(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 108027713 B
(45)授权公告日 2021.10.12

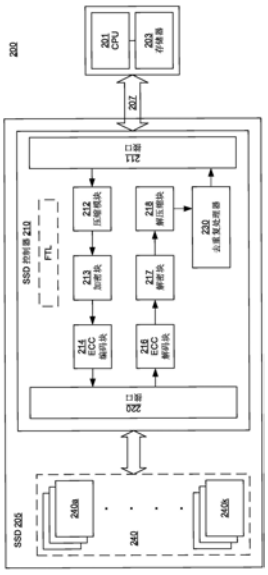
(21)申请号 201680054387.6
(22)申请日 2016.09.16
(65)同一申请的已公布的文献号
申请公布号 CN 108027713 A
(43)申请公布日 2018.05.11
(30)优先权数据
14/858257 2015.09.18 US
(85)PCT国际申请进入国家阶段日
2018.03.19
(86)PCT国际申请的申请数据
PCT/US2016/052222 2016.09.16
(87)PCT国际申请的公布数据
WO2017/049142 EN 2017.03.23
(73)专利权人 阿里巴巴集团控股有限公司
地址 英属开曼群岛大开曼资本大厦一座四层847号邮箱
(72)发明人 李舒 李勇 牛功彪
(74)专利代理机构 北京清源汇知识产权代理事务所(特殊普通合伙) 11644
代理人 冯德魁 窦晓慧
(51)Int.Cl.
G06F 3/06(2006.01)
(56)对比文件
US 2012173656 A1,2012.07.05
US 7761425 B1,2010.07.20
US 2010114833 A1,2010.05.06
US 2014304464 A1,2014.10.09
TW 201007452 A,2010.02.16
CN 102591947 A,2012.07.18
US 2014074804 A1,2014.03.13
US 2015074065 A1,2015.03.12
CN 101882141 A,2010.11.10
CN 103473266 A,2013.12.25
CN 103547991 A,2014.01.29
邢玉轩 等."一种基于历史信息的一致性Hash集群重复数据删除路由策略".《计算机研究与发展》.2014,(第2014年第S2期),
审查员 赵识谦
权利要求书2页 说明书9页 附图7页

(54)发明名称

用于固态驱动器控制器的重复数据删除

(57)摘要

一种由控制器执行的用于固态驱动器的重复数据删除方法。控制器接收一个数据块的签名。控制器执行签名与所述签名库的信息的比较,并确定签名是否匹配该信息。控制器发送指示比较结果的信号。如果签名和信息匹配,则所述信号具有指示数据块已经存储在SSD上的第一值;如果签名和信息不匹配,则信号具有与第一值不同的第二值。



CN 108027713 B

1.一种由控制器执行的用于固态驱动器SSD的重复数据删除方法,所述方法包括:

由所述控制器从中央处理单元CPU接收数据块的签名;

由所述控制器确定所述签名是否与签名库中的信息相匹配;以及

从所述控制器向所述CPU发送指示所述确定的结果的信号,其中,如果所述签名与所述签名库中的信息相匹配,则所述信号具有指示所述数据块已经存储在所述SSD上的第一值,并且如果所述签名与所述签名库中的信息不匹配,则所述信号具有不同于所述第一值的第二值;

其中,所述控制器为SSD控制器,所述SSD控制器包括去重复处理器,所述去重复处理器比较签名和所述签名库中的信息。

2.根据权利要求1所述的方法,其中,所述方法还包括:

由所述控制器从CPU接收与所述签名关联的地址;

所述控制器使用所述地址来索引在所述签名库中的条目;以及

由所述控制器确定所述签名是否与在所述签名库中的所述条目处的所述信息相匹配。

3.根据权利要求1所述的方法,其中,所述信号包括单个二进制比特,其中,如果所述签名与所述签名库中的信息相匹配,则所述比特是所述第一值,并且如果所述签名与所述签名库中的信息不匹配,则所述比特是所述第二值。

4.根据权利要求1所述的方法,其中,如果所述签名与所述签名库中的信息相匹配,则该方法还包括递增与所述数据块相关联的计数器。

5.根据权利要求1所述的方法,其中,如果所述签名与所述签名库中的信息不匹配,则所述方法还包括:

接收所述数据块并将所述数据块写入所述SSD;以及

将所述签名添加到所述签名库并递增与所述数据块相关联的计数器。

6.根据权利要求1所述的方法,其中,所述去重复处理器包括多个门,所述多个门执行逐比特比较所述签名与所述签名库中的信息。

7.根据权利要求1所述的方法,其中,所述SSD是串行高速技术附件SATA SSD。

8.一种由包括中央处理单元CPU和固态驱动器SSD的服务器执行的重复数据删除方法,所述SSD包括控制器,所述方法包括:

从客户端接收数据块的签名;

基于所述签名确定地址;

利用所述控制器并使用所述地址来定位签名库的条目;

利用所述控制器比较所述签名与在所述签名库中的所述条目处的信息;

利用所述控制器产生指示所述签名是否与所述信息匹配的信号,其中,如果所述签名和所述信息匹配,则所述数据块已经存储在所述SSD上,并且其中,如果所述签名和所述信息不匹配,则该方法还包括:

给客户端发信号;

从所述客户端接收响应于所述信号的所述数据块;和

将所述数据块写入所述SSD;

其中,所述控制器为SSD控制器,所述SSD控制器包括去重复处理器,所述去重复处理器比较签名和所述签名库中的信息。

9.根据权利要求8所述的方法,其中,所述信号包括单个二进制比特,其中,如果所述签名与所述信息相匹配,则所述比特是第一值,并且如果所述签名与所述信息不匹配,则所述比特是第二值。

10.根据权利要求8所述的方法,其中,如果所述签名和所述信息匹配,则所述方法进一步包括递增与所述数据块相关联的计数器。

11.根据权利要求8所述的方法,其中,如果所述签名和所述信息不匹配,则所述方法进一步包括将所述签名添加到所述签名库,并递增与所述数据块相关联的计数器。

12.根据权利要求8所述的方法,其中,所述控制器包括多个异或XOR门,多个门中的每一个门将来自所述签名的比特值和来自所述签名库中的所述信息的每一比特值进行比较。

13.根据权利要求8所述的方法,其中,所述SSD是串行高速技术附件SATA SSD。

14.一种固态驱动器SSD,其包括:

多个存储元件;和

联接到所述存储元件的控制器;

所述控制器从中央处理单元CPU可操作地接收用于数据块的签名和基于所述签名的地址,其中,由所述CPU从所述数据块的源接收所述签名;

所述控制器可操作以使用所述地址来定位签名库中的条目,用于比较所述签名与所述签名库中的所述条目的信息,并且用于生成指示所述签名是否与所述信息匹配的信号并用于将所述信号发送给所述CPU;

该控制器进一步可操作用于,在响应于指示所述签名与所述签名库中的信息不匹配的信号接收到所述数据块时,将所述数据块写入至所述多个存储元件中的一个存储元件;

其中,所述控制器为SSD控制器,所述SSD控制器包括去重复处理器,所述去重复处理器比较签名和所述签名库中的信息。

15.根据权利要求14所述的SSD,其中,所述信号包括单个二进制比特,其中,如果所述签名与所述信息相匹配,则所述比特是第一值,并且如果所述签名与所述信息不匹配,则所述比特是第二值。

16.根据权利要求14所述的SSD,其中,如果所述签名和所述信息匹配,则与所述数据块相关联的计数器递增。

17.根据权利要求14所述的SSD,其中,如果所述签名和所述信息不匹配,则所述签名被添加到所述签名库,并且与所述数据块相关联的计数器递增。

18.根据权利要求14所述的SSD,其中,所述去重复处理器包括多个异或XOR门,多个门中的每一个门将来自所述签名的比特值和来自所述签名库中的所述信息的每一比特值进行比较。

19.根据权利要求14所述的SSD,其中,所述SSD是串行高速技术附件SATA SSD,并且所述存储元件包括NAND管芯。

用于固态驱动器控制器的重复数据删除

技术领域

[0001] 本发明涉及存储技术领域，具体涉及用于固态驱动器SSD的重复数据删除方法。本发明同时涉及一种固态驱动器SSD。

背景技术

[0002] 当代商业在数据库中积累了巨大的数据(例如，PB，petabyte)，这些数据库存储于诸如磁带，硬盘驱动器，固态驱动器(SSD)等各种介质中。法律规定，政府规则和规章，商业规则及最佳实践需要数据库经常存档并备份。结果，数千PB的数据已被存储，且存储的数据量不断激增。

[0003] 重复数据删除方法和系统用于减少数据量以提高效率并降低成本。通常，重复数据删除系统会在不同的数据文件中找到相同的部分，并将这些相同的部分仅存储一次。重复数据删除系统也维护元数据，以便数据文件可以在稍后访问时组织和重建。然而，大量的数据被存储以检验现有重复数据删除方法和系统的极限。现有的重复数据删除方法和系统适用于数PB的数据，但并不适用于数千PB量级数据量的数据。

[0004] 使用固态硬盘SSD(替代其它存储介质，例如随机存取存储器(RAM))来存储大量数据带来了一些挑战。SSD相对于例如双数据速率三型同步动态随机存储器(double data rate type three synchronous dynamic RAM，DDR3 DRAM)有较长的读取和写入延迟。并且，SSD在写入之前会被擦除，并且在耗尽之前只能擦除有限的次数。

[0005] 在另一方面，固态硬盘具有很多优点，使它们成为存储大量数据的不错选择。对于重复数据删除，文件被分割成通常称为与元数据相关联的“块(chunks)”(例如，4千字节(KB)，16KB或256KB块)的块(blocks)或片段。每个独有的块都与其元数据一起存储。元数据的大小例如可以是16字节(B)，32B，128B或256B。对于512PB数据，假设每个块的大小为16KB，并假定每块的元数据为32B，元数据的存储空间为一个PB。使用RAM来存储这些数据是不实际的，但使用SSD是有效的。

[0006] 而且，为了满足上述法规和要求，元数据需要被硬写入到存储器中。当RAM的电源断电或中断时，RAM所保存的数据将丢失。SSD使用基于NAND的闪速存储器，例如，其无需电源保持数据。

[0007] 因此，SSD的优势包括其容量和非易失性。为了缓解其较长的访问时间(读取和写入延迟)，数据并行写入。每个SSD读/写操作的基本单位被称为页面。对于16KB的页面大小，假设每块大小为128B的元数据，则可以在页面内并行读取或写入128个块的元数据。

[0008] 针对每块中的元数据包括唯一标识该块的哈希值或签名。因此，为了确定是否有必要存储新的块(以确定先前是否存储了相同的块)，可以将新块的签名与已经存储的块的签名进行比较。如果新块的签名与现有签名相匹配，则新的块无需被存储。

[0009] 如上所述，一SSD读/写操作的基本单元是页面。为获得一块的签名与其它签名进行比较，整个页面(例如，16K字节)被读出并由SSD传输至中央处理单元(CPU)。该传输会占用CPU以及内存带宽和总线带宽的大量资源。

[0010] 更具体地，具有要存储在存储服务器上的数据的客户端将数据分成块并计算每个块的签名。在一实现中，客户端将每个签名发送到签名服务器，该签名服务器包含已存储在存储服务器上的块的签名库。签名服务器的作用是确定来自客户端的签名是否与签名库中的任何签名相匹配。为了达到此目的，针对每个签名将整个页面(例如，16KB)传输到内存中，并且CPU将在页面内定位和提取签名并将提取的签名与来自客户端的签名进行比较。但是，签名可能只有32B的大小。因此，要获得签名与其它签名进行比较，高达所需数据500倍的数据被读出并被传输(例如，16KB的数据被读取以获得32B签名)。

[0011] 此外，基于请求签名比较的客户端的数量和签名服务器的数量，可以估计每个签名服务器的比较数量。每个比较至少需要两次输入/输出(I/O)访问，因此，还可以估计每个签名服务器的每秒I/O操作数(IOPS)。考虑到CPU和SSD功能，IOPS要求变得非常大，需要大量的签名服务器，而且还需要使用更昂贵的、更高带宽的快速外设组件互连标准(PCIe)SSD来提供必要的容量。

[0012] 总之，传统的重复数据删除的方法是低效的、昂贵的，并且占据大量的CPU、存储器和总线的资源。

发明内容

[0013] 根据本发明的实施例，通过在SSD中完成签名比较来解决上述问题。这样可以减少CPU的工作负载，并减少SSD与CPU之间传输的数据量，从而减少对内存和总线带宽的影响。

[0014] 在根据本发明的实施例中，由SSD的控制器执行重复数据删除方法。所述控制器接收来自CPU的块或片段数据的第一签名(“块”)。控制器执行第一签名和签名库中的信息的比较，并确定第一签名是否匹配该信息。控制器向CPU发送指示比较结果的信号。如果第一签名与签名库中的信息匹配，则该信号具有第一值，该第一值指示该块数据已经存储于SSD;如果第一签名与签名库中的信息不匹配，则该信号具有与第一值不同的第二值。如果第一签名与签名库中的信息不匹配(例如，如果信号具有第二值)，则从其源(例如，客户端)接收数据块并将其写入SSD。

[0015] 在一个实施例中，控制器还从CPU接收与第一签名相关联的地址。在这样的实施例中，控制器使用该地址来定位签名库中的条目。然后控制器将签名与该条目中的信息进行比较。该信息可能是第二个签名，也可能是空值(例如，该条目可能不包含任何信息)。控制器然后向CPU发送信号以指示上述比较结果。

[0016] 在一个实施例中，由SSD控制器向CPU的信号由单个二进制比特组成，该单个二进制比特在第一签名与签名库中的信息匹配时具有第一值，在第一签名与签名库中的信息不匹配时具有第二值。

[0017] 在一个实施例中，如果第一签名与签名库中的信息匹配，则与该数据块相关联的计数器递增。如果第一签名与签名库中的信息不匹配，则第一签名被添加到签名库，并且与该数据块相关联的计数器被初始化并递增。

[0018] 在一个实施例中，除了常规组件之外，控制器还包括比较第一签名和来自签名库的信息的去重复处理器。在一个这样的实施例中，去重复处理器包括多个门，它们执行对第一签名和来自签名库的信息的逐比特比较。

[0019] 在一个实施例中，SSD是串行高级技术附件(Serial Advanced Technology

Attachment,SATA)SSD。

[0020] 根据本发明的实施例可以用于大规模数据应用中的高效重复数据删除,其中存储的数据量以EB(Exabyte,EB)为单位进行度量。需要更少的IOPS,因此根据本发明的实施例可以使用较便宜的SATA SSD来实现。

[0021] 本领域的普通技术人员在阅读各个附图中所示的实施例的以下详细描述之后将认识到本发明的各种实施例的这些和其它目的和优点。

附图说明

[0022] 附图被并入本说明书并形成本说明书的一部分,并且其中相同标号表示相同元件,附图示出了本公开的实施例,并且与详细描述一起用于解释本公开的原理。

[0023] 图1是示出在根据本发明的实施例中去重复系统可被实现的系统的示例的框图。

[0024] 图2是示出根据本发明的实施例中的计算机系统的元件的框图。

[0025] 图3是示出根据本发明的实施例中的SATA SSD和中央处理单元之间的连接的框图。

[0026] 图4是根据本发明的实施例中的去重复方法的流程图。

[0027] 图5示出了可用于实现根据本发明的实施例中的去重复处理器的电路。

[0028] 图6是根据本发明实施例的由控制器为SSD执行的重复数据删除方法的流程图。

[0029] 图7为根据本发明实施例中由包括CPU和SSD的签名服务器执行的重复数据删除方法的流程图。

[0030] 图8是示出在根据本发明一个实施例中去重复系统的实现方式的示例的框图。

具体实施方式

[0031] 现在将详细说明本公开的各种实施例,其示例在附图中示出。在结合这些实施例描述时,但应当理解的是,它们并非意图将本公开限制于这些实施例。相反,本公开旨在覆盖可包括在由所附权利要求限定的本公开的精神和范围内的替代、修改和等同物。此外,在本公开的以下详细描述中,阐述了许多具体细节以提供对本公开的透彻理解。然而,将理解的是,可以在没有这些具体细节的情况下实践本公开。在其他情况下,没有详细描述公知的方法、过程、组件和电路,以免不必要地模糊本公开的各方面。

[0032] 下面的详细描述的一些部分是按照程序、逻辑块、处理以及计算机存储器内的数据比特的操作的其他符号表示来呈现的。这些描述和表示是数据处理领域的技术人员用来将其工作的实质最有效地传达给本领域其他技术人员的手段。在本申请中,程序、逻辑块、处理等被认为是导致期望结果的步骤或指令的自洽序列。这些步骤是利用物理量的物理操作的步骤。通常,虽然不一定,但这些量采取能够在计算机系统中存储、传输、组合、比较和以其他方式操作的电或磁信号的形式。主要由于普遍使用的原因,将这些信号称为事务、比特、值、元素、符号、字符、样本、像素等已经被证明有时是方便的。

[0033] 但应该记住,所有这些和类似的术语都与适当的物理量相关联,并且仅仅是适用于这些量的便利标签。除非从下面的讨论特别说明,否则如显而易见的,应理解,在整个本公开中,讨论使用诸如“接收”、“存储”、“读”、“写”、“索引”、“执行”、“发送”、“比较”、“添加”、“访问”、“定位”、“使用”、“确定”、“生成”、“递增”、“信令”、或类似术语是装置或计算机

系统或类似电子计算设备或处理器(例如,计算机系统的图2的200)的指动作和过程(例如,流程图图4、6和7中400,600,700)。计算机系统或类似的电子计算设备操作并转换在存储器、寄存器的表示为物理(电子)量数据或其他这样的存储、传输或显示设备内的信息。

[0034] 这里描述的实施例可以在驻留在某种形式的计算机可读存储介质(例如程序模块)上的计算机可执行指令的一般场景中讨论,计算机可执行指令由一个或多个计算机或其他设备执行。作为示例而非限制,计算机可读存储介质可以包括非暂时性计算机存储介质和通信介质。通常,程序模块包括执行特定任务或实现特定抽象数据类型的例程、程序、对象、组件、数据结构等。程序模块的功能可以根据需要在各种实施例中组合或分配。

[0035] 计算机存储介质包括在任何方法或技术实现的、用于存储诸如计算机可读指令、数据结构、程序模块或其他数据的易失性和非易失性、可移动和不可移动介质。计算机存储介质包括但不限于随机存取存储器(RAM),只读存储器(ROM),电可擦除可编程ROM(EEPROM),闪存(例如SSD)或其他存储器技术,光盘ROM(CD-ROM),数字通用盘(DVD)或其他光存储器,盒式磁带,磁带,磁盘存储器或其他磁存储设备或可用于存储所需信息并可被访问以取回信息的介质。

[0036] 通信介质可以收录计算机可执行指令、数据结构和程序模块,并且包括任何信息传递介质。作为示例而非限制,通信媒体包括诸如有线网络或直接有线连接的有线介质,以及诸如声学,射频(RF),红外和其它无线介质的无线介质。上述任何组合也可以被包括在计算机可读介质的范围内。

[0037] 图1是示出在根据本发明的实施例中可以实现去重复系统的网络或系统100的示例的框图。在图1的示例中,系统100包括:多个客户端101a、101b、101c、...、101m,其可以被单独地称为客户端101m或统称为客户端101;一个或多个配置服务器102;以及可被单独称为签名服务器103n并统称为签名服务器103的多个签名服务器103a、103b 103c...103n。客户端101、签名服务器103和配置服务器102经由网络104(例如,因特网,但并不局限于此)通信地联接(有线或无线)。

[0038] 客户端101本身可以是服务器。通常,客户端101具有它们已经生成的、或者由其它设备或系统(未示出)接收的数据文件。客户端101m将其具有的数据文件分割成小块(例如,块或片段,通常被称为块)。对于每个这样的块,客户端101m计算签名SCP,它唯一地标识块,并将签名发送到签名服务器103。在一个实施例中,来自客户端101m的签名以批处理模式发送;因此可以将多个签名(例如,数百个签名)分组为单个数据分组并发送至签名服务器103以供比较。

[0039] 在一个实施例中,块的签名是通过将散列函数应用于块的内容来计算的。在一个实施例中,元数据与每个块相关联,并且块的签名被包括在用于该块的元数据中。块的元数据可以包括除了块的签名之外的信息。例如,元数据可以包含这样的信息,即,该信息可以用来将数据块和与数据文件关联的其它块组合起来,以重构数据文件。例如,除了签名之外,元数据还可以包含地址,该地址指向存储块的数据所在的位置,以及附录,用来标识该数据块如何(例如,按照什么顺序)与其他块组合以重构数据文件。

[0040] 配置服务器102调度和控制客户端101和签名服务器103之间的通信。配置服务器102将来自客户端101的签名引导到适当的签名服务器。

[0041] 签名服务器103n接收配置服务器102引导的、来自客户端101的签名,在其签名库

查找条目(内容或信息,例如,签名),比较客户的签名和签名库中的信息,针对每一个其接收到的签名Scp,通知客户关于签名Scp是否匹配其签名库中的信息。

[0042] 如前所述,去重复包括通过比较已经被存储的块的签名与新块的签名,来确定与新块相同的块是否先前已被存储,进而确定是否有必要存储新块。如果新块的签名与现有签名匹配,则新块不需要被存储。

[0043] 图2是示出根据本发明的实施例中的计算机系统200的元件的框图。在一个实施例中,计算机系统200表示用于实现签名服务器103(图1)的平台。在图2的示例中,计算机系统200包括中央处理单元(CPU)201,存储器203和固态驱动器(SSD)205。存储器203可以是例如动态随机存取存储器(DRAM)。在一个实施例中,SSD 205是SATA SSD,并且经由SATA总线207联接到CPU 201。计算机系统200可以包括除了所示的那些之外的元件。

[0044] 图3是示出根据本发明的实施例中的SATA SSD 301、302和303与CPU 201之间的连接的框图。SSD 301-303分别经由SATA总线311、312和313联接到主机总线适配器(HBA)320。在一个实施例中,HBA适配器320包括SATA接口321,高级主机控制器接口(AHCI)引擎322(例如,允许软件与SATA设备通信的硬件机制;总线主控器到系统存储器),PCIe到AHCI桥接器323(例如,用于在SATA和PCIe格式之间转换数据)以及PCIe接口324。HBA 320经由PCIe总线330联接到CPU 201。

[0045] 再次参考图2,SSD 205包括控制器210和多个存储元件,具体地为用于存储数据的多个管芯(die)或芯片(chip)240a-240k。管芯240a-240k可以被单独地称为管芯240k并统称为管芯240。在一个实施例中,管芯240是NAND管芯,因此SSD 205可以被称为NAND闪存器件。

[0046] 控制器210可以被实现为被嵌入在SSD 205的专用集成电路(application-specific integrated circuit,ASIC)或现场可编程门阵列(field-programmable gate array,FPGA)。在图2的实施例中,控制器210包括闪存转换层(FTL),其可以实现为固件或软件。控制器210还包括写入路径和读取路径。写入路径开始于接口211,其包括例如物理层(PHY)接口,和在模拟域和数字域(从模拟到数字,以及从数字到模拟)之间转换数据的串行器/解串器。写路径可以包括一个数据压缩模块212、加密块213以及纠错码(error correction code,ECC)编码块214。SSD控制器通过接口220(例如,开放NAND接口,ONFI)联接到所述管芯240。使用同步和异步触发模式(切换)将数据移动到管芯240。

[0047] 数据通过相同的触发机制和接口220从管芯240移动到读取路径。读取路径可以包括ECC解码块216、解密块217和解压缩块218。

[0048] 重要的是,与传统SSD控制器相比,控制器210在读取路径中包括去重复处理器230。如将在下面更充分地描述的,去重复处理器230执行至少两个主要功能:它从管芯240提取信息(例如,签名),并且将这些签名与从客户端101m接收的签名进行比较,以确定来自客户端的该签名与提取的信息是否匹配(例如,签名是否匹配任何提取的签名)。

[0049] 图4是根据本发明的实施例中的去重复方法的流程图400。由流程图400中的框代表的所有或一些操作可以实现为驻留在某种形式的非暂时性计算机可读存储介质上的计算机可执行指令,并且由签名服务器或计算机系统执行,诸如签名服务器103n或图1和2的计算机系统200。

[0050] 在图4的框402中,从客户端(例如,图1的客户端101m)接收针对数据的块(例如,

块)的签名Scp。

[0051] 在图4的框404中,在一个实施例中,基于签名Scp来计算读取地址(Raddr)。

[0052] 在框406中,签名Scp被发送到SSD控制器210(图2)。在一个实施例中,读取地址Raddr也被发送到SSD控制器210。在这样的实施例中,控制器使用读取地址Raddr来定位存储在管芯240(图2)上的签名库中的条目。更具体地说,控制器210可以使用读地址Raddr作为存储在其中一个管芯240(例如,管芯240k)的签名库中的特定条目的索引。条目可以是签名,也可以是空值(例如,条目可以是空的)。一般来说,在由读地址Raddr索引的条目中有内容或信息Rssd;该内容可能是也可能不是签名,甚至由缺乏内容(例如,无效或空白条目)传递信息。

[0053] 在图4的框408中,SSD控制器210比较签名Scp与信息Rssd。在一个实施例中,所述比较使用去重复处理器230(图2)执行。将在下面的图5中提供附加的信息。

[0054] 如果签名Scp匹配信息Rssd,则流程图400进入框410;否则,流程图进入框414。

[0055] 在图4的框410中,签名Scp匹配信息Rssd。在这种情况下时,信息Rssd构成与第一签名Scp相同的第二签名,这表明与签名相关联的Scp已经被存储在SSD 205,没有必要重写数据块到SSD。因此,控制器210将信号发送到CPU 201。在一个实施例中,信号由单个二进制比特组成。该比特值具有第一值(例如二进制1或高)以指示签名Scp匹配信息Rssd。

[0056] 继续参照图4,在框412中,与签名Scp相关联的元数据被更新并且映射计数被递增。元数据被用于恢复或帮助恢复数据文件,该数据文件包括与所述签名Scp相关联的块。映射计数标识该块使用了多少次(例如,有多少数据文件包含该块)。

[0057] 在框414中,签名Scp与信息Rssd不匹配,这表明与签名Scp相关联的数据块当前未被存储在SSD 205上,并且,因此可以将数据块写入SSD。如框410中那样,控制器210向CPU 201发送信号。在一个实施例中,所述信号同样包含单个二进制位。然而,比特值具有第二值(例如,二进制零或低)以指示签名Scp与信息Rssd不匹配。

[0058] 在框416,控制器210将签名Scp添加到签名库。在一个实施例中,控制器210将签名Scp添加到由读取地址Raddr标识的签名库中的条目中。

[0059] 在框418中,数据块被写入SSD 205。在一个实施例中,数据块被写入存储由读取地址Raddr索引的签名库的管芯240k。而且,通过控制器210,与签名Scp相关联的元数据被更新并且该块的映射计数开始(递增)。控制器210还可以确认该签名库已更新,并且确认该块被保存。

[0060] 在框420中,如果存在针对另一个数据块的另一个签名,则流程图400返回到框402。

[0061] 根据本发明的实施例不限于使用读取地址Raddr来找到用于与签名Scp进行比较的条目,该比较用于确定该数据块是否已经存储在SSD 205上。其他技术可以用于将签名Scp与签名库中的信息进行比较。一种这样的技术被称为布谷鸟搜索算法。其他被称为粒子群优化、微分进化和人工蜂群算法的技术也可以使用。

[0062] 如上所述,来自客户端101的签名可以以批处理模式发送;因此可以将多个签名(例如,数百个签名)分组为单个数据分组并发送到签名服务器103,在这种情况下,可以针对每个签名并行执行流程图400的操作。在一个实施例中,这可以通过使用SSD和CPU之间的多比特总线来实现,其中所述总线的每个比特表示是否单个签名已经与如上所述的签名库

中的信息匹配。这个功能可以用一个多位寄存器来实现，而这个寄存器可以定期轮询以获得它的状态。

[0063] 同样如上所述，返回参考图2，控制器210在读取路径中包括去重复处理器230。图5示出了可用于实现根据本发明的实施例中的去重复处理器230的电路。在图5中，使用“A”指签名Scp，并且，使用“B”指与签名Scp相比较的信息(例如，在由读地址Raddr索引的条目处的信息，或使用诸如上述算法搜索的签名库中的信息)。A中的比特被标识为A[0]，...，A[n-1]，并且B中的比特被标识为B[0]，...，B[n-1]。

[0064] 在图4的框406描述的操作中，在控制器210的写入路径中，签名Scp(A)可以在缓冲器(未示出)中保持一定量的时间，直到信息(B)被读取为止。然后，在读取信息(B)之后，可以使用图5的去重复处理器230对签名Scp(A)和信息(B)进行逐比特比较。

[0065] 在图5的实施例中，去重复处理器230包括以XOR门501和502例示的多个异或(XOR)门。异或门501将签名Scp(A)的第一位(A[0]A)与信息(B)的第一位(B[0])相比较，另一异或门(未示出)将签名Scp(A)的第二位与信息(B)的第二位进行比较，以此类推，通过异或门502将签名Scp(A)的最后一位(A[n-1])与信息(B)的最后一位(B[n-1])进行比较。如果所有位都匹配，则或门511将输出二进制零；否则，或门的输出将是二进制1。这可以表示为：Out＝(A＝＝B)?1’b0:1’b1。

[0066] 因此，去重复处理器230可以使用基本电路元件(例如，门)来实现。因而，去重复处理器230可以很容易地且低成本地加入至常规的SSD控制器的设计。此外，去重复处理器230可以与SSD控制器中的其他模块共享一些电路元件，诸如ECC解码块216、解密块217和/或解压缩块218，从而进一步帮助控制成本。

[0067] 综上，在根据本发明的实施例中，SSD的内部处理电路被设计成使得签名比较可以由SSD(具体地，由SSD控制器)来执行，而不是在CPU上。因此，CPU的工作负载减少，SSD与CPU之间的数据传输量大幅减少。

[0068] 所述CPU发送一个签名SCP至SSD。在一个实施例中，CPU还发送基于签名Scp的读取地址Raddr到SSD。SSD向CPU发送信号，指示签名Scp是否与签名库中的信息匹配。多个签名、读取地址和信号可以以批处理模式发送并且如前所述并行处理。

[0069] 根据本发明的实施例，有效利用了SSD的块写入和块读取特性(例如，NAND闪存设备)。此外，与传统方法相反，将非使用数据从SSD传输至CPU上不再消耗带宽。因此，SSD上的每秒输入/输出操作(IOPS)数量显著降低，从约500，000到约9，000。要求的性能得以维持，但在CPU、存储器和带宽消耗方面成本降低。

[0070] 另外，通过提高处理数据的效率，SATA SSD可以在签名服务器中使用，这比PCIe SSD要便宜。此外，还可以配置更多的SATA SSD，而非PCIeSSD，并将其与每个CPU核心进行连接。因此，硬件成本也降低了。

[0071] 图6是在根据本发明一个实施例中的由SSD控制器执行的重复数据删除方法(例如，图2的控制器210)流程图600。

[0072] 在图6的块602中，由控制器从CPU接收针对数据块(片段、片、块)的签名Scp。在一个实施例中，也通过控制器从CPU接收与签名相关联的地址Raddr。

[0073] 在框604中，通过控制器来访问在签名库的信息。在一个实施例中，地址Raddr被控制器用来索引一个在签名库中的条目。

[0074] 在框606中，由控制器判定所述签名Scp是否与签名数据库中的信息匹配。也就是说，控制器确定签名Scp是否与签名库中的任何其他签名相匹配。在一个实施例中，控制器确定签名Scp是否与由地址Raddr寻址的条目的信息相匹配。

[0075] 在块608中，信号用于指示来自框606的结果从控制器发送给CPU。如果签名Scp与签名库中的信息匹配，则该信号具有指示数据块已经存储在SSD上的第一值。如果签名Scp与签名库中的信息不匹配，那么信号具有不同于第一值的第二值。

[0076] 如果该签名Scp与签名库中信息不匹配，则将数据块写入到SSD。如果签名Scp与签名库中的信息不匹配，则签名被添加到签名库中，并且与该数据块相关联的计数器被初始化并递增。如果签名Scp与签名库中的信息匹配，则与该数据块相关联的计数器递增。

[0077] 图7是在根据本发明的一个实施例中的通过包含CPU和SSD的签名服务器103n(图1)与SSD控制器执行重复数据删除方法的流程图700。

[0078] 在图7的框702，从客户端101m接收用于块(片，片段，块)的数据签名Scp。CPU将签名发送给SSD控制器。

[0079] 在框704，在一实施例中，基于签名的地址Raddr被确定(例如，由CPU)。在这样的实施例中，CPU发送地址Raddr到SSD控制器。

[0080] 在框706中，通过SSD控制器访问在签名库中的信息。在一实施例中，由SSD控制器使用地址Raddr来定位在签名库中的一个条目。

[0081] 在框708中，SSD控制器比较签名Scp和来自签名库的信息。在一个实施例中，SSD控制器比较签名和在签名库中的由地址Raddr寻址的信息。

[0082] 在框710，控制器产生表示该签名与签名库中的信息相匹配的信号，并将该信号发送到CPU。如果签名Scp与签名库中的信息相匹配，则该数据块已经存储在SSD中。如果签名Scp与签名库中的信息不匹配，则该信号从签名服务器发送到客户端，响应于所述信号，来自客户端的该数据块在签名服务器处被接收，并将该数据块写入SSD。

[0083] 图8是示出了根据本发明的一个实施例中的重复数据删除系统800的一个实现方式的示例的框图。去重复系统800可以部署在存储集群上，并且可以直接在由应用程序802存储的备份副本804、805、806和807上工作。在全局层面分析数据的冗余，则去重复系统800移除重复块，仅保留独特的块，同时也更新相关的元数据。之后，当特定数据被访问或更新时，所述元数据和独特的块被修改。通过这种机制，备份副本804-807所消耗的存储量(例如，按数量级)显著减少。

[0084] SATA SSD而非PCIe SSD可如上述那样使用。每个签名服务器都能够驱动更多SSD，例如，12个SATA SSD与4个PCIe卡。由于本发明的结果，每个SATA SSD在满足去重复需求方面的性能和PCIe SSD一样好。因此，每个签名服务器的性能都是提高了三倍。换句话说，对于相同的性能，签名服务器的数量可以减少三分之一。因此，降低成本是因为SATA SSD的成本低于PCIeSSD，并且需要更少的签名服务器。此外，通过消除回收大量不必要数据的需要，节省了计算机资源。

[0085] 尽管前面的公开使用特定的框图，流程图，以及实例展示的每个框图部件，流程图步骤，操作阐述的各种实施例和/或组件描述和/或示出可被单独地和/或共同，使用各种硬件，软件或固件(或其任何组合)配置。此外，其他组件中包含的组件的任何公开都应被视为示例，因为可以实现许多其他体系结构以实现相同的功能。

[0086] 本文所描述和/或说明的步骤参数和步骤仅以示例方式给出，并可根据需要进行更改。例如，尽管本文所描述和/或描述的步骤可以以特定的顺序示出或讨论，但这些步骤并不一定需要在演示或讨论的顺序中执行。本文所描述和/或说明的各种示例方法也可以忽略本文中描述或说明的一个或多个步骤，或者除了所公开的步骤之外还包括附加步骤。

[0087] 虽然，已经在全功能计算系统的上下文中描述和/或说明了各种实施例，但是这些示例实施例中的一个或多个可以作为各种形式的程序产品来分发，而不管所使用的实际执行分发的计算机可读介质的具体类型。这里公开的实施例还可以使用执行特定任务的软件模块来实现。这些软件模块可以包括可以存储在计算机可读存储介质上或计算系统中的脚本、批处理或其他可执行文件。这些软件模块可以配置计算系统以执行这里公开的示例实施例中的一个或多个。这里公开的一个或多个软件模块可以在云计算环境中实现。云计算环境可以通过互联网提供各种服务和应用程序。这些基于云的服务器(例如，软件作为服务器，平台作为服务器，基础架构作为服务器等)可通过Web浏览器或其他远程接口访问。这里描述的各种功能可以通过远程桌面环境或任何其他基于云的计算环境来提供。

[0088] 尽管本主题已经以特定于结构特征和/或方法动作的语言进行了描述，但是应当理解的是，在本发明中定义的主题不必限于上述具体特征或动作。而是，上面描述的具体特征和行为被公开为实现本公开的示例形式。

[0089] 根据本发明的实施方式。尽管已经在特定实施例中描述了本公开，但应该认识到，本发明不应该被解释为受这些实施例的限制，而是根据本申请的权利要求来解释。

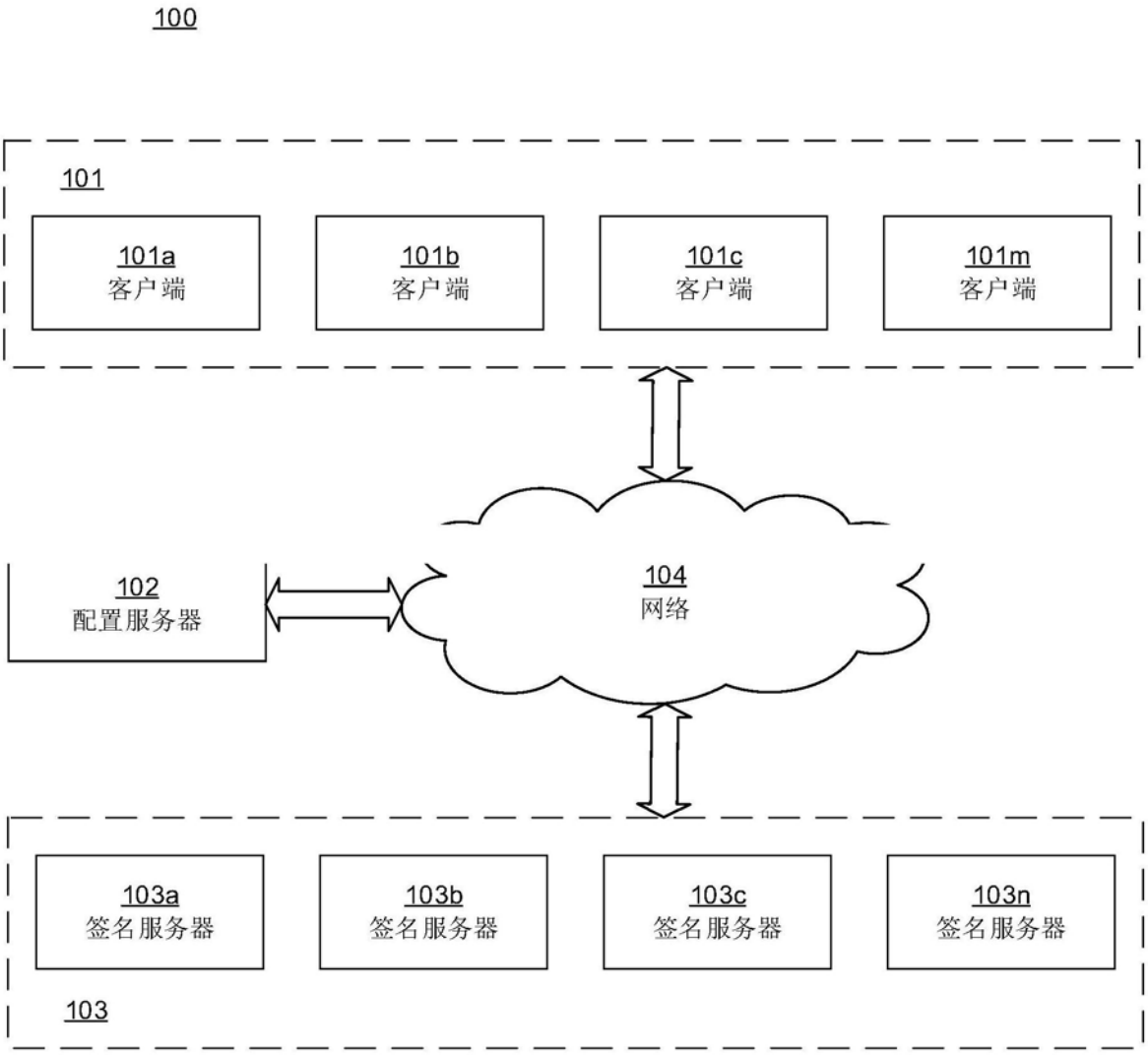

100
101
101a
客户端
101b
客户端
101c
客户端
101m
客户端
102
配置服务器
104
网络
103a
签名服务器
103b
签名服务器
103c
签名服务器
103n
签名服务器
103

图1

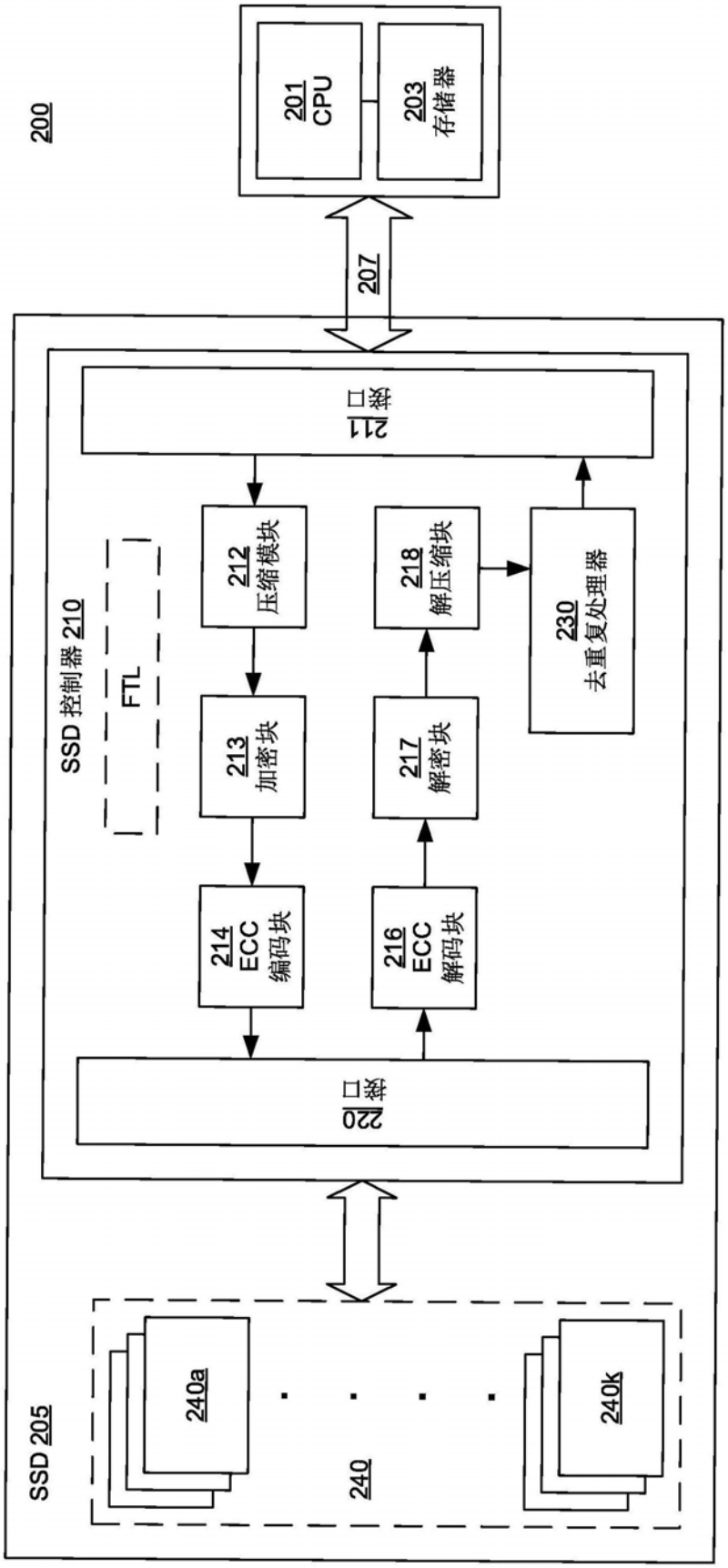

200
201 CPU
203 存储器
207
SSD 控制器 210
FTL
211 接口
212 压缩模块
213 加密块
214 ECC 编码块
216 ECC 解码块
217 解密块
218 解压缩块
230 去重复处理器
220 接口
SSD 205
240a
240
240k

图2

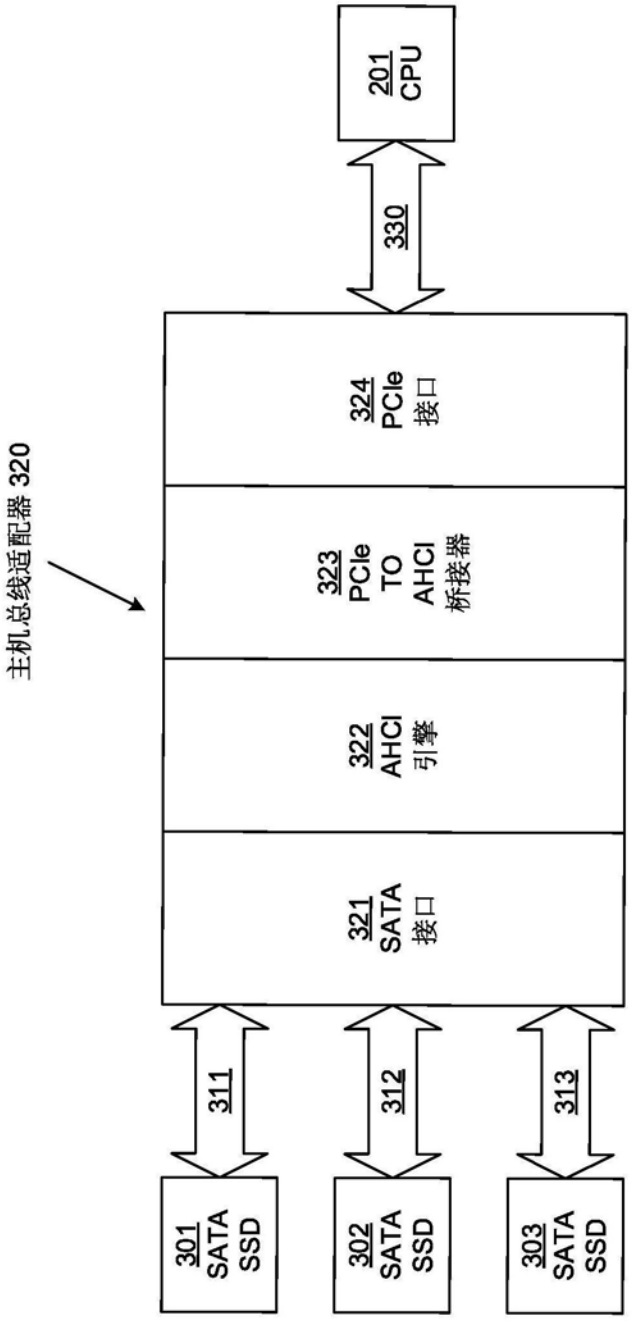

201 CPU
330
主机总线适配器 320
324 PCIe 接口
323 PCIe TO AHCI 桥接器
322 AHCI 引擎
321 SATA 接口
311
312
313
301 SATA SSD
302 SATA SSD
303 SATA SSD

图3

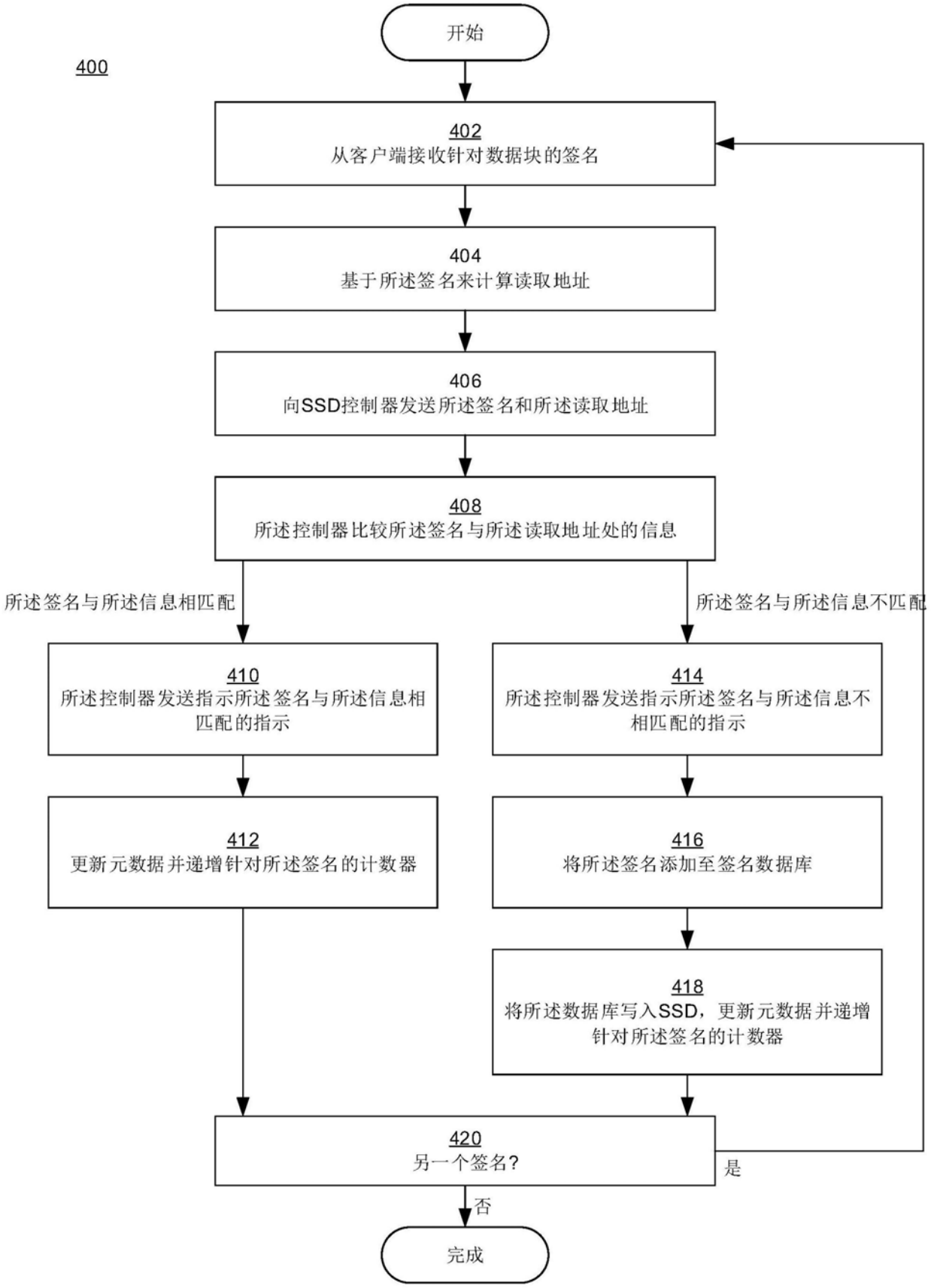

400
开始
402
从客户端接收针对数据块的签名
404
基于所述签名来计算读取地址
406
向SSD控制器发送所述签名和所述读取地址
408
所述控制器比较所述签名与所述读取地址处的信息
所述签名与所述信息相匹配
所述签名与所述信息不匹配
410
所述控制器发送指示所述签名与所述信息相匹配的指示
412
更新元数据并递增针对所述签名的计数器
414
所述控制器发送指示所述签名与所述信息不相匹配的指示
416
将所述签名添加至签名数据库
418
将所述数据库写入SSD，更新元数据并递增针对所述签名的计数器
420
另一个签名?
是
否
完成

图4

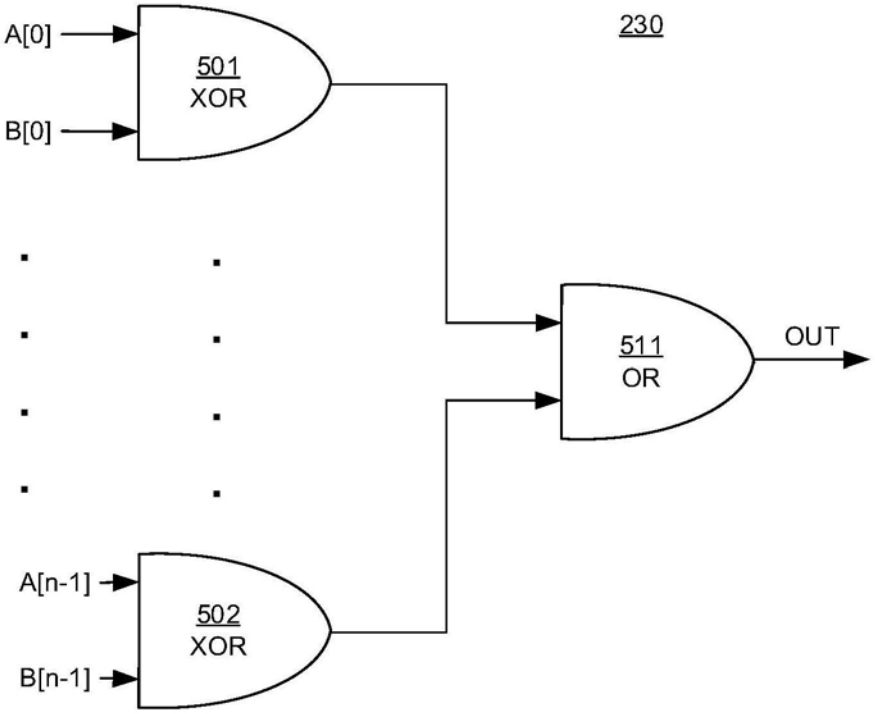

230
A[0]
B[0]
501
XOR
A[n-1]
B[n-1]
502
XOR
511
OR
OUT

图5

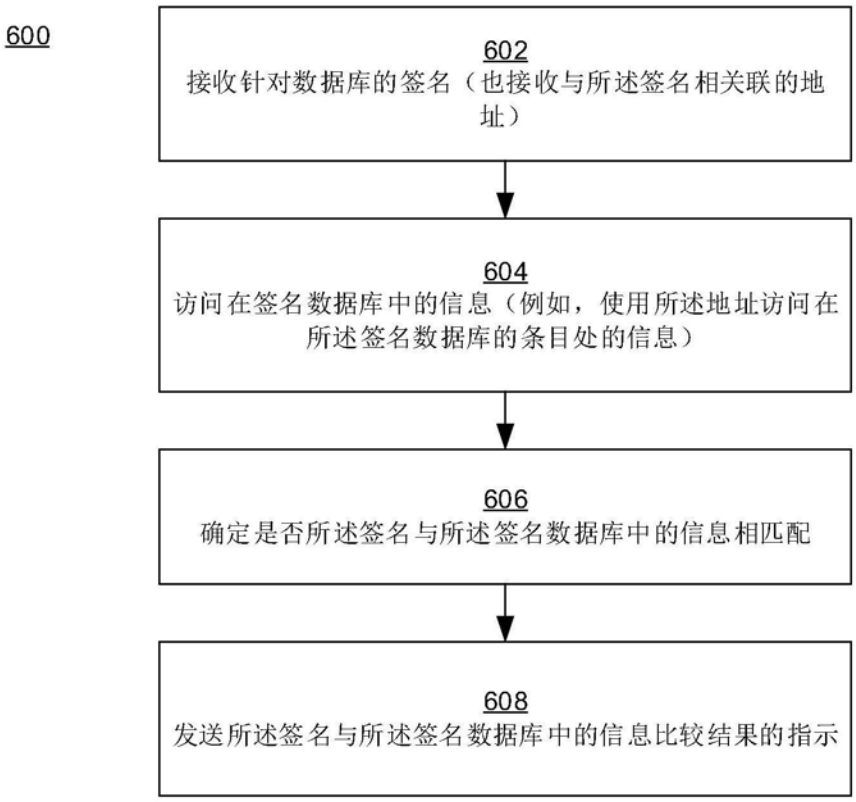

600
602
接收针对数据库的签名（也接收与所述签名相关联的地址）
604
访问在签名数据库中的信息（例如，使用所述地址访问在所述签名数据库的条目处的信息）
606
确定是否所述签名与所述签名数据库中的信息相匹配
608
发送所述签名与所述签名数据库中的信息比较结果的指示

图6

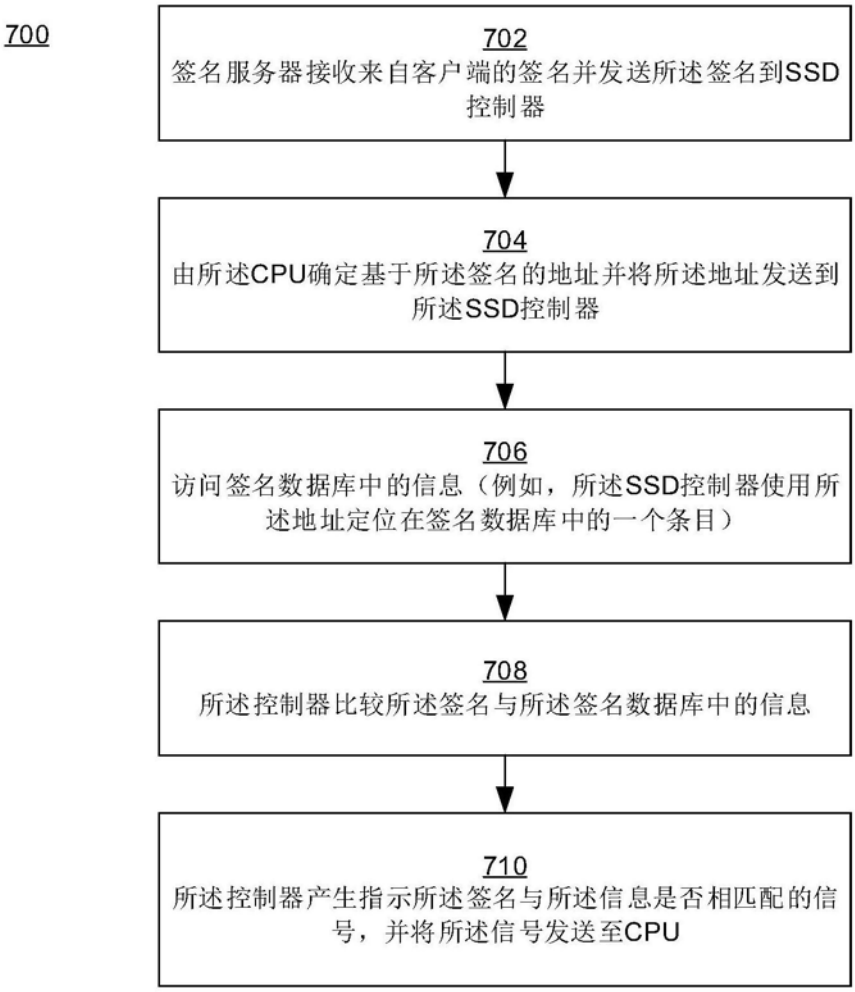

700
702
签名服务器接收来自客户端的签名并发送所述签名到SSD控制器
704
由所述CPU确定基于所述签名的地址并将所述地址发送到所述SSD控制器
706
访问签名数据库中的信息（例如，所述SSD控制器使用所述地址定位在签名数据库中的一个条目）
708
所述控制器比较所述签名与所述签名数据库中的信息
710
所述控制器产生指示所述签名与所述信息是否相匹配的信号，并将所述信号发送至CPU

图7

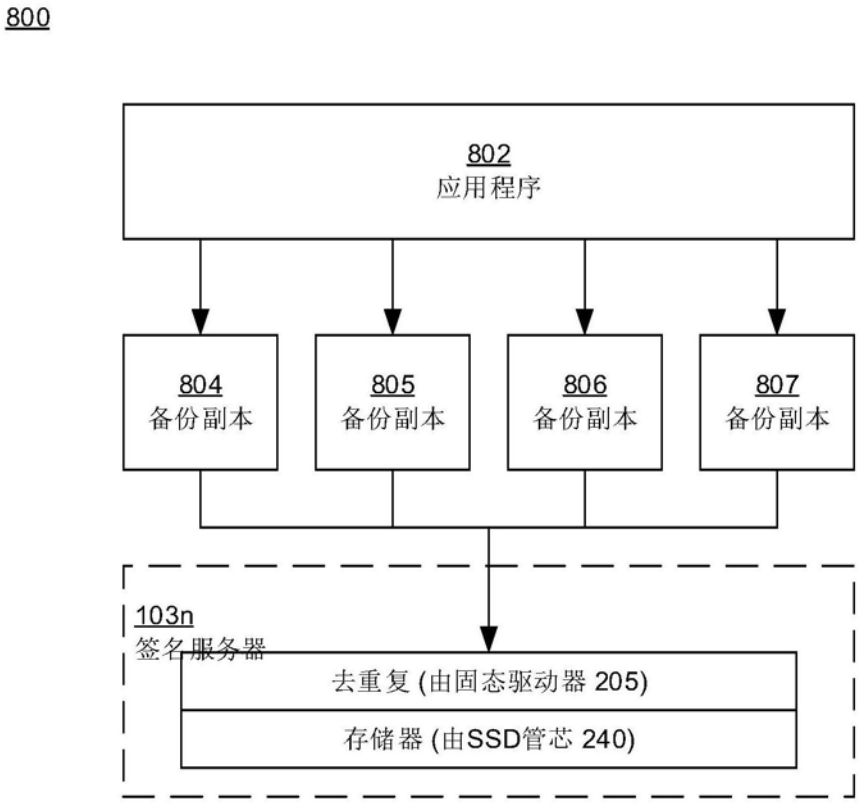

800
802
应用程序
804
备份副本
805
备份副本
806
备份副本
807
备份副本
103n
签名服务器
去重复 (由固态驱动器 205)
存储器 (由SSD管芯 240)

图8