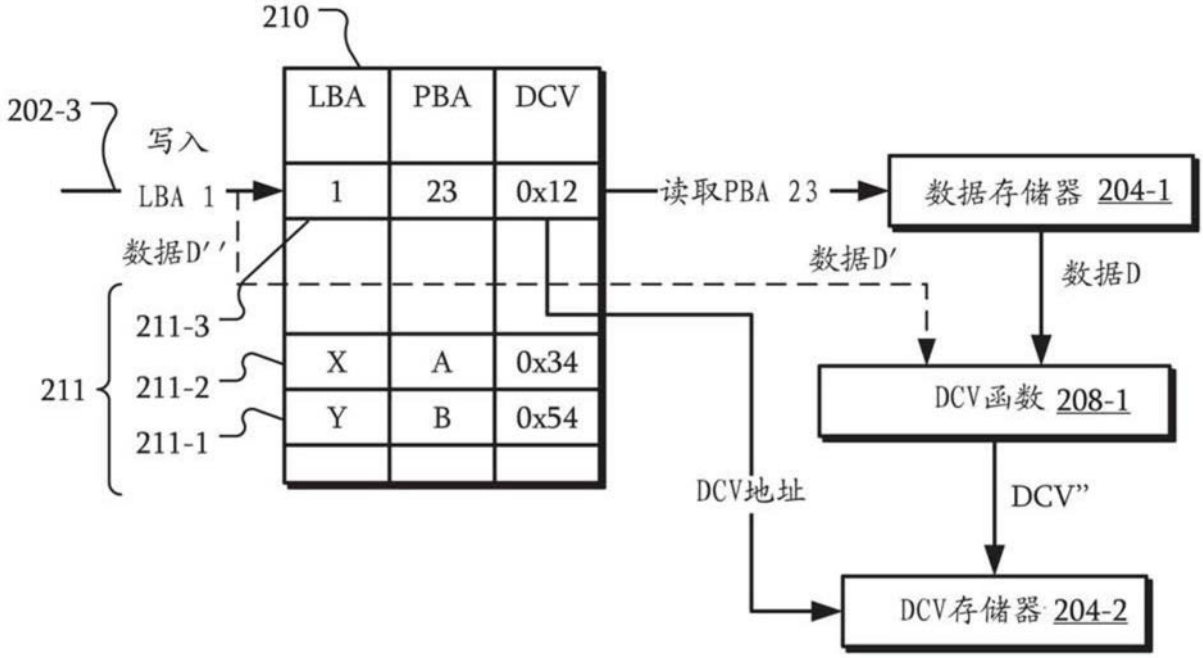

210
LBA
PBA
DCV
202-3
写入
LBA 1
1
23
0x12
读取PBA 23
数据存储器 204-1
数据D′′
数据D′
数据D
211-3
211
X
A
0x34
211-2
DCV函数 208-1
Y
B
0x54
211-1
DCV地址
DCV"
DCV存储器 204-2

图2C

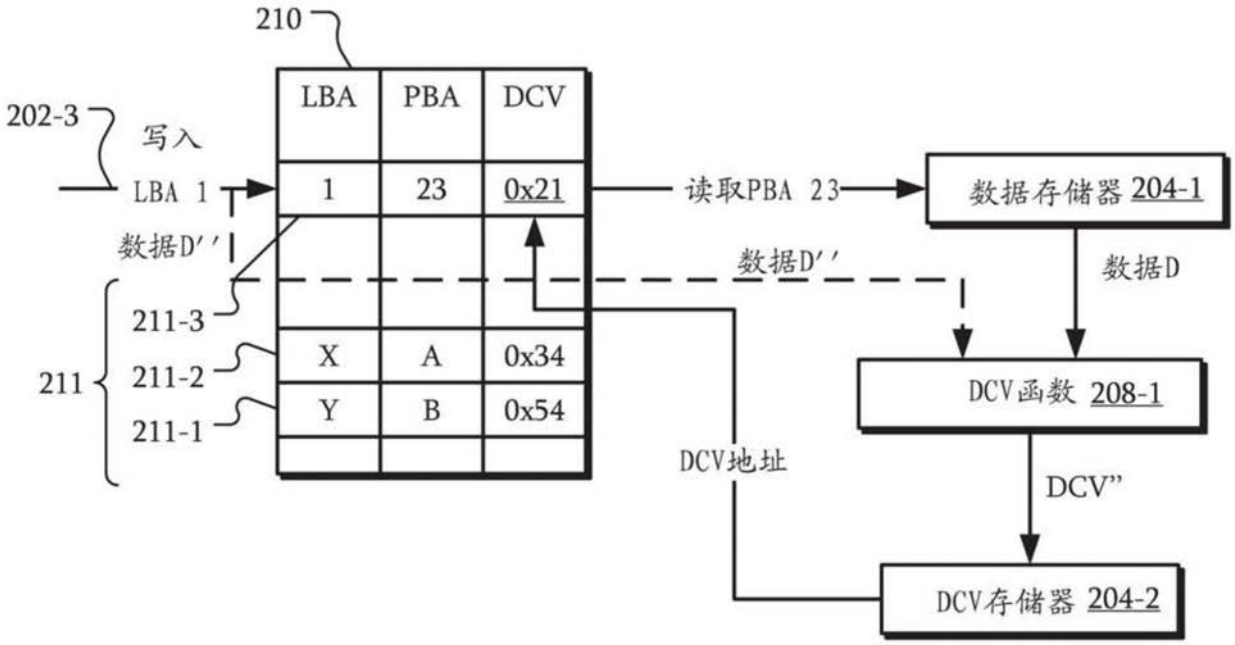

210
LBA
PBA
DCV
202-3
写入
LBA 1
1
23
0x21
读取PBA 23
数据存储器 204-1
数据D''
数据D''
数据D
211-3
211
211-2
X
A
0x34
211-1
Y
B
0x54
DCV函数 208-1
DCV地址
DCV"
DCV存储器 204-2

图2D

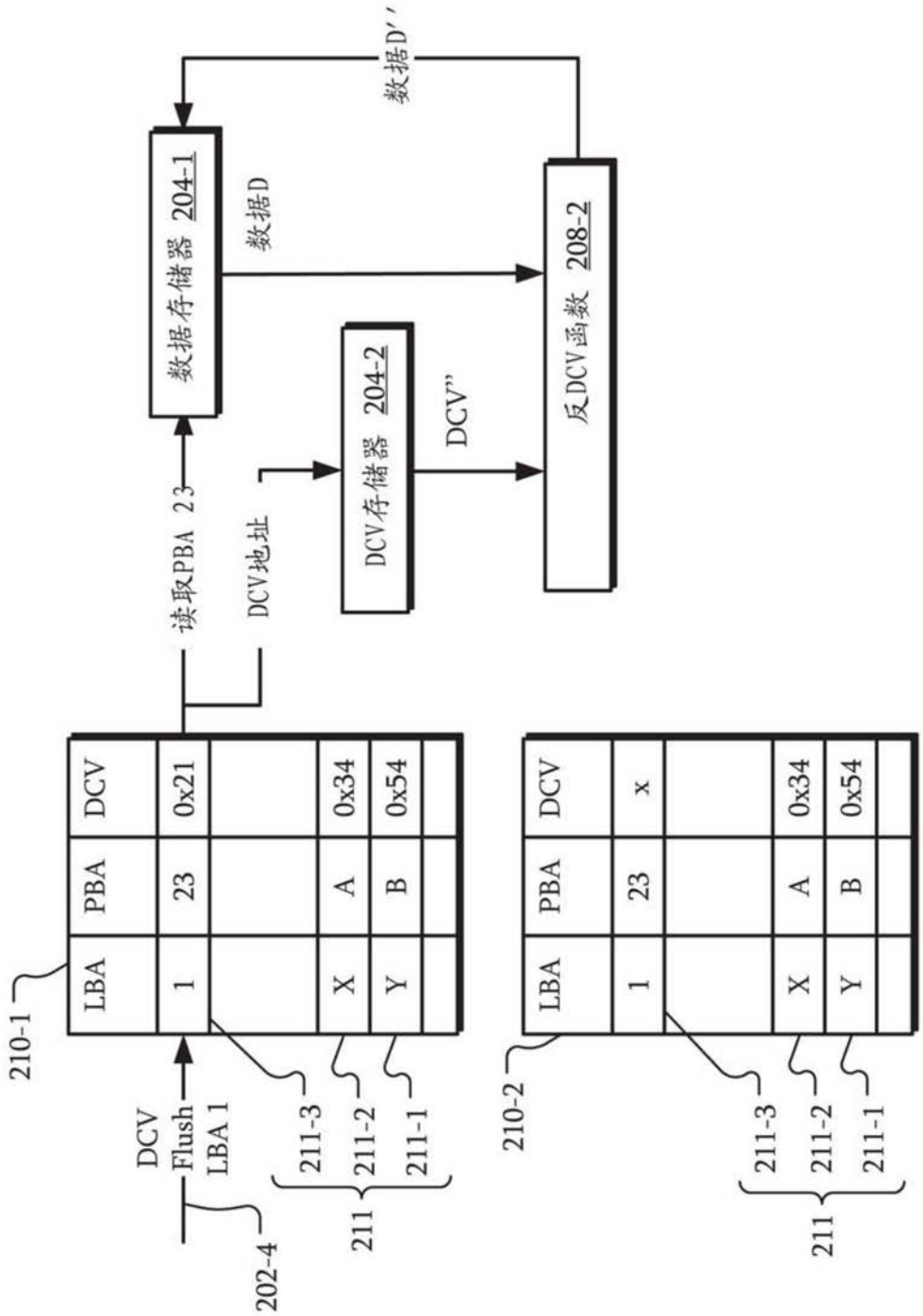

数据D''
数据存储器 204-1
数据D
反DCV函数 208-2
DCV存储器 204-2
DCV''
读取PBA 23
DCV地址
210-1
LBA
PBA
DCV
1
23
0x21
X
A
0x34
Y
B
0x54
DCV
Flush
LBA 1
202-4
211-3
211-2
211-1
211
210-2
x

图2E

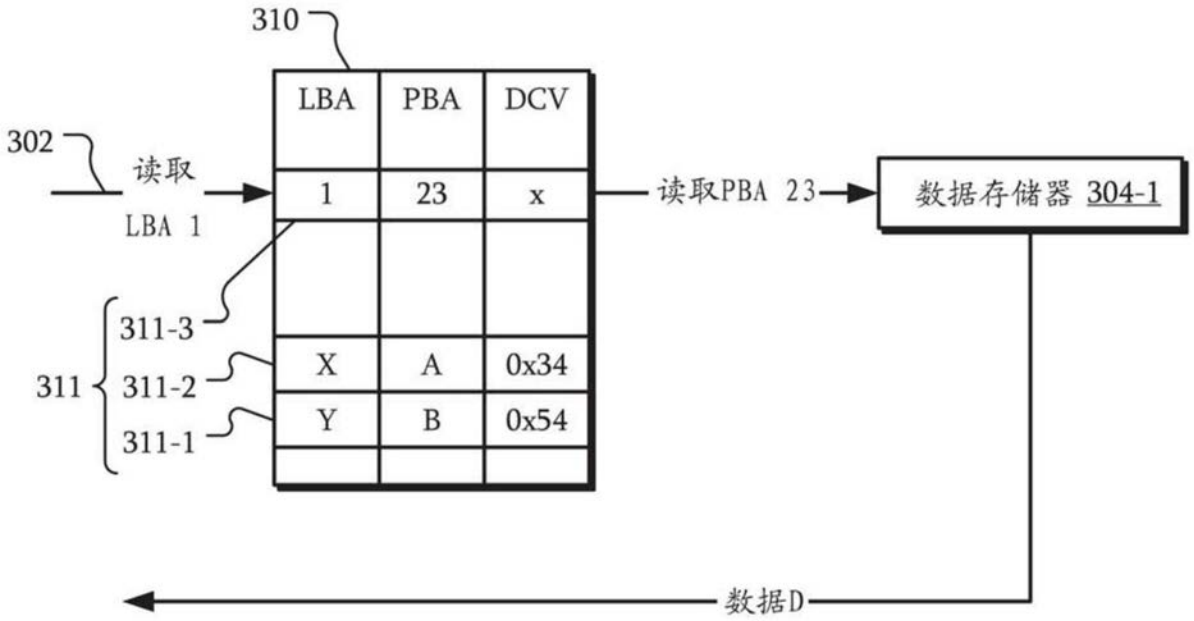

310
LBA
PBA
DCV
302
读取
LBA 1
1
23
x
读取PBA 23
数据存储器 304-1
311-3
311
311-2
311-1
X
A
0x34
Y
B
0x54
数据D

图3A

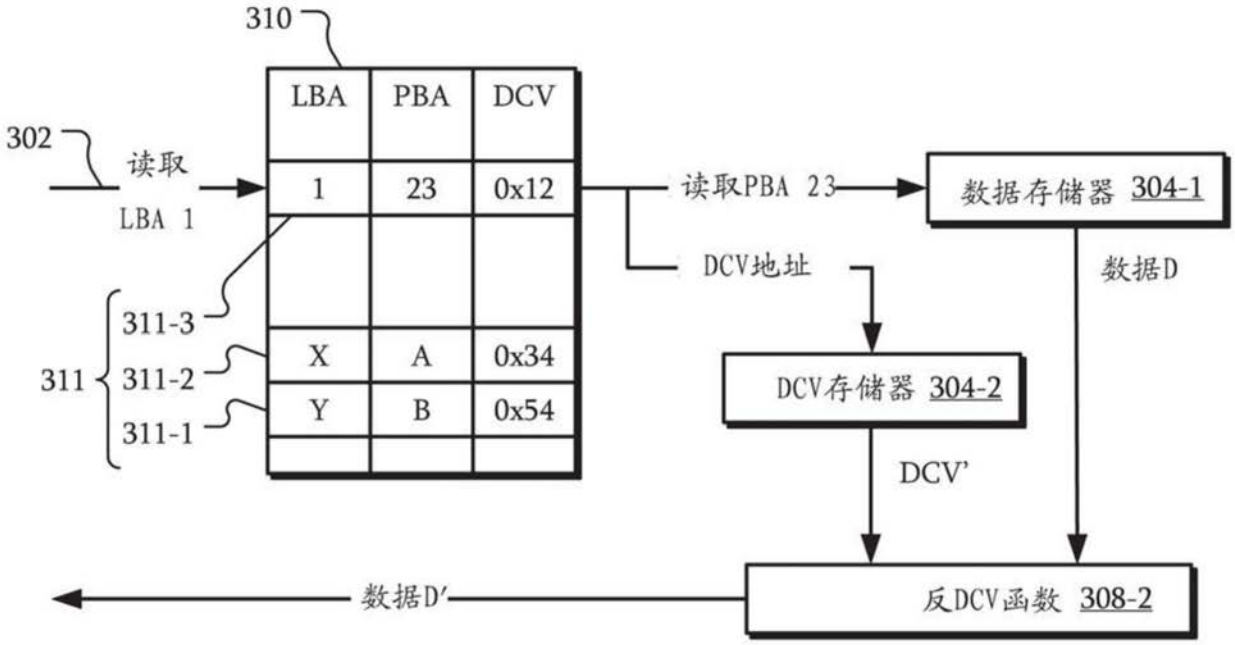

310
LBA
PBA
DCV
302
读取
LBA 1
1
23
0x12
读取PBA 23
数据存储器 304-1
DCV地址
数据D
311-3
311
311-2
311-1
X
A
0x34
Y
B
0x54
DCV存储器 304-2
DCV'
数据D'
反DCV函数 308-2

图3B

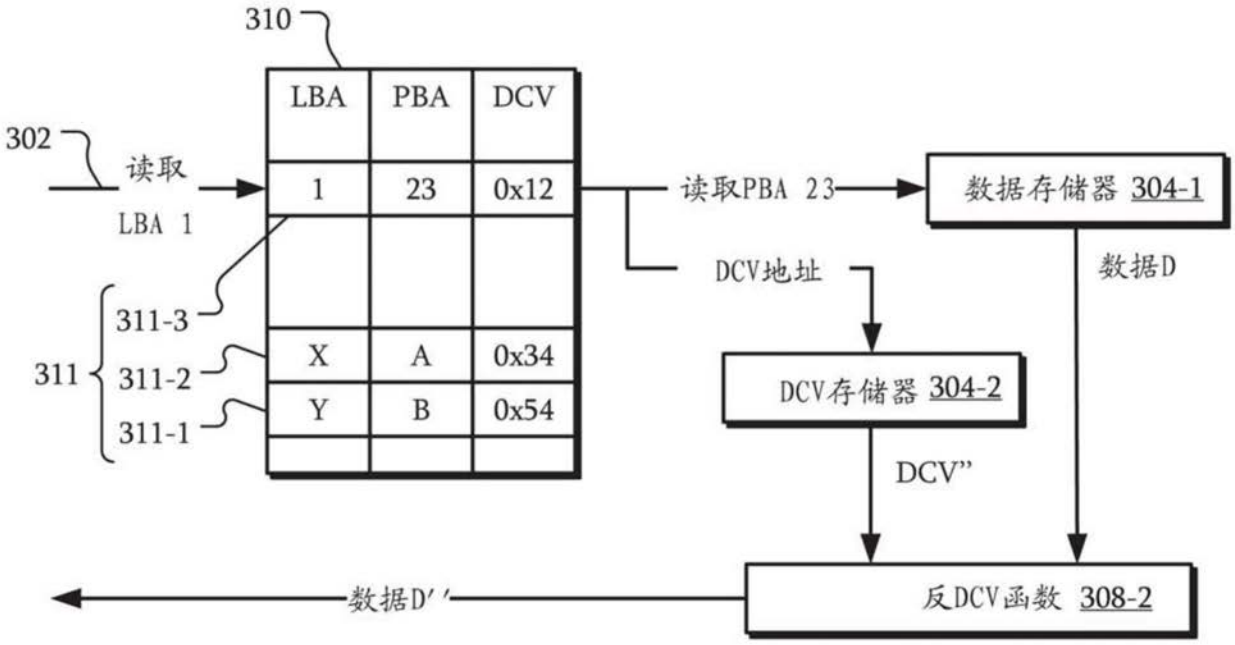

310
LBA
PBA
DCV
302
读取
LBA 1
1
23
0x12
读取PBA 23
数据存储器 304-1
DCV地址
数据D
311-3
311
311-2
311-1
X
A
0x34
Y
B
0x54
DCV存储器 304-2
DCV"
数据D''
反DCV函数 308-2

图3C

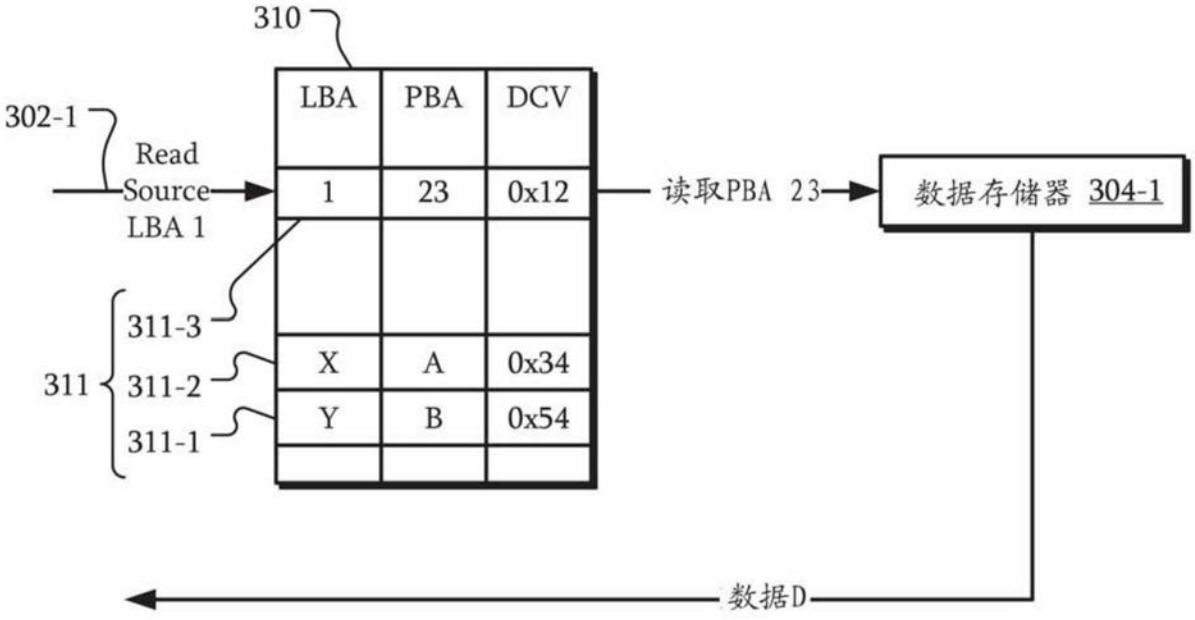

310
LBA
PBA
DCV
302-1
Read
Source
LBA 1
1
23
0x12
读取PBA 23
数据存储器 304-1
311-3
311
311-2
311-1
X
A
0x34
Y
B
0x54
数据D

图3D

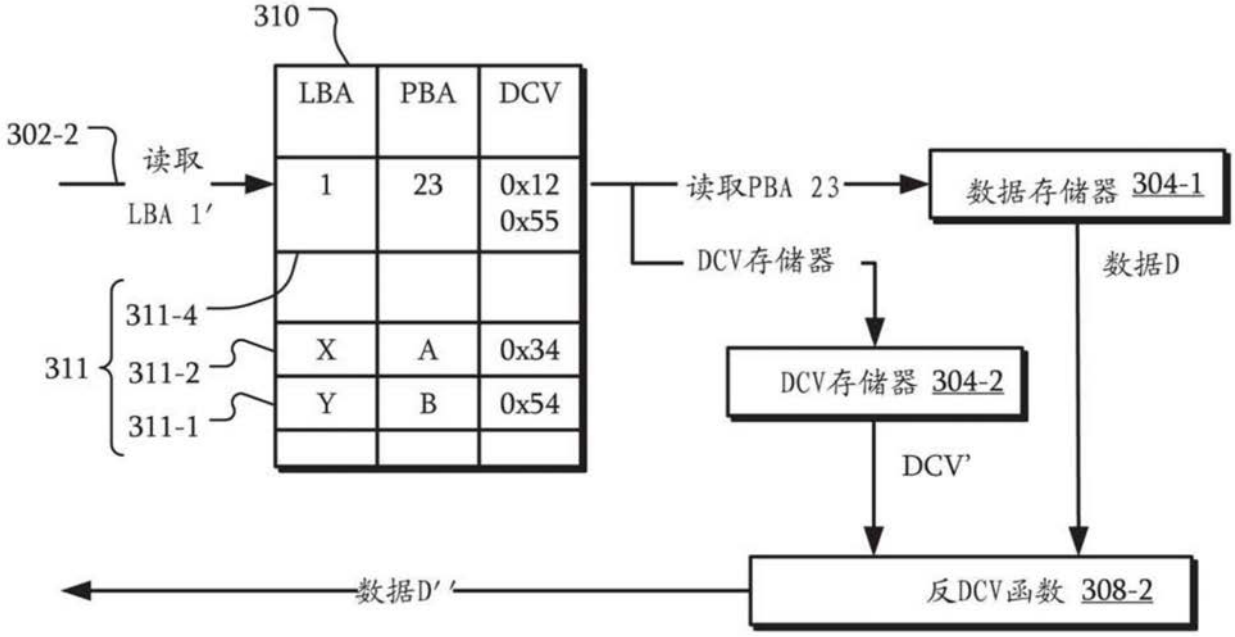

310
LBA
PBA
DCV
302-2
读取
LBA 1′
1
23
0x12
0x55
读取PBA 23
数据存储器 304-1
DCV存储器
数据D
311-4
311
311-2
311-1
X
A
0x34
Y
B
0x54
DCV存储器 304-2
DCV'
数据D′
反DCV函数 308-2

图3E

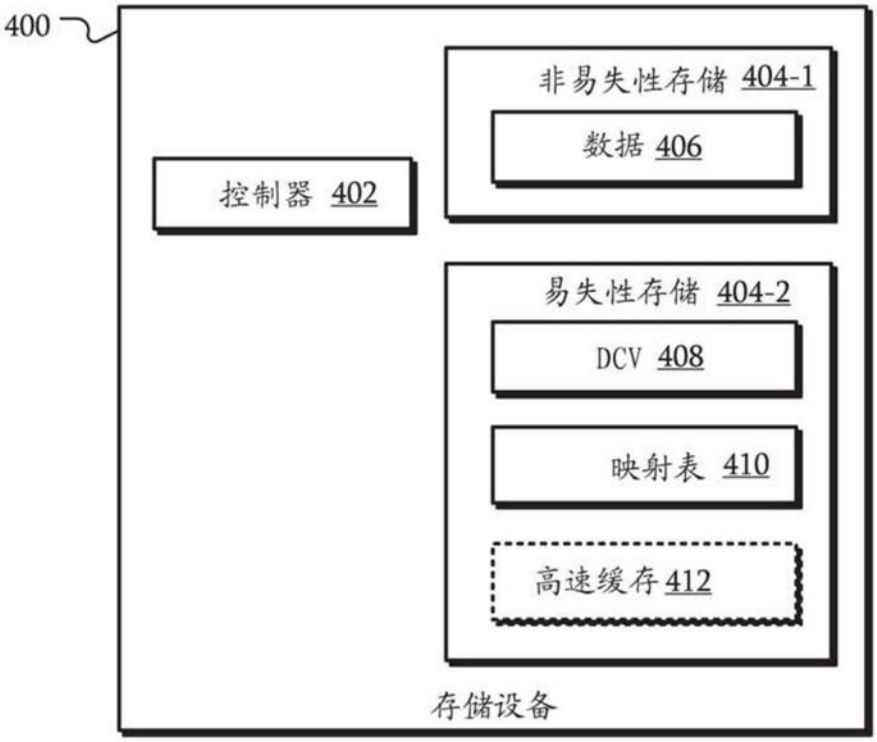

400
非易失性存储 404-1
数据 406
控制器 402
易失性存储 404-2
DCV 408
映射表 410
高速缓存 412
存储设备

图4A

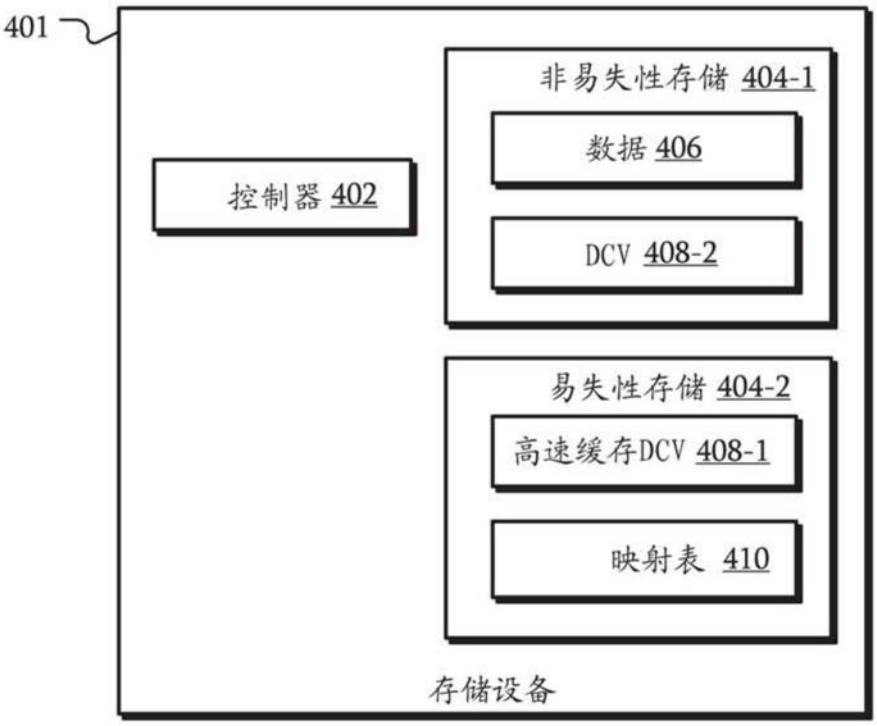

401
非易失性存储 404-1
数据 406
控制器 402
DCV 408-2
易失性存储 404-2
高速缓存DCV 408-1
映射表 410
存储设备

图4B

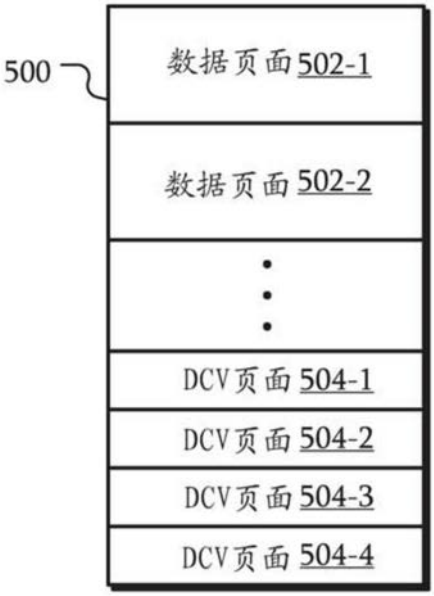

500
数据页面 502-1
数据页面 502-2
DCV页面 504-1
DCV页面 504-2
DCV页面 504-3
DCV页面 504-4

图5

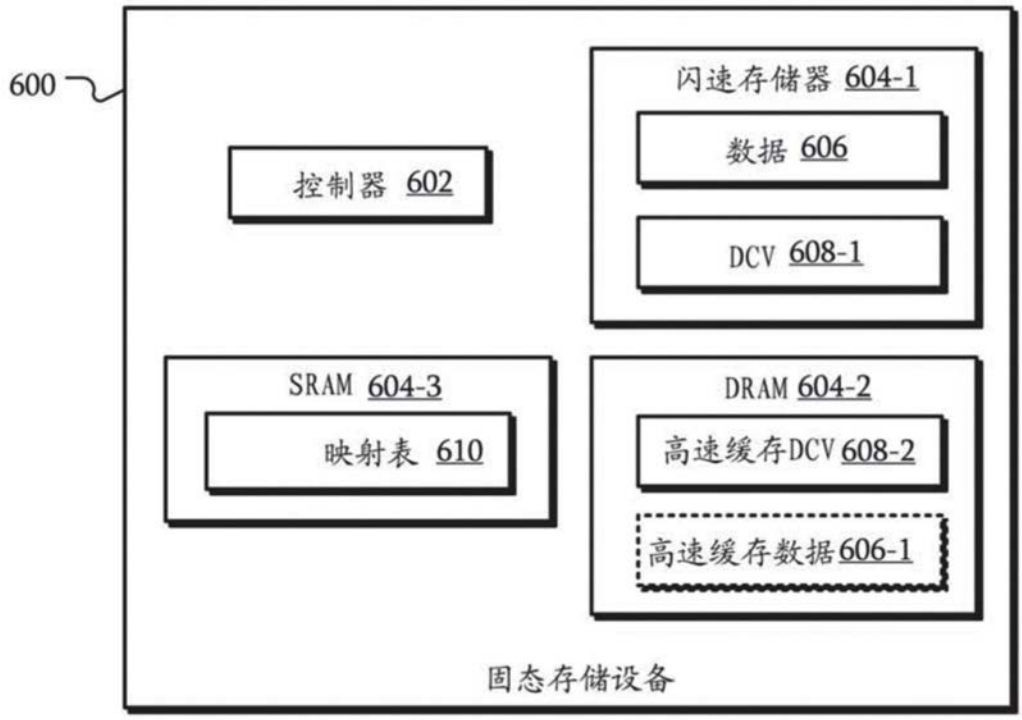

600
闪速存储器 604-1
数据 606
控制器 602
DCV 608-1
SRAM 604-3
映射表 610
DRAM 604-2
高速缓存DCV 608-2
高速缓存数据606-1
固态存储设备

图6

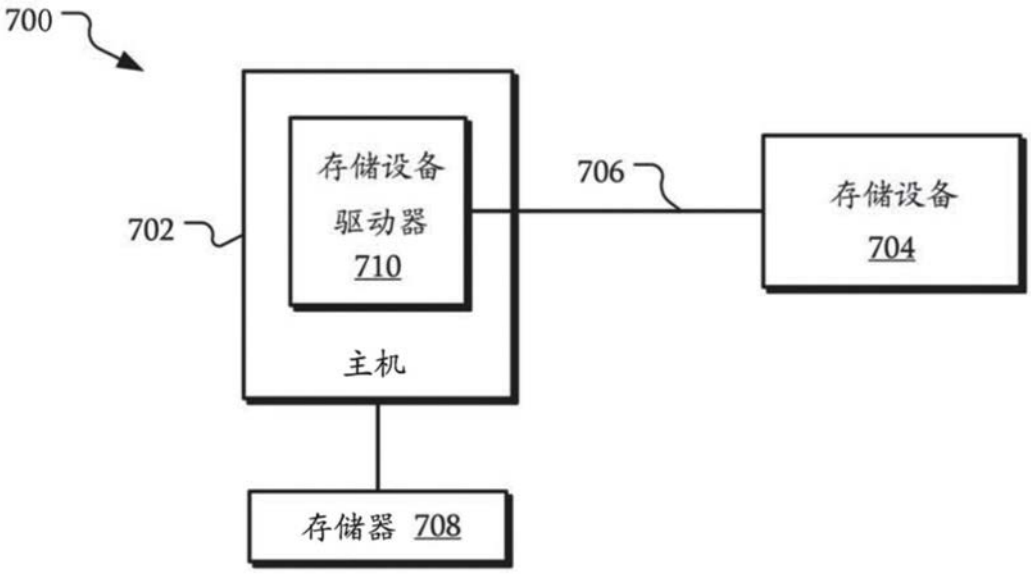

700
存储设备
驱动器
710
706
存储设备
704
702
主机
存储器 708

图7

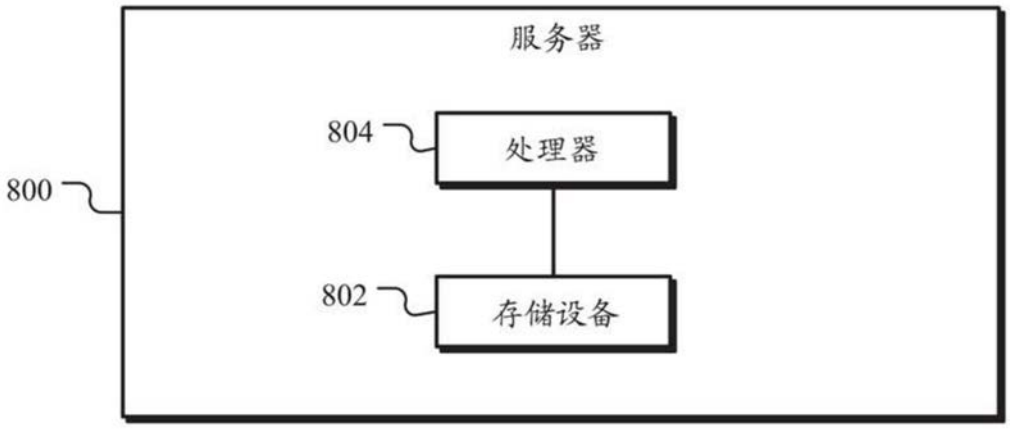

服务器
804
处理器
800
802
存储设备

图8

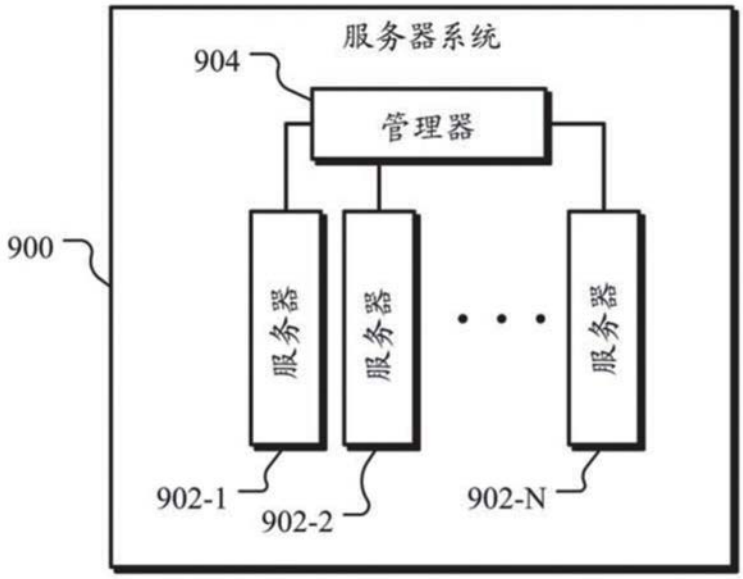

服务器系统
904
管理器
900
服务器
服务器
服务器
902-1
902-2
902-N

图9

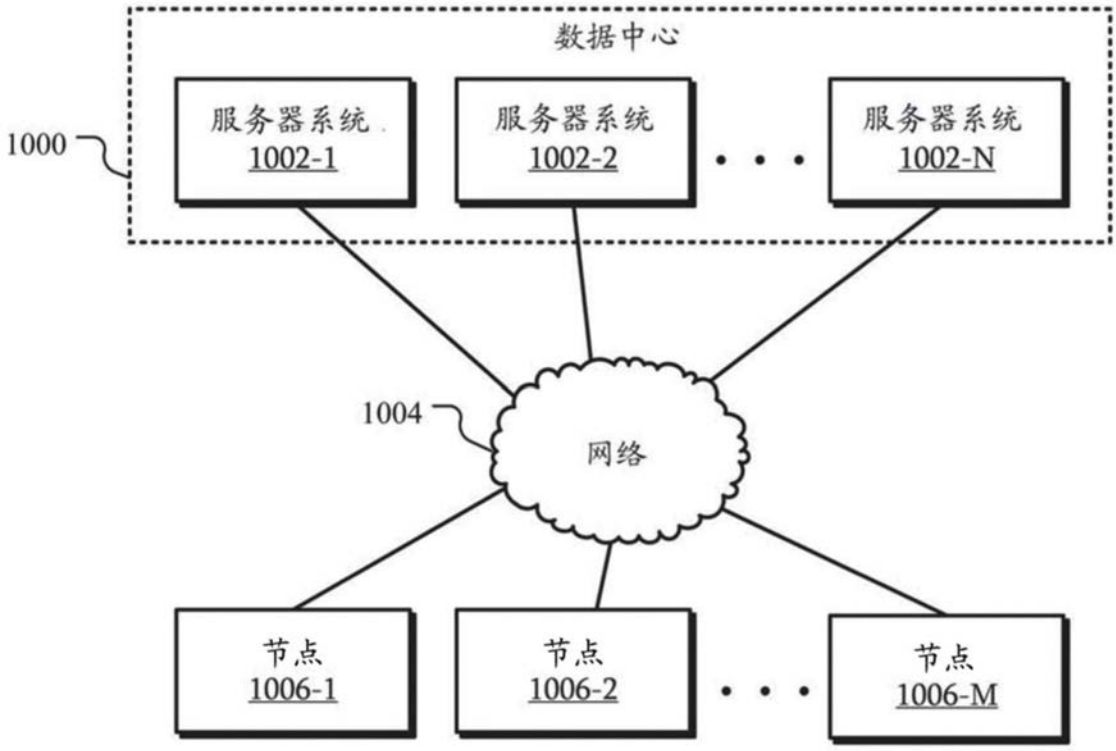

数据中心
1000
服务器系统
1002-1
服务器系统
1002-2
服务器系统
1002-N
1004
网络
节点
1006-1
节点
1006-2
节点
1006-M

图10