[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
G06F 11/20 (2006.01)
G06F 11/14 (2006.01)

[12] 发明专利说明书

专利号 ZL 200480012741.6

[45] 授权公告日 2007 年 12 月 19 日

[11] 授权公告号 CN 100356336C

[22] 申请日 2004.6.11
[21] 申请号 200480012741.6
[30] 优先权
[32] 2003.6.18 [33] US [31] 10/465,118
[86] 国际申请 PCT/EP2004/051044 2004.6.11
[87] 国际公布 WO2004/114136 英 2004.12.29
[85] 进入国家阶段日期 2005.11.10
[73] 专利权人 国际商业机器公司
地址 美国纽约
[72] 发明人 W·斯坦利 W·F·米奇卡
S·C·沃纳 S·塔尔
G·A·斯皮尔 T·M·布朗
M·桑切斯 S·拉哈乌
T·C·贾维斯 A·哈亚拉丹尼
D·察内法瑞尔 S·菲恩布利特
R·M·马托舍维奇 S·舒克维奇
I·努里尔
[56] 参考文献
US6092066A 2000.7.18
CN1010490A 1987.4.15
审查员 牛晓丽
[74] 专利代理机构 北京市中咨律师事务所
代理人 于 静 李 峥

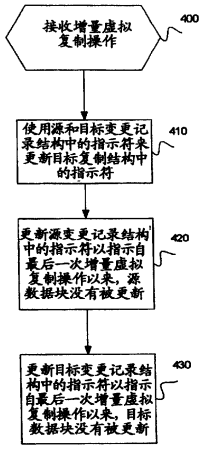
权利要求书 2 页 说明书 18 页 附图 8 页

[54] 发明名称

用于增量虚拟复制的方法和系统

[57] 摘要

本发明公开了一种用于减少传输的数据量的技术。 为每个源数据块维护第一指示符以指示自所述源数据块被最后一次传输到目标存储器以来，所述源数据块是否已在源存储器中被更新。 为所述目标存储器中的每个目标数据块维护第二指示符以指示自所述目标数据块被对应的源数据块覆盖以来，所述目标数据块是否已在目标存储器中被更新。 当从所述源存储器向所述目标存储器传输数据时，传输已为其设置第一指示符以指示源数据块已被更新的每个源数据块，并且传输与已为其设置第二指示符以指示目标数据块已被更新的目标数据块对应的每个源数据块。



1. 一种用于减少在源与目标之间传输的数据量的方法，所述方法包括：为每个源数据块维护第一指示符以指示自所述源数据块被最后一次传输到目标存储器以来，所述源数据块是否已在源存储器中被更新；为所述目标存储器中的每个目标数据块维护第二指示符以指示自所述目标数据块被对应的源数据块覆盖以来，所述目标数据块是否已在目标存储器中被更新；为每个目标数据块维护第三指示符；当从所述源存储器向所述目标存储器传输数据时，传输已为其设置第一指示符以指示源数据块已被更新的每个源数据块；以及传输与已为其设置第二指示符以指示目标数据块已被更新的目标数据块对应的每个源数据块；其中从所述源存储器向所述目标存储器传输数据包括传输所述第三指示符为其指示源数据块将被传输到所述目标存储器的每个源数据块；所述方法的特征在于包括：

使所述源和所述目标形成增量虚拟复制关系；

使用所述第三指示符来指示在增量虚拟复制操作之前的运行期间相应的源数据块或所述目标数据块是否将被检索以用于登台操作；

当源数据块已被复制到所述目标存储器时，为所述目标数据块更新所述第三指示符；

接收增量虚拟复制操作；以及

响应于接收到所述增量虚拟复制操作，根据所述源数据块的所述第一指示符、与所述源数据块对应的所述目标数据块的所述第二指示符以及所述第三指示符的当前值来更新每个目标数据块的所述第三指示符。

2. 一种用于减少在源与目标之间传输的数据量的系统，所述系统包括：用于为每个源数据块维护第一指示符以指示自所述源数据块被最后一次传输到目标存储器以来，所述源数据块是否已在源存储器中被更新的装置；用于为所述目标存储器中的每个目标数据块维护第二指示符以指示自所述目标数据块被对应的源数据块覆盖以来，所述目标数据块是否已在目标存储器中被更新的装置；用于为每个目标数据块维护第三指示符的装置；用于当从所述源存储器向所述目标存储器传输数据时，传输已为其设置第

一指示符以指示源数据块已被更新的每个源数据块的装置；以及用于传输与已为其设置第二指示符以指示目标数据块已被更新的目标数据块对应的每个源数据块的装置；其中所述用于从所述源存储器向所述目标存储器传输数据的装置包括用于传输所述第三指示符为其指示源数据块将被传输到所述目标存储器的每个源数据块的装置；所述系统的特征在于包括:

用于使所述源和所述目标形成增量虚拟复制关系的装置;

用于使用所述第三指示符来指示在增量虚拟复制操作之前的运行期间相应的源数据块或所述目标数据块是否将被检索以用于登台操作的装置;

用于当源数据块已被复制到所述目标存储器时，为所述目标数据块更新所述第三指示符的装置;

用于接收增量虚拟复制操作的装置；以及

用于响应于接收到所述增量虚拟复制操作，根据所述源数据块的所述第一指示符、与所述源数据块对应的所述目标数据块的所述第二指示符以及所述第三指示符的当前值来更新每个目标数据块的所述第三指示符的装置。

用于增量虚拟复制的方法和系统

技术领域

本发明涉及增量虚拟复制。

背景技术

计算系统通常包括一个或多个用于处理数据和运行应用程序的主计算机（“主机”）、用于存储数据的直接存取存储设备（DASD）以及用于控制主机与 DASD 之间的数据传输的存储控制器。存储控制器，也被称为控制单元或存储导向器，管理对存储空间（包括以环形架构相连的大量硬盘驱动器）的访问，另外也被称为直接存取存储设备（DASD）。主机可以通过存储控制器传送到存储空间的输入/输出（I/O）请求。

在许多系统中，存储设备（例如 DASD）上的数据可以被复制到同一或另一存储设备，以便可以从两个不同的设备来提供对数据卷的访问。即时复制包括将所有数据从源卷物理地复制到目标卷，以便目标卷即时具有数据的副本。也可以通过逻辑地制作数据的副本然后只在需要时完全复制数据来进行即时复制，从而有效地推迟物理复制。此逻辑复制操作被执行以便最小化其间不可访问目标卷和源卷的时间。

许多直接存取存储设备（DASD）子系统能够执行“即刻虚拟复制”操作，也被称为“快速复制功能”。即刻虚拟复制操作通过修改元数据（例如关系表或指针）以将源数据对象同时视为原件和副本来工作。为了响应主机的复制请求，存储子系统在没有对数据进行任何物理复制的情况下立即报告副本的创建。仅创建了“虚拟”副本，而主机完全不知道并不存在附加的物理副本。

随后，当存储系统接收到对原件和副本的更新时，所述更新被分别存储并仅被交叉引用到更新的数据对象。此刻，原件和副本数据对象开始出

现差异。首要的好处是即刻虚拟复制几乎瞬时地发生，比正常的物理复制操作完成的快得多。这释放了主机和存储子系统以便可以执行其他任务。主机或存储子系统甚至可以在后台处理期间继续创建原始数据对象的实际物理副本，或在其他时间创建该副本。

一个此类即刻虚拟复制操作是熟知的 FlashCopy 7 操作。FlashCopy 7 操作包括在同一或不同设备上的源与目标卷之间建立逻辑的即时关系。FlashCopy 7 操作保证直到 FlashCopy 7 关系中的磁道已被硬化到其在目标磁盘上的位置时，该磁道才驻留在源磁盘上。关系表被用来维护与子系统中所有现有 FlashCopy 7 关系有关的信息。在 FlashCopy 7 关系的建立阶段，为参与正在被建立的 FlashCopy 7 的源和目标在源和目标关系表中记录一个表项。每个所添加的表项都维护关于 FlashCopy 7 关系的全部所需信息。当来自源范围的全部 FlashCopy 7 磁道都已被物理地复制到目标范围或者当接收到撤消命令时，从关系表中删除所述关系的两个表项。在某些情况下，即使所有磁道都已被从源范围复制到目标范围，所述关系仍然继续存在。

所述目标关系表还包括一个位图，该位图标识了哪些涉及 FlashCopy 7 关系的磁道尚未被全部复制并且因此是受保护的磁道。目标设备中的每个磁道由所述位图中的一个位来表示。当相应的磁道被建立为 FlashCopy 7 关系的目标磁道时，目标位被设置。当相应的磁道已被从源位置复制并且离台（destage）到目标设备（由于源或目标设备上的写入，或后台复制任务）时，目标位被重置。

在现有技术中，作为在 FlashCopy 7 操作期间逻辑即时关系的建立的一部分，包括在 FlashCopy 7 关系中的源高速缓存内的所有磁道都必须被离台到物理源卷（例如源 DASD），并且包括在 FlashCopy 7 中的目标高速缓存内的所有磁道都必须被丢弃。

一旦建立了逻辑关系，主机就可以立即访问源和目标卷上的数据，并且数据可以作为后台操作的一部分被复制。对是 FlashCopy 7 关系中的目标并且不在高速缓存中的磁道的读取将触发登台（stage）拦截，当源磁

道尚未被复制完毕并且在访问被从目标高速缓存提供给该磁道之前，这将导致与所请求的目标磁道对应的源磁道被登台到目标高速缓存。这确保了所述目标具有来自在 FlashCopy 7 操作的时刻存在的源的副本。进而，任何对尚未被复制完毕的源设备上的磁道的离台都将触发离台拦截，这将导致源设备上的磁道被复制到目标设备。

已经开发了即刻虚拟复制技术，以便（至少部分地）快速创建数据的重复副本而不中断或减缓前台过程。即刻虚拟复制技术（如 FlashCopy 7 操作）提供了一种即时复制工具。即刻虚拟复制技术可被用于各种应用，包括例如数据备份、数据迁移、数据挖掘、测试等。例如，即刻虚拟复制技术可用于创建源数据的物理“备份”副本以帮助灾难恢复。

尽管即刻虚拟复制技术对于复制大量数据是很有用的，但是常规的即刻虚拟复制技术仍可以被改进。具体地说，本领域中需要避免物理地复制大量数据的改进的即刻虚拟复制技术。

发明内容

在第一个方面中，本发明提供了一种用于减少传输的数据量的方法，所述方法包括：为每个源数据块维护第一指示符以指示自所述源数据块被最后一次传输到目标存储器以来，所述源数据块是否已在源存储器中被更新；为所述目标存储器中的每个目标数据块维护第二指示符以指示自所述目标数据块被对应的源数据块覆盖以来，所述目标数据块是否已在目标存储器中被更新；以及当从所述源存储器向所述目标存储器传输数据时，传输已为其设置第一指示符以指示源数据块已被更新的每个源数据块；以及传输与已为其设置第二指示符以指示目标数据块已被更新的目标数据块对应的每个源数据块。

所述方法优选地还包括：为每个目标数据块维护第三指示符以指示相应的源数据块或所述目标数据块是否将被检索以用于登台操作。

所述方法优选地还包括：当源数据块已被复制到所述目标存储器时，为所述目标数据块更新所述第三指示符。

所述方法优选地还包括：接收增量虚拟复制操作；以及根据所述源数据块的所述第一指示符和与所述源数据块对应的所述目标数据块的所述第二指示符来更新每个目标数据块的所述第三指示符，其中从所述源存储器向所述目标存储器传输数据包括传输第三指示符为其指示源数据块将被传输到所述目标存储器的每个源数据块。

所述方法优选地还包括：在建立了增量虚拟复制关系后，更新每个目标数据块的所述第三指示符以指示与所述目标数据块对应的所述源数据块将被从所述源存储器中检索出来以用于登台操作。

所述方法优选地还包括：在从所述源存储器向所述目标存储器传输数据后，更新每个源数据块的所述第一指示符以指示自所述源数据块被最后一次传输到所述目标存储器以来，所述源数据块尚未在所述源存储器中被更新。

所述方法优选地还包括：在从所述源存储器向所述目标存储器传输数据后，更新每个目标数据块的所述第二指示符以指示自所述源数据块被最后一次传输到所述目标存储器以来，所述目标数据块尚未在所述目标存储器中被更新。

优选地，每个源数据块和每个目标数据块包括卷的磁道。

优选地，每个源数据块和每个目标数据块包括卷的扇区。

所述方法优选地还包括：通过在源数据部分与目标数据部分之间执行增量虚拟复制操作来在所述源数据部分与所述目标数据部分之间创建增量虚拟复制关系，其中为所述增量虚拟复制关系中的所述源和目标数据部分中的每个源数据块和目标数据块分别维护所述第一指示符和所述第二指示符。

所述方法优选地还包括：更新每个源数据块的所述第一指示符以指示所述源数据块没有被更新；更新每个目标数据块的所述第二指示符以指示所述目标数据块没有被更新；以及更新每个目标数据块的所述第三指示符以指示所述源数据块需要被传输到所述目标存储器。

所述方法优选地还包括：当源数据块被更新时，更新所述源数据块的

所述第一指示符以指示所述源数据块已被更新。

所述方法优选地还包括：当目标数据块被更新时，更新所述目标数据块的所述第二指示符以指示所述目标数据块已被更新。

还可以提供一件用于减少传输的数据量的制品，其中所述制品导致操作，所述操作包括：为每个源数据块维护第一指示符以指示自所述源数据块被最后一次传输到目标存储器以来，所述源数据块是否已在源存储器中被更新；为所述目标存储器中的每个目标数据块维护第二指示符以指示自所述目标数据块被对应的源数据块覆盖以来，所述目标数据块是否已在目标存储器中被更新；以及当从所述源存储器向所述目标存储器传输数据时，传输已为其设置第一指示符以指示源数据块已被更新的每个源数据块；以及传输与已为其设置第二指示符以指示目标数据块已被更新的目标数据块对应的每个源数据块。

所述制品优选地包括操作，所述操作还包括：为每个目标数据块维护第三指示符以指示相应的源数据块或所述目标数据块是否将被检索以用于登台操作。

所述制品优选地包括：当源数据块已被复制到所述目标存储器时，为所述目标数据块更新所述第三指示符。

所述制品优选地包括：接收增量虚拟复制操作；以及根据所述源数据块的所述第一指示符和与所述源数据块对应的所述目标数据块的所述第二指示符来更新每个目标数据块的所述第三指示符，其中从所述源存储器向所述目标存储器传输数据包括传输第三指示符为其指示源数据块将被传输到所述目标存储器的每个源数据块。

所述制品优选地包括：在建立了增量虚拟复制关系后，更新每个目标数据块的所述第三指示符以指示与所述目标数据块对应的所述源数据块将被从所述源存储器中检索出来以用于登台操作。

所述制品优选地包括：在从所述源存储器向所述目标存储器传输数据后，更新每个源数据块的所述第一指示符以指示自所述源数据块被最后一次传输到所述目标存储器以来，所述源数据块尚未在所述源存储器中被更

新。

所述制品优选地包括：在从所述源存储器向所述目标存储器传输数据后，更新每个目标数据块的所述第二指示符以指示自所述源数据块被最后一次传输到所述目标存储器以来，所述目标数据块尚未在所述目标存储器中被更新。

优选地，每个源数据块和每个目标数据块包括卷的磁道。

优选地，每个源数据块和每个目标数据块包括卷的扇区。

所述制品优选地包括：通过在源数据部分与目标数据部分之间执行增量虚拟复制操作来在所述源数据部分与所述目标数据部分之间创建增量虚拟复制关系，其中为所述增量虚拟复制关系中的所述源和目标数据部分中的每个源数据块和目标数据块分别维护所述第一指示符和所述第二指示符。

所述制品优选地包括：更新每个源数据块的所述第一指示符以指示所述源数据块没有被更新；更新每个目标数据块的所述第二指示符以指示所述目标数据块没有被更新；以及更新每个目标数据块的所述第三指示符以指示所述源数据块需要被传输到所述目标存储器。

所述制品优选地包括：当源数据块被更新时，更新所述源数据块的所述第一指示符以指示所述源数据块已被更新。

所述制品优选地包括：当目标数据块被更新时，更新所述目标数据块的所述第二指示符以指示所述目标数据块已被更新。

在第二个方面中，本发明提供了一种用于减少传输的数据量的系统，所述系统包括：用于为每个源数据块维护第一指示符以指示自所述源数据块被最后一次传输到目标存储器以来，所述源数据块是否已在源存储器中被更新的装置；用于为所述目标存储器中的每个目标数据块维护第二指示符以指示自所述目标数据块被对应的源数据块覆盖以来，所述目标数据块是否已在目标存储器中被更新的装置；以及当从所述源存储器向所述目标存储器传输数据时，用于传输已为其设置第一指示符以指示源数据块已被更新的每个源数据块的装置；以及用于传输与已为其设置第二指示符以指

示目标数据块已被更新的目标数据块对应的每个源数据块的装置。

所述系统优选地包括：用于为每个目标数据块维护第三指示符以指示相应的源数据块或所述目标数据块是否将被检索以用于登台操作的装置。

所述系统优选地包括：当源数据块已被复制到所述目标存储器时，用于为所述目标数据块更新所述第三指示符的装置。

所述系统优选地包括：用于接收增量虚拟复制操作的装置；以及用于根据所述源数据块的所述第一指示符和与所述源数据块对应的所述目标数据块的所述第二指示符来更新每个目标数据块的所述第三指示符的装置，其中从所述源存储器向所述目标存储器传输数据包括传输第三指示符为其指示源数据块将被传输到所述目标存储器的每个源数据块。

所述系统优选地包括：在建立了增量虚拟复制关系后，用于更新每个目标数据块的所述第三指示符以指示与所述目标数据块对应的所述源数据块将被从所述源存储器中检索出来以用于登台操作的装置。

所述系统优选地包括：在从所述源存储器向所述目标存储器传输数据后，用于更新每个源数据块的所述第一指示符以指示自所述源数据块被最后一次传输到所述目标存储器以来，所述源数据块尚未在所述源存储器中被更新的装置。

所述系统优选地包括：在从所述源存储器向所述目标存储器传输数据后，用于更新每个目标数据块的所述第二指示符以指示自所述源数据块被最后一次传输到所述目标存储器以来，所述目标数据块尚未在所述目标存储器中被更新的装置。

优选地，每个源数据块和每个目标数据块包括卷的磁道。

优选地，每个源数据块和每个目标数据块包括卷的扇区。

所述系统优选地包括：用于通过在源数据部分与目标数据部分之间执行增量虚拟复制操作来在所述源数据部分与所述目标数据部分之间创建增量虚拟复制关系的装置，其中为所述增量虚拟复制关系中的所述源和目标数据部分中的每个源数据块和目标数据块分别维护所述第一指示符和所述第二指示符。

所述系统优选地包括：用于更新每个源数据块的所述第一指示符以指示所述源数据块没有被更新的装置；用于更新每个目标数据块的所述第二指示符以指示所述目标数据块没有被更新的装置；以及用于更新每个目标数据块的所述第三指示符以指示所述源数据块需要被传输到所述目标存储器的装置。

所述系统优选地包括：当源数据块被更新时，用于更新所述源数据块的所述第一指示符以指示所述源数据块已被更新的装置。

所述系统优选地包括：当目标数据块被更新时，用于更新所述目标数据块的所述第二指示符以指示所述目标数据块已被更新的装置。

在第三个方面中，本发明提供了一种包括计算机程序代码的计算机程序，当所述计算机程序被加载到计算机系统中并在其上执行时，所述计算机程序代码将导致所述计算机系统执行根据所述第一个方面的方法的诸步骤。所述计算机程序的优选特征包括与所述第一个方面的诸优选方法步骤对应的计算机程序代码。

因此，优选地提供了一种用于减少传输的数据量的方法、系统和程序。为每个源数据块维护第一指示符以指示自所述源数据块被最后一次传输到目标存储器以来，所述源数据块是否已在源存储器中被更新。为所述目标存储器中的每个目标数据块维护第二指示符以指示自所述目标数据块被对应的源数据块覆盖以来，所述目标数据块是否已在目标存储器中被更新。当从所述源存储器向所述目标存储器传输数据时，传输已为其设置第一指示符以指示源数据块已被更新的每个源数据块，并且传输与已为其设置第二指示符以指示目标数据块已被更新的目标数据块对应的每个源数据块。

本发明的所述实现提供了一种用于创建增量虚拟复制的方法、系统和程序。

附图说明

现在将仅通过实例的方式并参考附图对本发明的一个优选实施例进行描述，这些附图是：

图 1A 和 1B 以方块图的形式示出了根据本发明的特定实现的计算环境;

图 2 示出了根据本发明的特定实现的各种结构;

图 3 示出了根据本发明的特定实现的用于更新结构的逻辑;

图 4 示出了根据本发明的特定实现的用于执行增量虚拟复制操作的逻辑;

图 5 示出了根据本发明的特定实现的在处理写入操作的写入过程中实现的逻辑;

图 6 示出了根据本发明的特定实现的在处理读取操作的读取过程中实现的逻辑;

图 7 示出了根据本发明的特定实现的后台复制过程;

图 8 示出了根据本发明的特定实现的可以使用的计算机系统的体系结构。

具体实施方式

在以下描述中，参考了形成本发明的一部分并示出了本发明的若干实现的附图。可以理解，在不偏离本发明的范围的情况下，可以使用其他实现并对结构和操作进行更改。

本发明的实现提供了一种增量虚拟复制操作，其是对即刻虚拟复制操作的增强。使用所述增量虚拟复制操作，只有自从源卷到目标卷的最后一次即刻虚拟复制操作以来在所述源和目标卷上被更新的数据块才被复制。增量虚拟复制操作减少了创建源卷的物理副本的持续时间并最小化了对其他应用（例如，最小化了用于将数据离台到物理存储的带宽利用率，从而允许更多的带宽用于从物理存储进行读取）的影响。

图 1A 和 1B 以方块图的形式示出了根据本发明的特定实现的计算环境。存储控制器 100 通过导向存储设备 120、130 的网络 190 分别接收来自主机 140a,b,...l（其中，a、b 和 l 可以是任意整数值）的输入/输出（I/O）请求，所述存储设备 120、130 被配置成具有卷（例如，逻辑单元号、逻辑

设备等）122a,b,...n 以及 132a,b,...m，其中 m 和 n 可以是不同的整数值也可以是相同的整数值。在特定实现中，目标存储器 130 的大小可以大于或等于源存储器 120 的大小。

源存储器 120 包括一个或多个卷 122a,b,...n，其可以被分成包含数据块的存储块 150，而存储块 150 进一步被分成包含数据子块的存储子块（150a-150p，其中 a 和 p 可以是任意整数值）。卷可以是存储器的任何逻辑或物理元素。在特定实现中，数据块是磁道的内容，而数据子块是磁道的扇区的内容。

目标存储器 130 保存源存储器 120 的卷 122a,b,...n 中的所有卷的副本或其子集的副本。此外，目标存储器 130 可以由例如主机 140 来修改。目标存储器 130 包括一个或多个卷 132a,b,...m，其可以被分成包含数据块的存储块 150，而存储块 150 进一步被分成包含数据子块的存储子块（150a-150p，其中 a 和 p 可以是任意整数值）。卷可以是存储器的任何逻辑或物理元素。在特定实现中，数据块是磁道，而数据子块是磁道的扇区。

为了便于引用，术语磁道和扇区将在此用作数据块和数据子块的实例，但这些术语的使用并非旨在将本发明的实现限制为磁道和扇区。本发明的实现可适用于任何类型的以任何方式划分的存储器、存储块或数据块。此外，尽管本发明的实现涉及数据块，但本发明的可替代实现可适用于数据子块。

存储控制器 100 包括源高速缓存 124，其中保存对源存储器 120 中的磁道的更新，直到写入源存储器 120 为止（即，所述磁道被离台到物理存储）。存储控制器 100 包括目标高速缓存 134，其中保存对目标存储器 130 中的磁道的更新，直到写入目标存储器 130 为止（即，所述磁道被离台到物理存储）。源高速缓存 124 和目标高速缓存 134 可以包括单独的存储设备或同一存储设备的不同部分。源高速缓存 124 和目标高速缓存 134 被用来缓存主机 140a,b,...l、源存储器 120 以及目标存储器 130 之间被传送的读写数据。进而，尽管高速缓存 124 和 134 被分别称为源和目标高速缓存以便保存即时复制关系中的源或目标数据块，但是高速缓存 124 和 134 可以

同时存储不同即时复制关系中的源和目标数据块。

此外，存储控制器 100 包括非易失性高速缓存 118。非易失性高速缓存 118 可以是例如后备电池的易失性存储器以保存数据更新的非易失性副本。

存储控制器 100 还包括系统存储器 110，其可以以易失性和/或非易失性设备来实现。系统存储器 110 包括用于读取数据的读取过程 112、用于写入数据的写入过程 114 以及增量虚拟复制过程 116。读取过程 112 在系统存储器 110 中执行以将数据分别从存储器 120 和 130 读取到高速缓存 124 和 134。写入过程 114 在系统存储器 110 中执行以将数据分别从高速缓存 124 和 134 写入存储器 120 和 130。增量虚拟复制过程 116 在系统存储器 110 中执行以便执行将数据从源存储器 120 传输到目标存储器 130 的增量虚拟复制操作。在本发明的特定实现中，可以有多个增量虚拟复制过程。在本发明的特定实现中，所述增量虚拟复制过程可以在与存储控制器 100 相连的其他存储控制器上执行，而不是在存储控制器 100 上执行，或者在除存储控制器 100 以外的存储控制器上执行。系统存储器 110 可以位于与高速缓存 124 和 134 分离的存储设备中，或者可以与高速缓存 124 和 134 中的一个或两个高速缓存共享存储设备。

本发明的实现可适用于在任意两个存储介质（为了便于引用，其在此处将被称为源存储器和目标存储器）之间传输数据。例如，如图 1A 所示，本发明的特定实现可以与位于单个存储控制器处的两个存储介质一起使用。此外，本发明的特定可替代实现可以与位于不同存储控制器、不同物理站点等的两个存储介质一起使用。此外，为了便于引用，源存储器中的数据块将被称为“源数据块”，而目标存储器中的数据块将被称为“目标数据块”。

在特定实现中，可移动存储器（而不是目标存储器 130 或除目标存储器 130 之外的存储器）可被用于保存所有源存储器 120 的副本或源存储器 120 的子集的副本，并且本发明的实现将数据传输到所述可移动存储器而不是所述目标存储器。所述可移动存储器可以位于存储控制器 100 处。

存储控制器 100 可以还包括处理器复合体（未示出）并且可以包括本领域公知的任何存储控制器或服务器，例如 Enterprise Storage Server 7（ESS）、39907 存储控制器等。主机 140a,b,...l 可以包括本领域公知的任何计算设备，例如服务器、大型机、工作站、个人计算机、手持计算机、膝上电话设备、网络家电等。存储控制器 100 和（多个）主机系统 140a,b,...l 通过网络 190 通信，网络 190 可以包括存储区域网络（SAN）、局域网（LAN）、广域网（WAN）、因特网、以太网等。源存储器 120 和目标存储器 130 可以各自包括一组存储设备，例如直接存取存储设备（DASD）、完全磁盘束（JBOD）、独立磁盘冗余阵列（RAID）、虚拟化设备等。

此外，尽管图 1A 示出了单独的存储控制器，但本领域的技术人员将理解，可以通过网络（例如局域网（LAN）、广域网（WAN）、因特网等）连接多个存储控制器，并且一个或多个存储控制器可以实现本发明。

当主机 140 希望更新源存储器 120 中的数据块时，主机 140 将数据写入源高速缓存 124 内的存储块。写入操作同步地修改源高速缓存 124 内的存储块（即，写入的主机 140 等待操作完成），然后，在后台过程中，源高速缓存 124 的内容被写入源存储器 120。写入操作可以更新数据、写入新数据或再次写入相同的数据。将源高速缓存 124 内的数据写入源存储器 120 被称为离台操作。将数据块的全部或一部分从源存储器 120 复制到源高速缓存 124 是登台操作。同样，数据可以在目标存储器 130 与目标高速缓存 134 之间登台和离台。此外，数据可以从源存储器 120 登台到目标高速缓存 134。

图 2 示出了根据本发明的特定实现的各种结构 200、210 和 220。非易失性高速缓存 118 包括目标复制结构 200。目标复制结构 200 可用于判定是否分别将数据从源存储器 120 或目标存储器 130 检索到高速缓存 124 或 134（即，用于登台操作）。此外，目标复制结构 200 可用于判定源存储器 120 中的哪些数据块将被复制到目标存储器 130。目标复制结构 200 包括用于例如卷中的每个数据块的指示符（例如，位）。当指示符被设置成第一个值（例如，1）时，所述设置指示将从源存储器 120 检索数据块以用于登台

操作，或指示数据块将被复制到目标存储器130以用于增量虚拟复制操作。当指示符被设置成第二个值（例如，0）时，所述设置指示将从目标存储器130检索数据块以用于登台操作，或指示不会将数据块从源存储器120复制到目标存储器130以用于增量虚拟复制操作。

源变更记录结构210被用于监视对源存储器120（已为其建立了增量虚拟复制关系）中的数据部分内的数据块的更新。源变更记录结构210包括用于源存储器120（其是所述增量虚拟复制关系的一部分）中的每个数据块的指示符（例如，位）。当指示符被设置成第一个值（例如，1）时，所述设置指示自最后一次增量虚拟复制操作以来所述数据块已被更新。当指示符被设置成第二个值（例如，0）时，所述设置指示自最后一次增量虚拟复制操作以来所述数据块尚未被更新。

目标变更记录结构220被用于在增量虚拟复制关系已被建立之后监视对目标存储器130中的数据块的更新。目标变更记录结构220包括用于目标存储器130（其是所述增量虚拟复制关系的一部分）中的每个数据块的指示符（例如，位）。当指示符被设置成第一个值（例如，1）时，所述设置指示自最后一次增量虚拟复制操作以来所述数据块已被更新。当指示符被设置成第二个值（例如，0）时，所述设置指示自最后一次增量虚拟复制操作以来所述数据块尚未被更新。

在本发明的特定实现中，每个结构200、210和220都包括位图，并且每个指示符都包括位。在每个结构200、210和220中，第n个指示符与第n个数据块对应（例如，每个结构200、210和220中的第一个指示符与第一个数据块对应）。尽管所述结构200、210和220被示为三个独立的结构，但是可以以任何形式对所述结构进行组合而不偏离本发明的范围。在本发明的特定实现中，具有用于每个卷的每个结构的副本。在本发明的特定实现中，具有用于所有卷的每个结构的单个副本。

图3示出了根据本发明的特定实现的在增量虚拟复制过程116中实现的用于更新结构的逻辑。控制开始于方块300，初始地建立增量虚拟复制关系。当在相应的数据部分之间执行了增量虚拟复制操作时，在源存储器

120 中的一个或多个数据部分（例如，源卷）与目标存储器 130 中的相应数据部分（例如，目标卷）之间形成了增量虚拟复制关系。第一增量虚拟复制操作可以例如将一个或多个源卷复制到目标卷。但是，随后的复制可以进行增量复制，避免重新复制自最后一次即刻虚拟复制操作以来尚未改变的源卷的任何部分。

在方块 310 中，增量虚拟复制过程 116 更新目标复制结构 200 中的指示符以指示将从源存储器检索与所述指示符对应的所有数据块以用于登台操作，并且指示将从源存储器向目标存储器复制所有数据块以用于增量虚拟复制操作或物理复制操作。在本发明的特定实现中，目标复制结构 200 中的指示符被设置成 1。

在方块 320 中，增量虚拟复制过程 116 更新源变更记录结构 210 中的指示符以指示自最后一次增量虚拟复制操作以来，与所述指示符对应的源数据块没有被更新。在本发明的特定实现中，源变更记录结构 210 中的所有指示符都被设置成 0。在方块 330 中，增量虚拟复制过程 116 更新目标变更记录结构 220 中的指示符以指示自最后一次增量虚拟复制操作以来，与所述指示符对应的目标数据块没有被更新。在本发明的特定实现中，目标变更记录结构 220 中的所有指示符都被设置成 0。

图 4 示出了根据本发明的特定实现的在增量虚拟复制过程 116 中实现的用于执行增量虚拟复制操作的逻辑。控制开始于方块 400，增量虚拟复制过程 116 接收增量虚拟复制操作。增量虚拟复制操作可以由主机 140 来发布。尽管在图 4 的流程中没有示出，但是在收到增量虚拟复制操作之前，源存储器 120 和/或目标存储器 130 中的一个或多个数据块可以已被例如主机 140 处的用户所更新。

在方块 410 中，增量虚拟复制过程 116 使用源和目标变更记录结构 210 和 220 中的指示符来更新目标复制结构 200 中的指示符。在本发明的特定实现中，使用“或”运算将源变更记录结构 210 与目标变更记录结构 220 合并，并且“或”运算的结果被与目标复制结构 200 相“或”。

在方块 420 中，在目标复制结构 200 已被更新之后，增量虚拟复制过

程 116 更新源变更记录结构 210 中的指示符以指示自最后一次增量虚拟复制操作以来，源数据块没有被更新。在本发明的特定实现中，源变更记录结构 210 中的所有指示符都被设置成 0。在方块 430 中，增量虚拟复制过程 116 更新目标变更记录结构 220 中的指示符以指示自最后一次增量虚拟复制操作以来，目标数据块没有被更新。在本发明的特定实现中，目标变更记录结构 220 中的所有指示符都被设置成 0。

图 5 示出了根据本发明的特定实现的在写入过程 114 中实现的用于处理写入操作的逻辑。控制开始于方块 500，写入过程 114 接收写入数据块的请求。在方块 520 中，写入过程 114 判定数据块是否在增量虚拟复制关系中。如果是，过程继续到方块 530，否则，过程继续到方块 560。在方块 530 中，写入过程 114 判定数据块是否在目标存储器 130 中。如果是，过程继续到方块 540，否则，过程继续到方块 550。

在方块 540 中，目标变更记录结构 220 中用于数据块的指示符被更新以指示自最后一次增量虚拟复制操作以来，目标数据块已被更改。在本发明的特定实现中，目标变更记录结构 220 中的指示符被设置成 1。在方块 550 中，源变更记录结构 210 中用于数据块的指示符被更新以指示自最后一次增量虚拟复制操作以来，源数据块已被更改。在本发明的特定实现中，源变更记录结构 210 中的指示符被设置成 1。在方块 560 中，由写入过程 114 来执行写入操作。

图 6 示出了根据本发明的特定实现的在读取过程 112 中实现的用于处理读取操作的逻辑。控制开始于方块 600，接收读取数据块的请求。在方块 620 中，读取过程 112 判定数据块是否是增量虚拟复制关系中的目标。如果是，过程继续到方块 630，否则，过程继续到方块 660。在方块 630 中，所述读取过程判定是否在目标复制结构中设置了用于所述数据块的指示符以指示将从源存储器 120 读取数据。如果是，过程继续到方块 640，否则，过程继续到方块 650。

在方块 640 中，读取过程 112 从源存储器 120 读取（即，登台）数据块。在方块 650 中，读取过程 112 从目标存储器 130 读取（即，登台）数

据块。在方块 660 中，读取过程 112 执行数据块的正常读取。

图 7 示出了根据本发明的特定实现的后台复制过程。控制开始于方块 700，判定已该复制数据块。在方块 710 中，判定目标复制结构 200 中的用于数据块的指示符是否指示数据块尚未被复制。如果是，过程继续到方块 730，否则，过程继续到 720。在方块 720 中，可以处理下一个数据块或者，如果没有其他将被处理的数据块，则此逻辑终止。

在方块 730 中，根据图 6 的逻辑来读取数据块。在方块 740 中，所述数据块被离台到目标存储器 130。在方块 750，目标复制结构 200 中用于数据块的指示符被更新以指示所述数据块已被复制。在本发明的特定实现中，目标复制结构 200 中的指示符被设置成 0。

因此，在本发明的特定实现中，通过对参与即刻虚拟复制关系的卷的磁道进行监视写入（即更新）和记录变更来完成所述增量虚拟复制操作。在初始的即刻虚拟复制操作之后，源或目标卷上已被更新的磁道可以被从源卷复制到目标卷，而不必复制整个源卷。

Enterprise Storage Server、FlashCopy 和 3990 是国际商业机器公司在美国和/或其他国家地区的注册商标或常用法律标记。

其他实现详细信息

使用生产软件、固件、硬件或它们的任意组合的标准编程和/或工程技术，所述用于增量虚拟复制的技术可以被实现为方法、装置或制品。此处所使用的术语“制品”指在硬件逻辑（例如，集成电路芯片、可编程门阵列（PGA）、专用集成电路（ASIC）等）或计算机可读介质中实现的代码或逻辑，所述计算机可读介质例如磁存储介质（例如，硬盘驱动器、软盘、磁带等）、光存储装置（CDROM、光盘等）、易失性和非易失性存储设备（例如，EEPROM、ROM、PROM、RAM、DRAM、SRAM、固件、可编程逻辑等）。所述计算机可读介质中的代码由处理器来存取和执行。其中实现优选实施例的代码还可以通过传输介质或从网络上的文件服务器来访问。在此情况下，其中实现所述代码的制品可以包括传输介质，例如网络传输线、无线传输介质、通过空间传播的信号、无线电波、红外信号等。

因此，所述“制品”可以包括其中包含所述代码的介质。此外，所述“制品”可以包括其中包含、处理和执行所述代码的硬件和软件组件的组合。当然，本领域的技术人员将认识到，在不偏离本发明的范围的情况下，可以对此配置进行修改，并且所述制品可以包括本领域公知的任何信息承载介质。

图 3-7 的逻辑描述了以特定顺序发生的特定操作。在可替代实现中，可以以不同的顺序执行、修改或删除某些逻辑操作。此外，操作可以被添加到上述逻辑并且仍然符合所述的实现。进而，此处所描述的操作可以顺序地发生或者某些操作可以被并行地处理，或者被描述为由单个过程执行的操作可以由分布式过程来执行。

图 3-7 所示的逻辑可以以软件、硬件、可编程和不可编程门阵列逻辑或硬件、软件或门阵列逻辑的某些组合来实现。

图 8 示出了根据本发明的特定实现的可以使用的计算机系统的体系结构。存储控制器 100 和/或主机 140 可以实现计算机体系结构 800。计算机体系结构 800 可以实现处理器 802（例如，微处理器）、存储器 804（例如，易失性存储设备）以及存储装置 810（例如，诸如磁盘驱动器、光盘驱动器、磁带驱动器之类的非易失性存储区域）。操作系统 805 可以在存储器 804 中执行。存储装置 810 可以包括内部存储设备或附加的或网络可访问的存储装置。存储装置 810 中的计算机程序 806 可以被加载到存储器 804 中并由处理器 802 以本领域公知的方式来执行。所述体系结构还包括允许与网络进行通信的网卡 808。输入设备 812 被用来提供到处理器 802 的用户输入，并且可以包括键盘、鼠标、触笔、麦克风、触敏显示屏或本领域公知的任何其他激活或输入机构。输出设备 814 能够呈现从处理器 802 或其他组件（例如显示监视器、打印机、存储装置等）传输过来的信息。所述计算机系统的计算机体系结构 800 可以包括比示出的组件更少的组件、此处未示出的其他组件或所示组件的某些组合以及附加的组件。

计算机体系结构 800 可以包括本领域公知的任何计算设备，例如，大型机、服务器、个人计算机、工作站、膝上型计算机、手持计算机、电话

设备、网络家电、可视化设备、存储控制器等。可以使用本领域公知的任何处理器 802 和操作系统 805。

出于示例和说明目的给出了对本发明的实现的以上描述。所述描述并非旨在是穷举的或是将本发明限于所公开的精确形式。根据上述教导，许多修改和变化都是可能的。其旨在本发明的范围并非由此详细描述来限制，而是由此后所附的权利要求书来限制。以上说明、实例和数据提供了对本发明的组成部分的制造和使用的完整描述。由于在不偏离本发明的精神和范围的情况下可以做出本发明的许多实现，因此本发明存在于此后所附的权利要求书中。

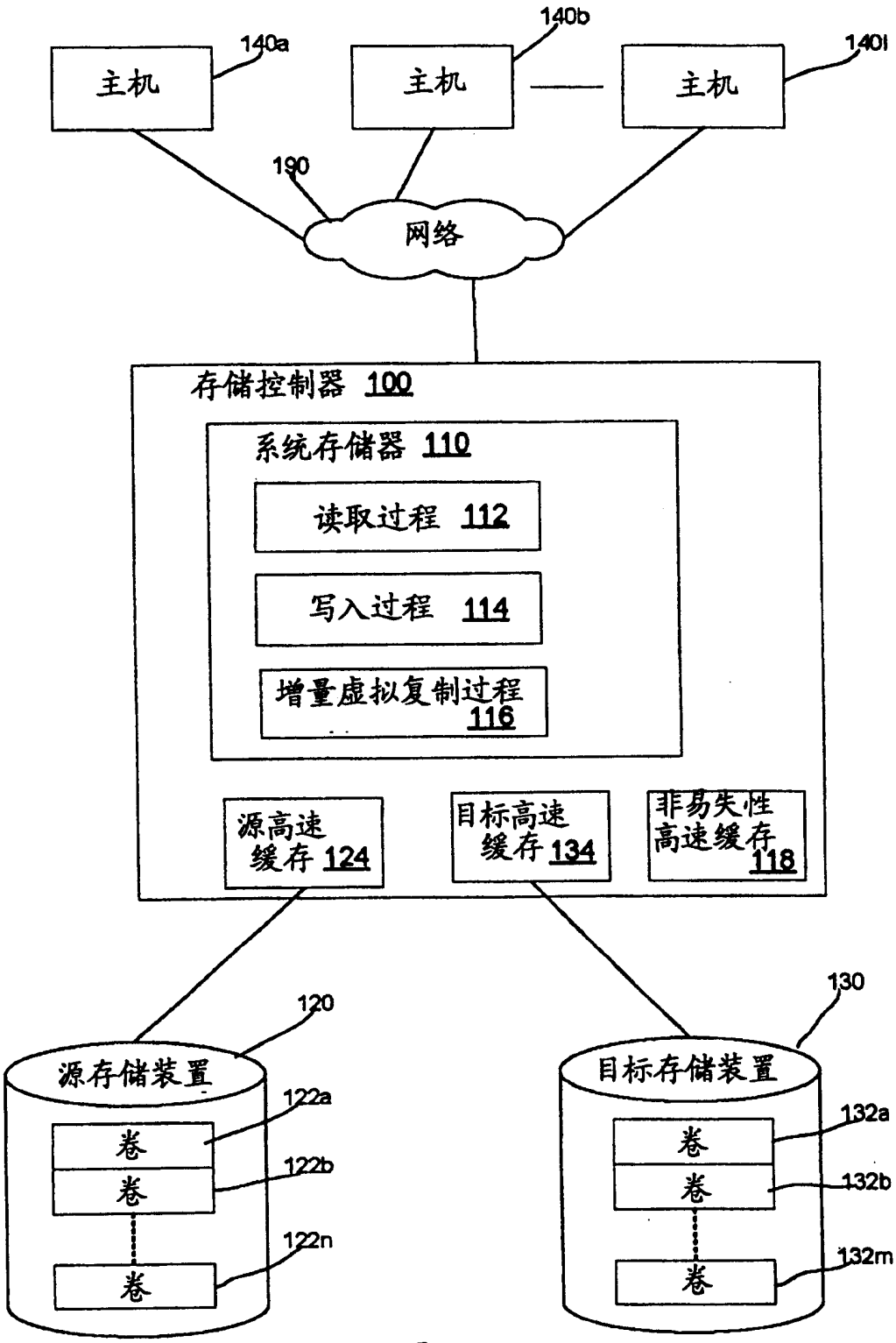

主机
140a
主机
140b
主机
140l
190
网络
存储控制器 100
系统存储器 110
读取过程 112
写入过程 114
增量虚拟复制过程 116
源高速缓存 124
目标高速缓存 134
非易失性高速缓存 118
120
源存储装置
122a
卷
122b
卷
122n
卷
130
目标存储装置
132a
卷
132b
卷
132m
卷

图 1A

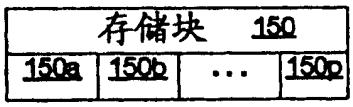

存储块 150
150a
150b
...
150p

图 1B

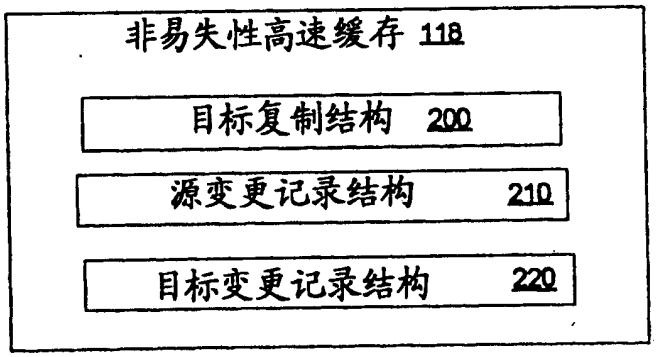

非易失性高速缓存 118
目标复制结构 200
源变更记录结构 210
目标变更记录结构 220

图 2

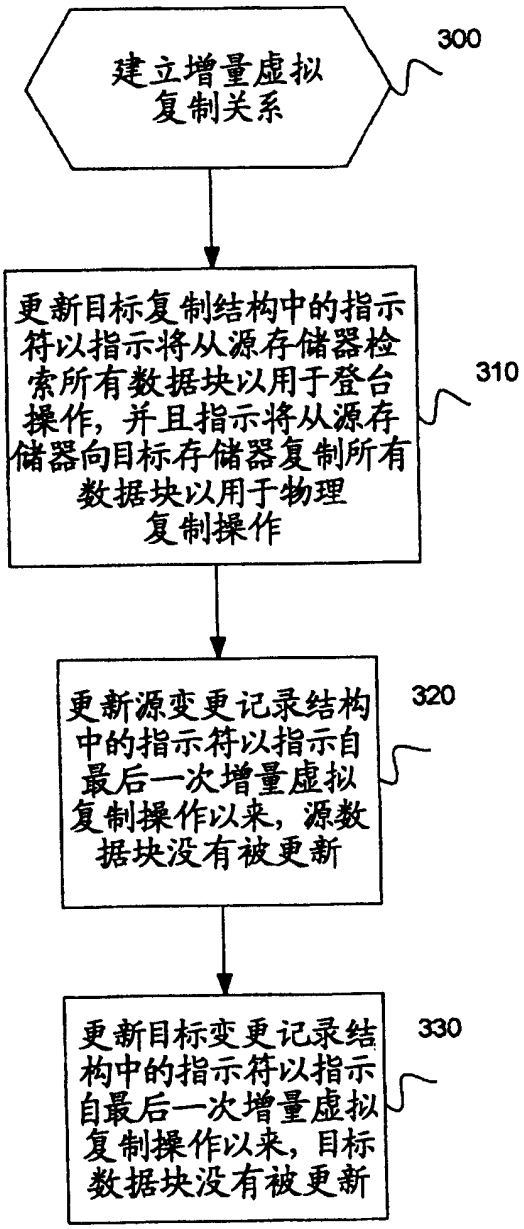

建立增量虚拟复制关系
300
更新目标复制结构中的指示符以指示将从源存储器检索所有数据块以用于登台操作，并且指示将从源存储器向目标存储器复制所有数据块以用于物理复制操作
310
更新源变更记录结构中的指示符以指示自最后一次增量虚拟复制操作以来，源数据块没有被更新
320
更新目标变更记录结构中的指示符以指示自最后一次增量虚拟复制操作以来，目标数据块没有被更新
330

图 3

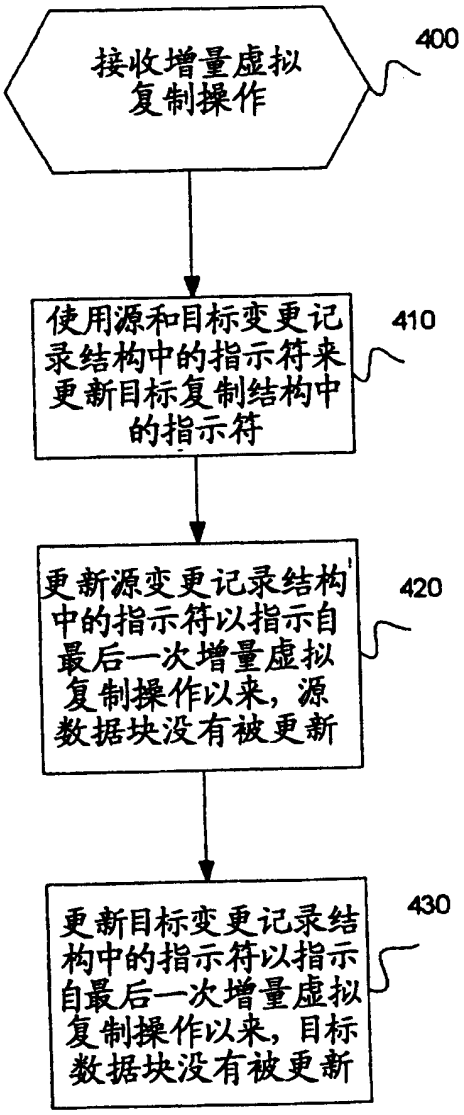

接收增量虚拟复制操作
400
使用源和目标变更记录结构中的指示符来更新目标复制结构中的指示符
410
更新源变更记录结构中的指示符以指示自最后一次增量虚拟复制操作以来，源数据块没有被更新
420
更新目标变更记录结构中的指示符以指示自最后一次增量虚拟复制操作以来，目标数据块没有被更新
430

图 4

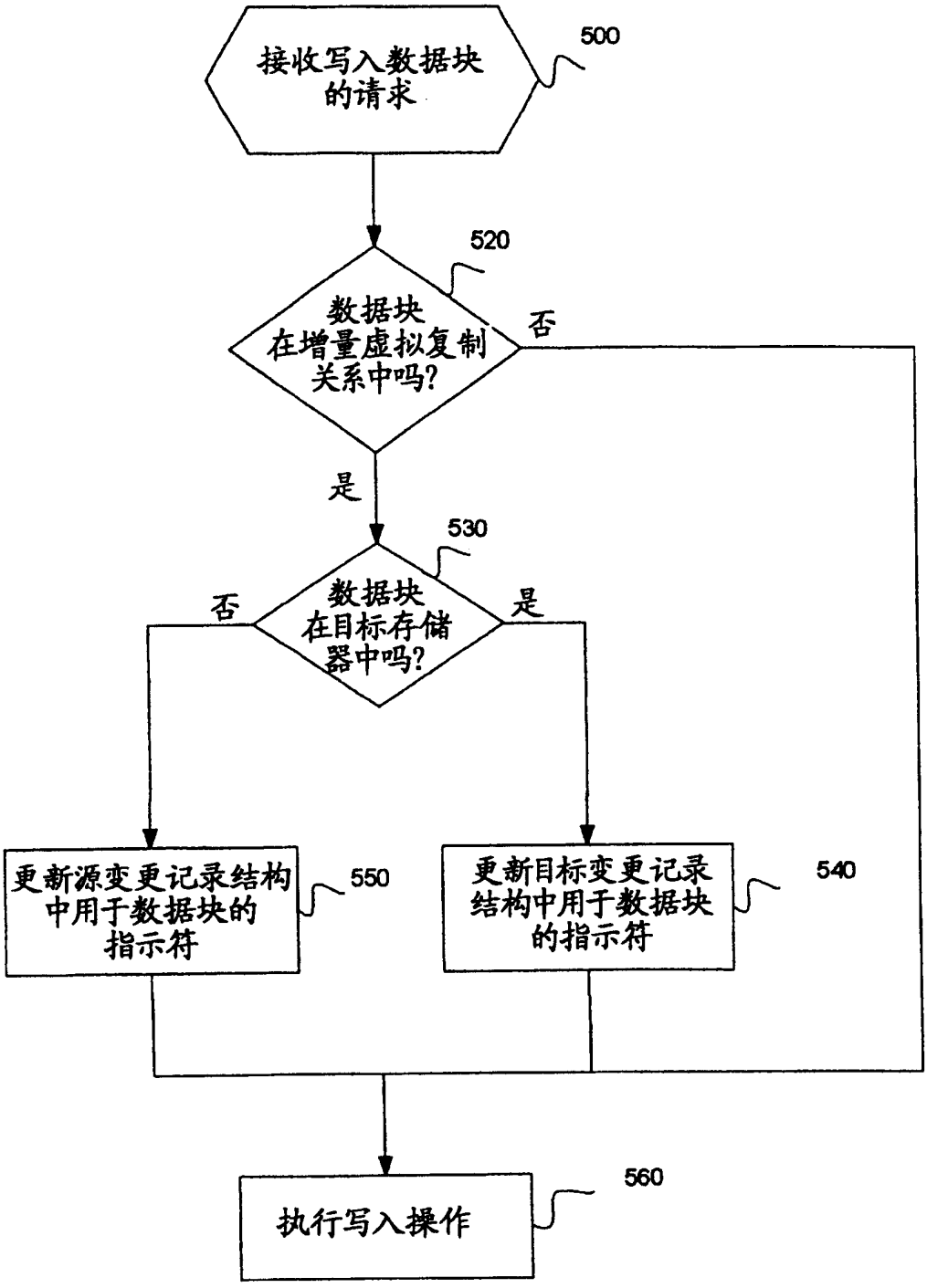

接收写入数据块的请求
500
数据块在增量虚拟复制关系中吗?
520
否
是
数据块在目标存储器中吗?
530
否
是
更新源变更记录结构中用于数据块的指示符
550
更新目标变更记录结构中用于数据块的指示符
540
执行写入操作
560

图 5

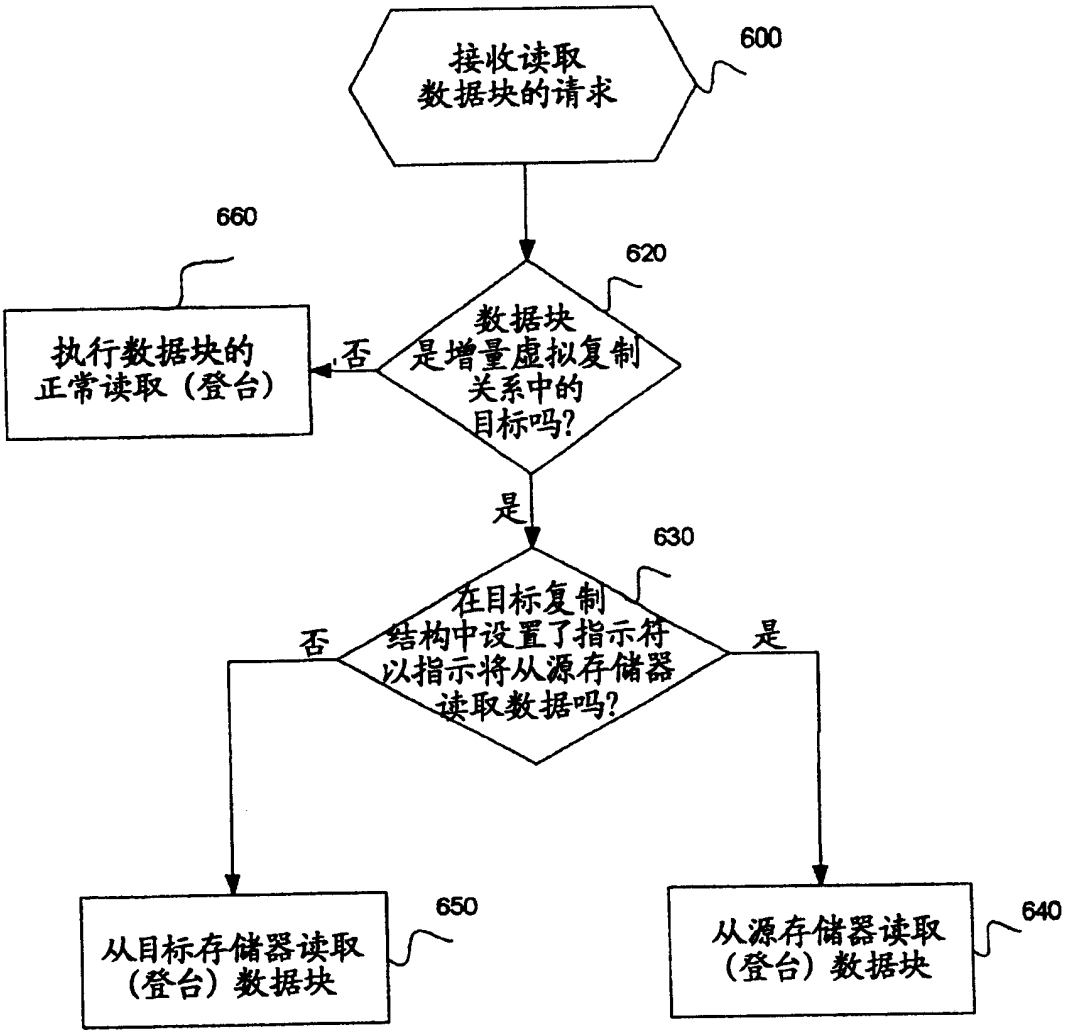

接收读取
数据块的请求
600
数据块
是增量虚拟复制
关系中的
目标吗?
620
否
执行数据块的
正常读取（登台）
660
是
在目标复制
结构中设置了指示符
以指示将从源存储器
读取数据吗?
630
否
是
从目标存储器读取
（登台）数据块
650
从源存储器读取
（登台）数据块
640

图 6

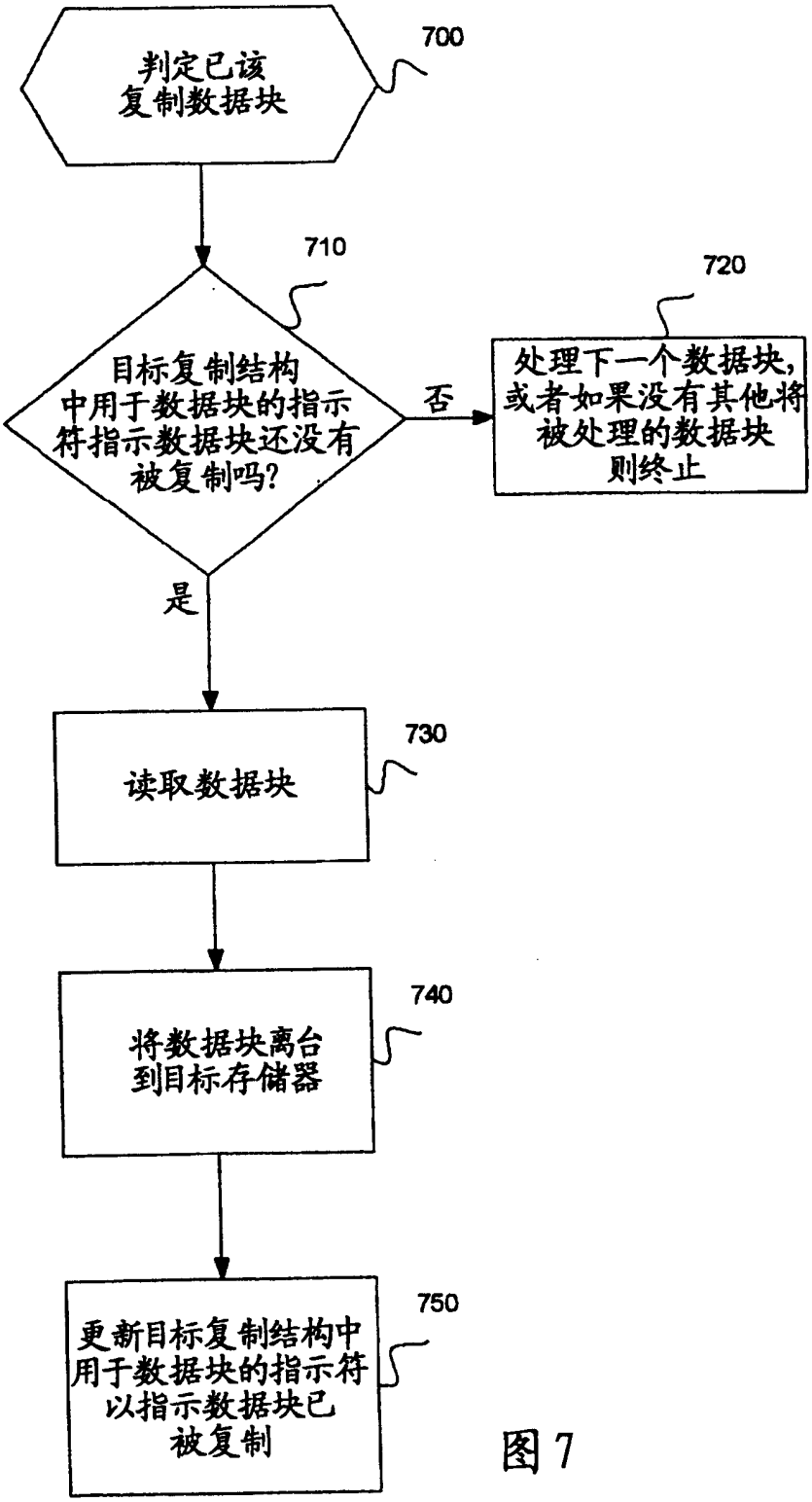

判定已该
复制数据块
700
710
目标复制结构
中用于数据块的指示
符指示数据块还没有
被复制吗?
否
720
处理下一个数据块,
或者如果没有其他将
被处理的数据块
则终止
是
730
读取数据块
740
将数据块离台
到目标存储器
750
更新目标复制结构中
用于数据块的指示符
以指示数据块已
被复制

图 7

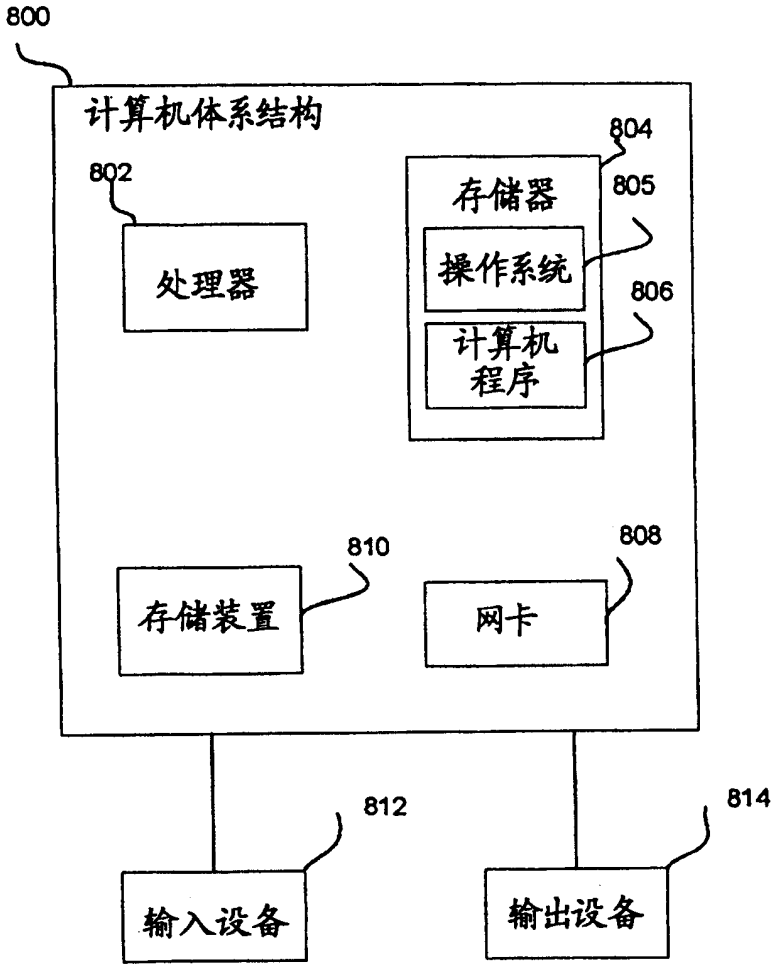

800
计算机体系结构
802
处理器
804
存储器
805
操作系统
806
计算机程序
810
存储装置
808
网卡
812
输入设备
814
输出设备

图 8