



(12) 发明专利

(10) 授权公告号 CN 103026343 B

(45) 授权公告日 2016. 05. 04

(21) 申请号 201180028891. 6

代理人 周亚荣 安翔

(22) 申请日 2011. 05. 03

(51) Int. Cl.

(30) 优先权数据

G06F 11/14(2006. 01)

12/773350 2010. 05. 04 US

G06F 11/20(2006. 01)

(85) PCT国际申请进入国家阶段日

审查员 马婷婷

2012. 12. 11

(86) PCT国际申请的申请数据

PCT/US2011/034909 2011. 05. 03

(87) PCT国际申请的公布数据

W02011/140025 EN 2011. 11. 10

(73) 专利权人 赛门铁克公司

地址 美国加利福尼亚州

(72) 发明人 A·蒂亚姆

(74) 专利代理机构 中原信达知识产权代理有限

责任公司 11219

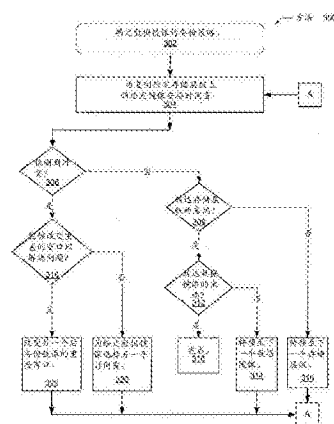
权利要求书2页 说明书10页 附图3页

(54) 发明名称

为数据镜像定义生命周期策略的方法和系统

(57) 摘要

一种用于为一个或多个数据镜像中的每一个定义一个独立生命周期策的系统和方法。一个备份服务器被配置用于选择一个第一时间点值,该第一时间点值指示何时开始将该数据镜像从第一数据存储介质向第二数据存储介质转移。该时间点值对应于一个日程表日期以及一天中的一个小时或多个小时。该备份服务器被配置用于继续为一个分层存储层级中的每个数据存储介质选择一个独立的时间点值,以定义生命周期策略。当检测到该数据镜像向目的数据存储介质的转移完成时,可以进一步配置该备份服务器用于删除该数据镜像在一个源数据存储介质上的副本。进一步地,该备份服务器可以删除两个或更多个生命周期策略之间的时序安排冲突。



1. 一种为一个或多个数据镜像中的每一个定义生命周期策略的方法,该方法包括:
选择在一个第一数据存储介质上存储的一个数据镜像以定义一个第一生命周期;
定义所述第一生命周期,其中所述定义包括:

识别一个第一时间点值,该第一时间点值指示何时开始将该数据镜像从该第一数据存储介质向一个第二数据存储介质转移;并且

识别一个第二时间点值,该第二时间点值指示何时允许将该数据镜像从该第二数据存储介质向一个第三数据存储介质进行初始转移;

在一个图形用户接口中向用户提供一个或多个可利用的时间窗以便将该数据镜像从该第二数据存储介质向该第三数据存储介质转移,其中该一个或多个可利用的时间窗由一个或多个条件来限制;以及

存储对应于所述第一生命周期的数据,其中所述数据包括识别该数据镜像的数据、该第一时间点值、该第二时间点值、该第一数据存储介质、以及该第二数据存储介质。

2. 根据权利要求1所述的方法,其中该数据镜像从该第一数据存储介质向该第二数据存储介质的转移包括一个备份操作,并且该数据镜像从该第二数据存储介质向该第三数据存储介质的转移包括一个复制操作。

3. 根据权利要求2所述的方法,其中该一个或多个条件至少包括该一个或多个可利用的时间窗不会与该备份操作的时间窗重叠多于一个预定阈值。

4. 根据权利要求2所述的方法,其中除了该识别的第二时间点值外,在该复制操作过程中允许该数据镜像的初始转移进一步由以下标准中的至少一项限制:该第二数据存储介质或该第三数据存储介质的当前可用容量、该数据镜像的业务值、该数据镜像的寿命、以及该数据镜像的大小。

5. 根据权利要求2所述的方法,进一步包括检测一个冲突,其中该冲突包括以下情况中的至少一项:两个或更多个生命周期策略包括一个给定数据存储介质的安排好的并行访问;一个给定数据存储介质的可用容量已经达到一个预定阈值;当完成安排好的转移时的一个给定数据存储介质的估计可用容量对应于一个预定的阈值;并且用于一个给定数据镜像的转移的估计持续时间超出了一个可利用时间窗。

6. 根据权利要求5所述的方法,其中响应于检测到所述冲突,该方法进一步包括提供一条对应于该冲突的消息并向用户提供修改对应于一个已有生命周期的一个或多个存储值的机会。

7. 根据权利要求4所述的方法,进一步包括实现响应于检测到一个单独的时间点值而在复制操作过程中转移该数据镜像并且实现任意其他的已识别限制值。

8. 根据权利要求7所述的方法,进一步包括实现响应于从独立源数据存储介质检测到该数据镜像的转移而删除该数据镜像在一个单独源数据存储介质上的一个副本。

9. 一种计算机系统,包括:

一个备份服务器;

一个网络;以及

通过该网络连接至该备份服务器的一个或多个客户端计算机;

其中该备份服务器被配置用于:

使用户能够选择存储在一个第一数据存储介质上的一个数据镜像以定义一个第一生

命周期；

使该用户能够定义所述第一生命周期，其中所述定义包括：

识别一个第一时间点值，该第一时间点值指示何时开始将该数据镜像从该第一数据存储介质向一个第二数据存储介质转移；并且

识别一个第二时间点值，该第二时间点值指示何时允许将该数据镜像从该第二数据存储介质向一个第三数据存储介质进行初始转移；

在一个图形用户接口中向用户提供一个或多个可利用的时间窗以便将该数据镜像从该第二数据存储介质向该第三数据存储介质转移，其中该一个或多个可利用的时间窗由一个或多个条件来限制；以及

存储对应于所述第一生命周期的数据，其中所述数据包括识别该数据镜像的数据、该第一时间点值、该第二时间点值、该第一数据存储介质、以及该第二数据存储介质。

10. 根据权利要求 9 所述的计算机系统，其中该数据镜像从该第一数据存储介质向该第二数据存储介质的转移包括一个备份操作，并且该数据镜像从该第二数据存储介质向该第三数据存储介质的转移包括一个复制操作。

11. 根据权利要求 10 所述的计算机系统，其中该一个或多个条件至少包括该一个或多个可利用的时间窗不会与该备份操作的时间窗重叠多于一个预定阈值。

12. 根据权利要求 10 所述的计算机系统，其中除了该识别的第二时间点值外，在该复制操作过程中允许该数据镜像的初始转移进一步由以下标准中的至少一项限制：该第二数据存储介质或该第三数据存储介质的当前可用容量、该数据镜像的业务值、该数据镜像的寿命、以及该数据镜像的大小。

13. 根据权利要求 10 所述的计算机系统，其中该备份服务器进一步被配置用于检测一个冲突，其中该冲突包括以下情况中的至少一项：两个或更多个生命周期策略包括一个给定数据存储介质的安排好的并行访问；一个给定数据存储介质的可用容量已经达到一个预定阈值；当完成安排好的转移时一个给定数据存储介质的估计可用容量对应于一个预定的阈值；并且一个给定数据镜像的转移的估计持续时间超出了可利用时间窗。

14. 根据权利要求 13 所述的计算机系统，其中响应于检测到所述冲突，该备份服务器被配置用于提供一条对应于该冲突的消息并向用户提供修改对应于一个已有生命周期的一个或多个存储值的机会。

15. 根据权利要求 12 所述的计算机系统，进一步包括实现响应于检测到一个单独的时间点值而在一个复制操作过程中转移该数据镜像并且实现任意其他的已识别限制值。

16. 根据权利要求 15 所述的计算机系统，其中该备份服务器进一步被配置用于实现响应于从一个独立源数据存储介质检测到该数据镜像的转移而删除该数据镜像在该独立源数据存储介质上的一个副本。

为数据镜像定义生命周期策略的方法和系统

技术领域

[0001] 本发明涉及计算机处理领域,更具体地是涉及为数据镜像定义基于时间的生命周期策略。

背景技术

[0002] 随着计算机存储器存储容量和数据带宽的增加,商业和工业中日常管理的数据的数量和复杂度也在增长。可以企业中管理的大部分数据包括为公司或工作群组管理的内容。这种信息可以包括电子文档、公告、任务、讨论主题等。根据信息的商业价值,访问某些信息可能是安全的和私有的,而其他信息可以具有更广泛的甚至公开的访问权。

[0003] 因为存储的数据的不同实例可能经常包含重要信息,可以将信息服务器和终端用户系统的内容备份至备份存储器以防止数据丢失。在这种情况下,专用服务器和存储子系统可以包括一个备份系统。备份系统试图减少数据量以及备份数据的时间以保护机器。一些方法使用去重复处理以减少备份的数据量,但这种方法仍然耗费实施的时间和网络带宽。例如,通过网络在终端用户的膝上计算机上进行一次计划的备份(甚至是一次增量备份)也需要耗费时间来扫描文件并检测变化的文件。在使用去重复操作或其他方法之前,文件的扫描和检测耗费了大量时间。

[0004] 除了在进行备份时要耗费大量时间外,管理和支持大量终端用户系统上的信息技术(IT)基础设施也提出了重大挑战。IT 管理员可以开发一种管理系统以用于数据备份和数据保护;数据恢复、还原、和重启;归档和长期保存;以及管理存储装置体系结构所必需的日常处理和程序。这种可以称为信息生命周期管理(ILM)的管理系统指的是管理计算装置上的存储系统的一组策略。ILM 的操作方面包括将某些策略在信息的整个有效期中应用于该信息的高效管理。

[0005] 采用 ILM 允许用户自动地将备份镜像从一个存储层级转移到另一个。当前的方法采用尽可能快(ASAP)的方式、可用容量的方式、或数据分类的方式来确定何时进行备份镜像的复制。然而,采用 ASAP 方式存在问题,原因是相应的存储资源由另一个备份或其他操作所使用。至于可用容量方式,同样是另一个操作可能正在访问存储资源。另外,或许很难确定提供一致结果的磁盘空间的可用容量。如果将一个预定的限制(例如高水准值)设置得太高,那么可用存储空间不足可能导致一个或多个备份操作失败。如果将该预定限制(例如低水准值)设置得太低,那么一些可用存储空间可能不会得到充分利用。对于基于数据分类的解决方案,数据可能会过早地移至另一个存储层级,这延长了恢复操作,因为数据位于其他存储层级中。其他时间,数据可能过晚地转移至另一个存储层,这导致当前存储层级变得近乎拥挤。目前,备份操作在持续时间上变长并且容量成本增加。

[0006] 鉴于以上情况,需要一些改进的系统和方法,以便为数据镜像定义基于时间的生命周期策略。

发明内容

[0007] 在设想了为数据镜像定义基于时间的生命周期策略的系统和方法。

[0008] 在一个实施方案中,一个计算机系统包括连接至多个数据存储介质的一个备份服务器和连接至一个数据存储介质的至少一个客户端计算机,该数据存储介质包含在该多个数据存储介质中。该数据存储介质存储对应于该客户端计算机的数据镜像的备份副本。该备份服务器被配置用于选择一个第一时间点值,该第一时间点值指示何时开始将该数据镜像转移至该多个数据存储介质的一个第二存储介质。该时间点值可以对应于一个日程表日期和一天中的一个小时或多个小时。在一个实施方案中,该第二数据存储介质与一个分层存储结构中的一个层级相关联,该层级比该第一数据存储介质更低。例如,该第二数据存储介质可以比该第一数据存储介质更少地受到访问。该第二数据存储介质可用于数据恢复操作,而不是工作期间的日常访问。

[0009] 该第二存储介质可以不对应于该分层存储层级内的一个最后存储介质。该备份服务器可以被配置用于为该分层存储层级内的每个数据存储介质连续选择一个各自的时间点值,以便定义该生命周期策略。该备份服务器可以被进一步配置用于在检测到该数据镜像向一个目的数据存储介质的转移完成时,删除该数据镜像在一个源数据存储介质上的副本。进一步地,该备份服务器可以检测时序安排冲突。定义一个时序安排冲突的标准可以包括以下内容中的至少一个:与其他生命周期策略相关联的一天中的日程表信息和小时数以及其他时间增量、数据镜像和对应的备份镜像中的每一个的数据类或业务值、这些存储介质中每一个的当前可用容量、当完成计划的每个生命周期的每个备份/复制操作的计划性转移时存储介质中每一个的预计的(估计的)可用容量、每个存储介质的每个数据格式中的每个数据镜像的大小、以及每个存储介质上的备份/复制操作的数据速率。

[0010] 这些以及其他实施方案将参见以下描述和附图得到理解。

附图说明

[0011] 图 1 的通用框图展示了信息生命周期的一个实施方案。

[0012] 图 2 的通用框图展示了基于时间的生命周期时序安排的一个实施方案。

[0013] 图 3 的流程图展示了一种方法用于为数据镜像定义基于时间的生命周期策略的另一个实施方案。

[0014] 图 4 的通用框图展示了一种网络体系结构的一个实施方案。

[0015] 尽管很容易做出各种修改和更改,但通过实例在附图中示出了具体实施方案并且在本文详细描述。然而应当理解的是,附图以及对其详细描述无意将本发明限制于所披露的具体形式,相反的是,本发明意在涵盖所附权利要求定义的本发明精神和范围之内的所有修改形式、等同体以及更改形式。

具体实施方式

[0016] 在以下描述中,给出了很多具体细节以提供对本发明的透彻理解。然而,本领域普通技术人员应当认识到可以在没有这些具体细节的条件下实施本发明。在一些实例中,未详细示出众所周知的电路、结构、信号、计算机程序指令、以及技术,以避免混淆本发明。

[0017] 参见图 1,图中示出了信息生命周期 100 的一个实施方案的通用框图。信息生命周期 100 或简单生命周期 100 展示了从创建和初始存储到数据变为过时并被删除时的信息系

统的数据和相关元数据的流动的管理。数据镜像可以包括信息、相关元数据以及相关存储安排信息（例如目录、文件夹以及子文件夹）中的一个或多个文件。数据镜像 150 可以存储在数据存储介质 110 中。在一个实施方案中，数据镜像 150 最初在时刻 t_1 存储在数据存储介质 110 上。在另一个实施方案中，可以选择时刻 t_1 作为生命周期 100 的初始时间点，尽管在时刻 t_1 之前数据镜像 150 已经在数据存储介质 110 中存储了一段时间。

[0018] 在一个实施方案中，数据存储介质 110 可以是与一个文件系统相关联的磁盘驱动器，该文件系统用于存储一个或多个客户端计算机的数据。计划的备份操作可以创建数据镜像 150 的副本以存储在独立的存储介质上，例如数据存储介质 120。快照镜像 152 可以代表数据镜像 150 的副本。快照镜像 152 中的一个或多个文件可以包括与用于存储数据镜像 150 的格式不同的一种文件格式。备份操作可以是由信息技术 (IT) 管理员预先确定的计划事件。创建快照镜像 152 的备份操作可以安排在可利用的特定时间窗内，其中确定在该可利用的特定时间窗内并未出现数据存储介质 120 的并发访问。例如，可以确定当出现备份操作时另一个生命周期策略（未示出）未同时访问数据存储介质 120。在一个实施方案中，在创建快照 152 之后，可以从数据存储介质 110 删除数据镜像 150 或使其在数据存储介质 110 中失效。根据 IT 管理员的安排，这种删除可以立即发生或在快照创建之后的一个预定时刻发生。

[0019] 对于数据镜像 150 的后续转移，可以重复上述过程。继续介绍生命周期 100 中所示的数据流动，在图 1 中示为复制 1 的一个预定后续复制操作可以从数据存储介质 120 向数据存储介质 130 转移快照镜像 152。在完成初始备份操作后，这种后续复制操作可以产生预定的延时，在图 1 中该延时示为延时 1。可以对在快照镜像 152 中存储的数据进行拷贝并在图 1 中由复制镜像 154 来表示。如本文所用，术语“复制”指代在数据镜像的初始备份操作之后进行的拷贝操作，其中复制操作在快照镜像或快照镜像的另一个后续副本上执行。

[0020] 创建复制镜像 154 的复制操作可以安排在可利用的特定时间窗内。例如，出现在时刻 t_3 和时刻 t_4 之间的时间窗是第一复制操作可利用的时间窗。在一个实施方案中，系统管理员可以利用包含图形用户接口 (GUI) 的备份操作来选择已识别的时间点值 t_3 和 t_4 。给定的已识别时间点值可以由一个或多个条件来限制。否则，可能会出现冲突，GUI 中的一条消息可以呈现给系统管理员以提供通过修改一个或多个存储值（对应于已存在的生命周期策略）来排除冲突的机会。例如，用于第一复制操作（在图 1 中示为复制 1）的时间窗的一个条件可以是该时间窗不会与备份操作的时间窗重叠多于一个预定阈值。在一个实施方案中，该条件不允许任何重叠。在另一个实施方案中，该条件可以允许重叠半个小时或其他选择的持续段。阈值可以表示备份操作和第一复制操作并行访问数据存储介质 130 的允许值（如果存在的话）。另外，可以出现针对其他数据镜像（未示出）的其他备份和复制操作。给定数据存储介质的并行访问可以减缓备份和复制操作二者，从而足以防止这些操作在预定的可接受持续时间内完成。

[0021] 除了上述重叠条件外，可以使用其他条件确定是否已经出现了一个时序安排冲突。例如，可以确定用于第一复制操作的两个或更多个时间窗（对应于两个或更多个数据镜像）被允许重叠。时刻 t_3 和时刻 t_4 之间的时间窗可用于快照镜像 152 和未示出的另一个快照镜像。时间点值 t_3 可以表示允许快照镜像何时最初从数据存储介质 120 转移至数据存储介质 130。首先选择哪一个快照镜像进行重复可以基于除已识别的时间点值 t_3 之外

的其他条件。这些其他条件可以包括以下内容中的至少一项：源数据存储介质或目的数据存储介质的当前可用容量、相应的数据镜像的业务值、数据镜像的寿命、以及数据镜像的大小。其他条件是可能的并且可以设想。类似于删除数据镜像 150，在创建复制镜像 154 之后可以从数据存储介质 120 删除快照镜像 152 或使其在数据存储介质中失效。在创建复制镜像 154 之后，这种删除可以立即或持续预定时间发生。

[0022] 上述转移过程可以重复多次（由系统管理员定义）。例如，图 1 中示为复制 2 的另一个复制操作可以将复制镜像 154 从数据存储介质 130 转移至数据存储介质 140。在完成前一次复制操作之后，这种后续复制操作会产生预定的延时，在图 1 中以延时 2 示出。可以对在复制镜像 154 中存储的数据进行备份并且由复制镜像 156 来表示。类似于删除快照镜像 152，在创建复制镜像 156 之后可以从数据存储介质 130 删除复制镜像 154 或使其在数据存储介质中失效。在创建复制镜像 156 之后，这种删除可以立即或持续一个预定时间发生。在一个实施方案中，数据存储介质 140 可以是在将数据镜像 150 发送至站外存储位置、站上磁带存储介质等之前的数据镜像 150 的生命周期中的一个最终数据存储介质。

[0023] 数据存储介质 110-140 可以代表分层存储器。分层存储器可以包括为不同类型的存储介质分配不同的数据类，以减少总的存储成本。数据类可以基于以下内容中的至少一项：预定的数据保护级别、系统性能目标、预定的业务值、以及数据的使用频率。在软件中可以采用一项或多项预定的策略来自动地为具体的介质分配数据。

[0024] 分层存储器的实例包括用于关键任务或最近访问文件的存储器的第一层级。第一层级存储器可以在昂贵的和高质量的介质（例如，双重奇偶校验 RAID（独立磁盘冗余阵列））上存储数据。分层存储器可以包括用于经济和分类文件的第二层级。第二层级可以在传统存储区域网络（SAN）的不那么昂贵的介质上存储数据。当分层储存系统中的层级数量增加时，数据的业务值和相应存储器的成本都可以降低。例如，分层存储器可以包括第三层级，该层级用于事件驱动的较少使用的或未分类的文件的存储。第三层级可以在可记录的紧凑光盘（CD-R）或磁带上存储数据。选择用于数据存储介质 110-140 的存储类型可以基于给定的数据镜像 150 的数据类。

[0025] 现在转到图 2，展示了基于时间的生命周期时序安排 200 的一个实施方案的通用框图。图中为四个数据镜像中每一个展示了一个生命周期的一部分。以一个数据镜像为例，例如镜像 2，可以看出备份操作被安排在时刻 t30 开始并在时刻 t30 到时刻 t32 之间持续。备份操作包括对数据存储介质（称为介质 1）的内存写访问。在时刻 t33 开始一个后续备份操作之前，出现了从时刻 t32 到时刻 t33 的预定延时。在一个实施方案中，如本文所用，术语“备份”是指一种处理，通过这种处理可以拷贝数据集，然后转换至另一种格式并进行存储（例如，以用于可能的未来恢复）。在一个实施方案中，术语“复制”是指一种处理，通过这种处理可以对数据集进行拷贝并存储，而无需改变格式。可替代地，复制可以承担将数据集从一个位置转移到另一个位置，无需在源位置上保留一个副本。在其他实施方案中，备份可以不承担格式的改变，或者复制可以承担某种格式改变。很多这种替代的实施方案是可能的并且可以设想。第一复制可以在时刻 t33 至时刻 t36 之间持续。该第一复制操作包括介质 1 的内存读访问和介质 4 的内存写访问。数据存储介质（介质 1 和介质 4）可以包含在分层存储系统的不同层级中。

[0026] 如图 2 中所示，继续介绍与数据镜像（镜像 2）有关的实例，在时刻 t36 和时刻 t38

之间出现第二预定延时。第二复制可以从时刻 t_{38} 至时刻 t_{41} 之间持续。第二复制操作包括介质 4 的内存读访问和介质 8 的内存写访问。数据存储介质（介质 4 和介质 8）可以包含在分层存储系统的不同层级中。此后还可以出现一个或多个延时和复制操作，它们并未示出。否则，在一个预定延时后，可以从分层存储器中删除数据镜像（镜像 2）。对于数据镜像（镜像 1、镜像 3 和镜像 4），示出了类似的操作和延时，尽管它们出现在图 2 所示的不同时刻上。

[0027] 计划的时间窗被示为在标记的时间点之间的持续时间。例如，镜像 1 的复制操作的时间窗在时刻 t_{31} 和时刻 t_{34} 之间。用于镜像 2 的备份操作的计划时间窗在时刻 t_{30} 和时刻 t_{32} 之间。可替代地，时间窗可以由开始于给定起始时刻的一段持续时间来指定，该给定开始时刻对应于日程表日期和一天中的一个小时或多个小时。例如， t_{31} 和 t_{34} 之间的持续时间可以是 2 个小时。时刻 t_{31} 可以对应于 3 月 14 的上午 5 点的时刻。因此，可以指定镜像 1 的复制操作开始于 3 月 14 的上午 5 点并结束于 3 月 14 的上午 7 点。所以，2 个小时的时间窗被指定出现在具体的日程表日上。

[0028] 图 2 中所示的时序安排信息可以在图形用户接口 (GUI) 上呈现给 IT 管理员。附加信息（未示出，以简化图示）可以在 GUI 中呈现给用户，这些信息例如是以下标准中的至少一项：日程表信息、一天中的多个小时或其他时间增量、数据镜像和对应的备份镜像中每一个的数据类或业务值、存储介质中的每一个的当前可用容量、当完成计划的每个生命周期的每个备份 / 复制操作的计划转移时存储介质中每一个的预计的（估计的）可用容量、每个存储介质内每个数据格式中的每个数据镜像的大小、每个存储介质上的备份 / 复制操作的数据速率、数据存储介质的相对或绝对经济成本、以及估计的加速窗口，其中备份 / 复制操作预计比计划的时间窗更快地完成（例如，对于镜像 2，预计复制可以在时刻 t_{33} 和时刻 t_{35} 之间的加速窗中完成，而不是在时刻 t_{33} 和时刻 t_{36} 之间的计划窗中完成）。除了为用户呈现以上标准外，可以使用以上标准检测时序安排冲突。以下描述时序安排冲突的检测。

[0029] 时序安排冲突的检测可以包括通过两个不同的生命周期策略至少识别访问给定存储介质的周期。例如，参考图 2 所示的时序安排 200，在时刻 t_{31} 和时刻 t_{32} 之间，对于镜像 1，标记为介质 1 的存储介质具有复制操作的内存读访问操作，而对于镜像 2，具有备份操作的内存写访问操作。并行访问可以延迟对应的访问中每一个的起始时刻并且还使得每个访问的各自持续时间延长。因此，操作可能放慢速度，从而在计划的时间窗内完全有可能不会完成。在此，镜像 1 的复制操作的计划时间窗位于时刻 t_{31} 和时刻 t_{34} 之间。镜像 2 的备份操作的计划时间窗位于时刻 t_{30} 和时刻 t_{32} 之间。

[0030] 冲突的检测还可以包括识别日常业务操作访问给定存储介质的周期。例如，可以不允许在工作日的上午 8 点和下午 6 点之间安排备份操作。用户访问将具有备份操作的内存读访问权的各个数据存储介质。类似地，可以不允许在这些办公时间期间安排第一复制操作。用户可以访问各个数据存储介质以进行恢复操作，这些恢复操作将具有复制操作的内存读访问权。在 GUI 的窗格中这些时间窗可以显示为阻断状态，以便向 IT 管理员表示在这些时间内并未安排操作。

[0031] 当为给定数据镜像的给定生命周期策略安排备份 / 复制操作时，删除被提议的时间窗的时序安排冲突还可以包括先前描述的呈现标准。例如，被提议的时间窗可能不会与

给定的数据存储介质的较早安排的生命周期的时间窗相重叠,但该给定数据存储介质可能已经达到了容量高水位值。利用上述标准进行检查的其他组合形式是可能的并且可设想。当自动检测到冲突并提交给 IT 管理员时,通过 GUI 的窗格,为数据镜像的生命周期安排备份 / 复制操作可以变得不那么繁重。

[0032] 在给定数据镜像的生命周期策略的定义阶段,通过 GUI 的基于时间的方法可以为检测到的冲突另外提供可能的解决方案。例如,软件应用程序可以为较早安排的操作的时间窗或数据存储介质确定一个新值。这种较早的备份 / 复制操作可能正在访问与被提议的备份 / 复制操作相同的数据存储介质,因而产生冲突。新值可以仍然允许较早的操作成功地完成并且当前允许安排已提议的操作。例如,用户可能以已经安排较早的操作持续 4 个小时。软件应用程序可以通过备份镜像大小和数据存储介质的数据速率来确定较早操作可能持续 3 个小时。因此,可以提前或延后一个小时来安排较早的操作,提前或延后依赖于哪个方向会解决冲突。较早操作的任意新值可以通过 GUI 提交给用户。

[0033] 继续描述为冲突找到解决方案,类似于上述解决方案,软件应用程序可以为时间窗或计划的备份 / 复制操作的数据存储介质确定新值。利用用户的选择,应用程序可以进行计算以确定是否一个或多个新值是可能的,并且如果是这样,通过 GUI 向用户提供新值。作为输入,计算还可以包括先前为检测冲突而列出的标准。

[0034] 如以上所述,在通过 GUI 向 IT 管理员提供信息方面,为数据镜像定义基于时间的生命周期策略可以包括多个时间窗。当确定信息生命周期时,时间窗可以与数据类、可用容量、数据速率、以及其他方法相组合。通过软件应用程序提供的并通过 GUI 或命令行接口呈现的时间窗允许用户确切地指定对给定数据镜像何时执行备份 / 复制操作以及在哪个数据存储介质上执行。

[0035] 用户可用利用基于日程表的方法并在备份应用程序(例如, Symantec NetBackup®)中指定各个备份 / 复制操作的日期和时间窗。备份应用程序可以将基于时间的计划表与每个目标或数据存储介质相关联或附加至它们。以此方式,当用户创建了包括至少一个快照目标和一个备份 / 复制目标的数据存储生命周期策略时,将一个计划表附加至各个备份 / 复制操作。计划表指定了何时进行这些操作中的每一个。例如,在创建快照镜像之后,复制操作可以在两周之后进行,以便将该快照镜像转移至磁盘存储器。在这两周中,可以利用该快照镜像进行数据恢复。所希望的是可以快照在两周之后失效并删除它。因此,并不保留各个数据镜像的多个副本。在复制操作中定义的第二计划表可用于将磁盘上的各个备份镜像转移至磁带,以便长期保存。该第二计划表可以是一个月后到期。

[0036] 参见图 3,图中示出了一种用于为数据镜像定义基于时间的生命周期策略的方法 300 的一个实施方案的概览。为了进行讨论,将对应于网站来描述数据。然而,本文描述的方法和机制还应用于其他的非网站的数据。在上述计算机系统 100 中实例化的组件通常可以根据方法 200 来操作。为了进行讨论,本实施方案中的步骤以顺序次序示出。然而,一些步骤可以采用不同于所示步骤的顺序出现,一些步骤可以同时执行,一些步骤可以与其他步骤组合,而一些步骤可以不出现在另一个实施方案中。

[0037] 在模块 302,为数据镜像确定生命周期内的备份和复制操作的策略。例如,这些策略可以包括确定分层存储器中的层级数量、存储介质的类型和数量、镜像存储在给定存储介质上的时间长度、每天中屏蔽特定操作的小时等等。在模块 304,用户可以提议基于时间

的计划表,以便对给定存储层级上的给定数据镜像进行备份/复制操作。该计划表可以包括指定时间点(例如日程表日期、一天中的一个小时),或在对应的数据镜像的相应前一次转移之后的预定持续时间。为了详细说明所提议的计划表,用户可以在 GUI 中选择窗口、在 GUI 中选择并拖拽窗口、在命令行上输入值或类似操作。时序安排冲突的检测可以使用先前描述的标准。

[0038] 如果对于提议的基于时间的计划表未检测到冲突(条件模块 306),并且未到达用于信息生命周期的存储层级的末端(条件模块 3308),那么在模块 310,可以选择生命周期中的下一个存储层级。当前安排的数据镜像的信息生命周期可以具有已提议的基于时间的计划表(与相应的数据存储介质相关联)。然后,方法 300 的控制流通过模块 A 返回至模块 304。在模块 304,可以输入下一个提议的基于时间的计划表。如果到达了用于信息生命周期的存储层级的末端(条件模块 308),并且如果存在更多数据镜像需要安排(条件模块 312),那么在模块 314,选择下一个数据镜像。当前安排的数据镜像的信息生命周期可以具有提议的基于时间的计划表(与相应的数据存储介质相关联)。然后,方法 300 的控制流通过模块 A 返回至模块 304。在模块 304,可以输入下一个提议的基于时间的计划表。如果没有更多数据镜像需要安排生命周期(条件模块 312),那么生命周期的基于时间的时序安排在模块 316 完成。可替代地,方法 300 的控制流可以返回至条件模块 306。例如,IT 管理员可以安排无冲突的多个生命周期策略。然而,在备份服务器上运行的应用程序未检测到的冲突可能稍后出现。尽管已经没有新的时序安排出现,但冲突可能仍然发生,那么可以修改一个或多个已存在的生命周期策略来排除冲突。

[0039] 如果针对提议的基于时间的计划表检测到了冲突(条件模块 306),并且确定先前冲突的备份/复制操作已经使它的计划表在不导致其他冲突的情况下成功地发生改变(条件模块 318),那么在模块 322,已确定的变化可以发生。变化的确定可以如先前描述的那样出现。可以进行修改或加速较早的备份/复制操作的信息生命周期以提供与相应的数据存储介质相关联的新的定义。另外,当前安排的数据镜像的信息生命周期可以具有提议的基于时间的计划表(与相应的数据存储介质相关联)。然后,方法 300 的控制流通过模块 A 返回至模块 304。在模块 304,可以输入下一个提议的基于时间的计划表。

[0040] 如果针对提议的基于时间的计划表检测到了冲突(条件模块 306),并且确定先前冲突的备份/复制操作并未使它的计划表在不导致其他冲突的情况下成功地发生改变(条件模块 318),那么在模块 320,改变当前数据镜像的已提议的计划表和当前数据存储介质。提出的改变可以利用先前描述的计算来提供。否则,可以例如通过 GUI 向用户提供系统的标准,以使用户调整提议的基于时间的计划表。然后,方法 300 的控制流通过模块 A 返回至模块 304。在模块 304,可以输入下一个提议的基于时间的计划表。

[0041] 现在转到图 4,图中示出了网络架构 400 的一个实施方案。网络架构 400 包括应用程序计算机系统 422 和客户端计算机系统 420,这些系统通过网络 480 彼此互联,连接至文件服务器 490,并且连接至备份服务器 450。磁盘存储器 460 和磁盘存储器 462 连接至备份服务器 450。网络 480 可以包括各种技术,这些技术包括无线连接、直接局域网(LAN)连接、广域网(WAN)连接(例如互联网)、路由器和其他装置。网络 480 可以包括一个或多个 LAN,这些 LAN 也可以是无线的。网络 480 可以进一步包括远程直接内存访问(RDMA)硬件和/或软件、传输控制协议/互联网协议(TCP/IP)硬件和/或软件、路由器、中继器、交换

机、网络、和 / 或其他设备。

[0042] 在网络架构 400 的各种实施方案中,应用程序计算机系统 422 表示任意数量的托管重要业务应用程序的服务器和数据库服务器。客户端计算机系统 420 表示任意数量的静态或移动计算机,例如服务器、桌上 PC、膝上计算机、手持式计算机等。客户端 420 可以包括至少一个单一处理器 402 和一个存储介质 430。客户端 420 可以使用处理器 402 来执行软件应用程序的指令。客户端 420 在其存储介质 430 上可以包含一个或多个软件应用程序。例如,客户端 420 可以包括备份代理 414。类似地,客户端 420 可以包括虚拟化代理 412,以用于在客户端中创建和维护虚拟环境。

[0043] 在一些实施方案中,系统 422 和 420 可以包括多个操作环境,例如操作系统(如基础 OS 416 和 / 或虚拟 OS 层 436)。在不同的实施方案中,所提供的基础 OS 416 可以是不同操作系统中的任意一种,这些操作系统包括

MS-DOS®、MS-WINDOWS®、OS/2®、UNIX®、Linux®、Solaris® 或另一种已知的操作系统。在一些实施方案中,系统 422 和 420 只具有虚拟 OS 层 436,该虚拟 OS 层由 IT 管理员利用该层的公共镜像来提供。在其他实施方案中,系统 422 和 420 同时具有基础 OS 416 和虚拟 OS 层 436。操作系统可以在系统硬件和软件应用程序之间对接。

[0044] 继续介绍系统 422 和 420 的组件,还可以包含文件系统 430。类似于文件服务器 490、备份服务器 450 或其他计算机系统或服务器上的文件系统,文件系统 440 可以作为专用数据库进行操作以进行数据的存储、组织、操控和检索。用户可以通过文件系统 440 浏览存储在相应的数据存储介质上的目录。

[0045] 网络架构 400 中的文件服务器 490 可以是配置用于存储数据的任意种类的服务器。在一个实施方案中,文件服务器 490 可以是网络附加存储(NAS)装置。文件服务器 490 可以是专用于向计算机系统 422 和 420 以及其他信息服务器(未示出)提供文件存储服务的计算机。文件服务器 490 上的操作系统和其他软件提供文件系统的功能、文件访问、以及对这些功能的管理。文件服务器 490 可以包括一个或多个硬盘,这些硬盘常常设置在逻辑冗余存储容器或者 RAID 阵列(廉价 / 独立磁盘的冗余阵列)中。数据 492 可以存储在文件服务器 490 的硬盘中的一个或多个上。

[0046] 文件服务器 490 可以在文件系统级别上提供数据存储。文件服务器 490 可以使用基于文件的协议,例如网络文件系统(NFS)(在 UNIX 系统上普及)、SMB/CIFS(服务器消息块 / 公共互联网文件系统)(与微软 Windows 系统一起使用)、或 AFP(与苹果 Macintosh 计算机一起使用)。在图 4 的实例中,文件服务器 490 上的文件可以通过包含在网络 480 中的网络文件系统(NFS)来访问。

[0047] 现在转向备份存储装置 460,装置 460 可用于存储网络架构 400 中的数据的备份副本。备份数据 494 可以存储数据 492 的副本、包含在计算机系统 422 和 420 中的数据的副本、以及包含在其他信息服务器(未示出)中的数据的副本。备份存储装置 460 可以包括或进一步连接至多种存储器,这些存储器由一个或多个硬盘、磁盘驱动器、刀片服务器、或专用装置组成,并且可以包括各种存储装置,例如 RAM、闪速 RAM、MEMS(微电子机械系统)存储器、电池支持的 RAM、和 / 或非易失 RAM(NVRAM)等。备份存储装置 462 可用于存储数据(在装置 460 上存储)的备份副本。备份存储装置 462 可以包括以上列出的用于备份存储装置 460 的存储装置中的一个或多个。在一个实施方案中,当在备份存储装置 460 拷贝一

个镜像时,可以删除存储在文件服务器 490 上的相应数据。同样,当在备份存储装置 462 上拷贝一个镜像时,就可以删除存储在装置 460 上的相应数据。用于如前所述定义给定数据镜像的信息生命周期的基于时间的计划表可以与备份存储装置 460 和 462 相关联。

[0048] 在一个实施方案中,网络架构 400 中的备份服务器 450 可以包含备份应用程序 458。备份应用程序 458 或“备份管理器”可以包括在存储介质 470 上存储并可由处理器 452 执行(以进行一个或多个备份操作)的程序指令。备份操作可以包括例如在计算机系统上或从一个计算机系统到另一个系统创建部分的或完整的备份;在计算机系统上或从一个计算机系统到另一个系统执行还原功能(例如部分或粒度还原功能或完全还原功能);备份操作还可以包括其他操作。备份管理器 458 可以允许用户执行基于时间的时序安排以便如先前描述的那样定义给定数据镜像的信息生命周期。

[0049] 备份服务器 450 上的备份管理器 458 可以被配置用于创建数据 492 的各种类型的备份。例如,可以创建基于图像的、基于文件的、或其他类型的备份。在一些实施方案中,备份管理器 458 可以接收指导备份管理器 458 的用户输入以创建特定系统(例如文件服务器 490)的备份。在一些实施方案中,备份管理器 458 可以自动创建特定系统的备份,例如可以安排备份管理器执行特定系统的定期备份。备份管理器 458 可以采用众所周知的文件格式(例如虚拟硬盘(VHD)格式、V2I 格式以及虚拟机磁盘格式(VMDK))来创建并存储备份或文件服务器镜像 474。

[0050] 继续描述数据备份操作,在一个实施方案中,备份服务器 450 上的备份管理器 458 可以创建数据 492 的备份,例如文件服务器镜像 474,并将它存储在备份存储装置 460 上。在另一个实施方案中,备份管理器 458 可以在备份服务器 450 上存储备份,例如,如图所示在备份服务器 450 的存储介质 470 上。在又其他实施方案中,备份服务器 458 相反可以存储另一个装置上(例如在网络 480 上连接的另一个计算机系统上)的备份。

[0051] 在一个实施方案中,备份管理器 458 可以直接创建计算机系统 422 和 420 的备份,无需使用文件服务器 490。执行这种备份可以保护计算机系统 422 和 420 上的数据(在文件服务器 490 上这些数据可能仍未进行更新)。在一些实施方案中,如果计算机系统 422 和 420 连接至网络 480,备份管理器 458 可以自动地创建计算机系统 422 和 420 的逐文件备份。在这种情况下,可以实施有计划的定期备份,例如计算机系统 422 和 420 的每天或每周计划备份。

[0052] 作为创建备份的一部分,备份管理器 458 还可以创建与相应的文件或数据镜像相关联的目录元数据 456。在备份操作过程中备份管理器 458 创建的目录元数据 456(例如客户端 420 上的数据的备份)可以包括用于文件的元数据。元数据可以包括以下内容中的至少一项或多项:数据名称、数据大小、数据指纹、批号,以及数据类型、版本号、所有权、权限、修改时间、错误代码等。批号可用于标识特定的备份操作,在该备份操作中创建相关联的备份项。元数据和/或标识符的其他形式对本领域普通技术人员是明显的。例如目录元数据 456 可以随后用于还原浏览。

[0053] 需要指出的是,在替代实施方案中,客户端、服务器和存储装置的数量和类型不限于图 4 中所示的那些。服务器、桌上计算机、膝上计算机以及移动客户端的几乎任意数量和组合可以通过调制解调器集群、直接 LAN 连接、无线连接、WAN 链路等在网络架构 400 中进行互联。另外,在不同时刻,一个或多个客户端可以离线操作。另外,在操作过程中,当移动

用户从一地到另一地连接、断开并且重新连接至网络架构 400 时,个体客户端连接类型可以改变。

[0054] 在不同的实施方案中,本文所述方法和机制的一个或多个部分可以构成云计算环境的一部分。在这样的实施方案中,根据一个或多个不同模型,可以通过互联网将资源作为服务来提供。这类模型可以包括基础设施即服务 (IaaS)、平台即服务 (PaaS)、软件即服务 (SaaS)。在 IaaS 中,计算机基础设施作为服务进行传送。在这种情况下,计算设备通常由服务提供商拥有并操作。在 PaaS 模式中,用于开发软件解决方案的开发者所用的软件工具和底层设备可以由服务提供者作为服务来提供并托管。SaaS 典型地包括按需作为服务的服务提供商许可软件。服务提供商可以托管软件,或者可以将软件调配给客户一段给定的时间。以上模式的很多组合是可能的并且是可以设想的。

[0055] 尽管已经相当详细地描述了以上这些实施方案,一旦以上披露得到充分理解,对于本领域普通技术人员而言众多的变化和修改将变得很清楚。在此的意图是将以下权利要求解释为涵盖所有此类变化和修改。

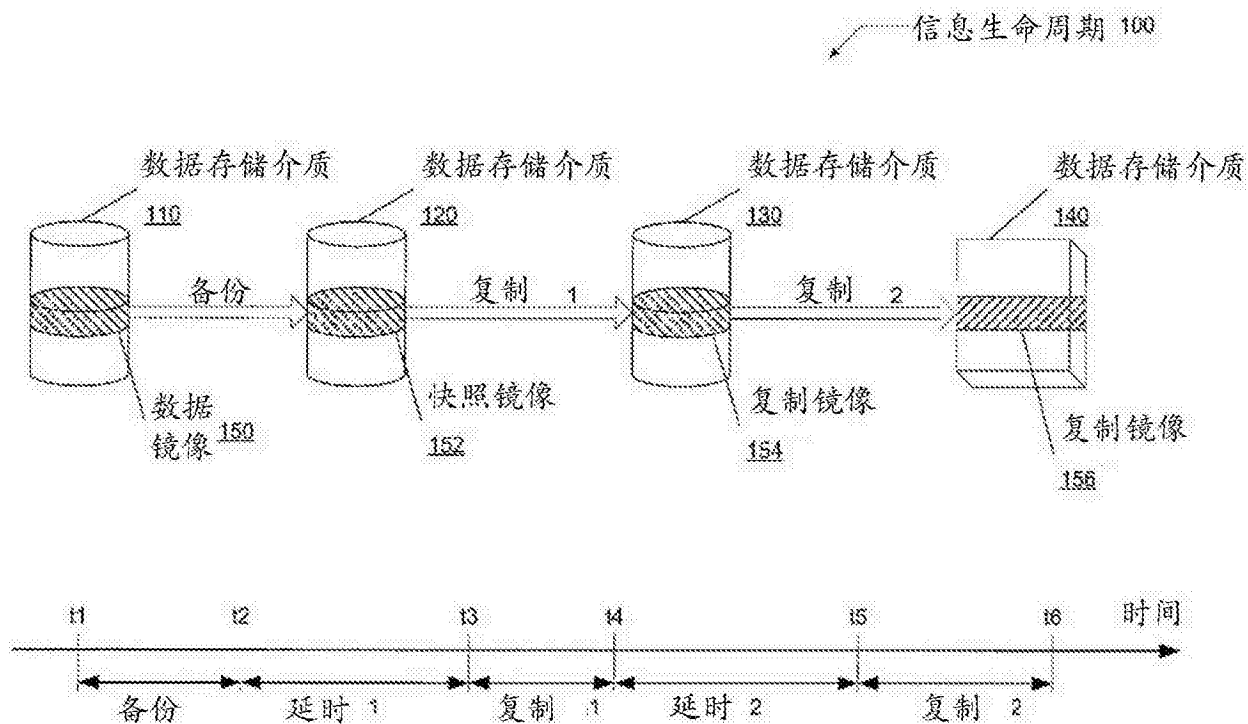


图 1

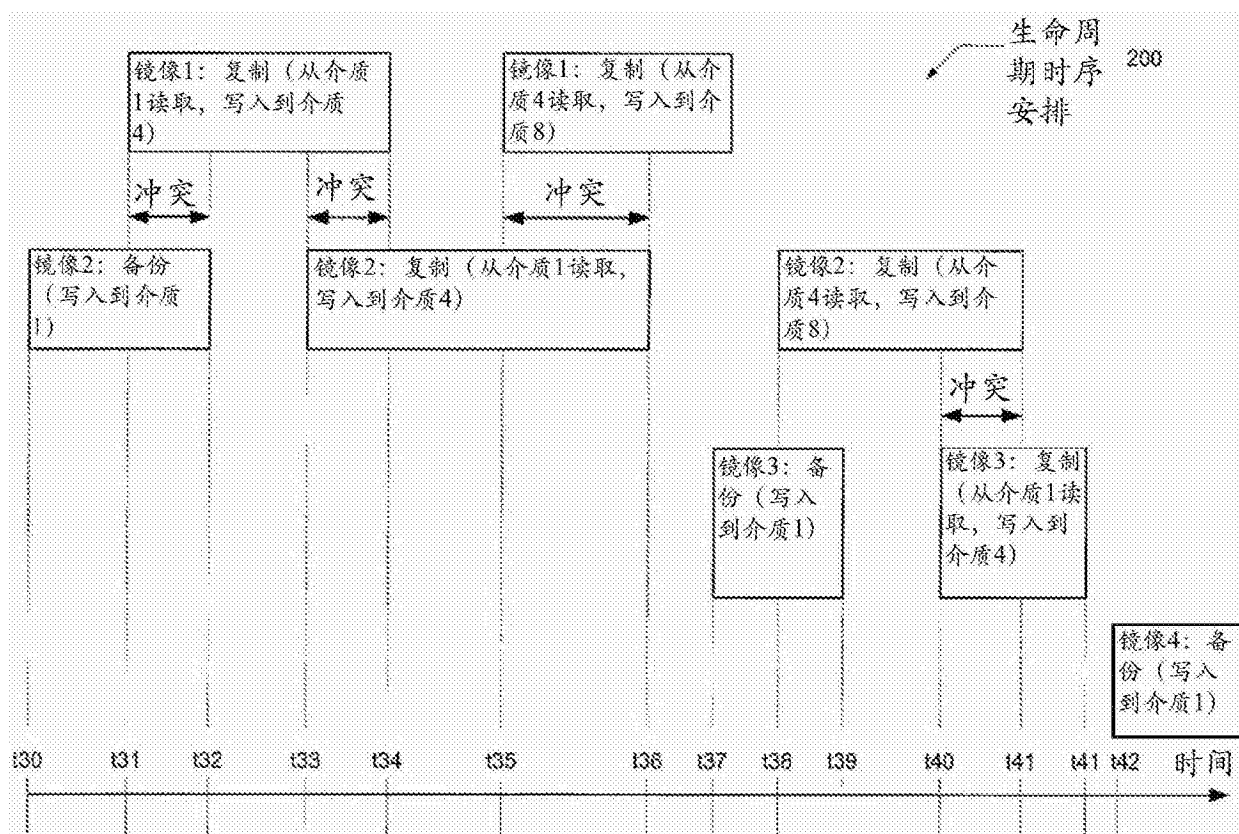


图 2

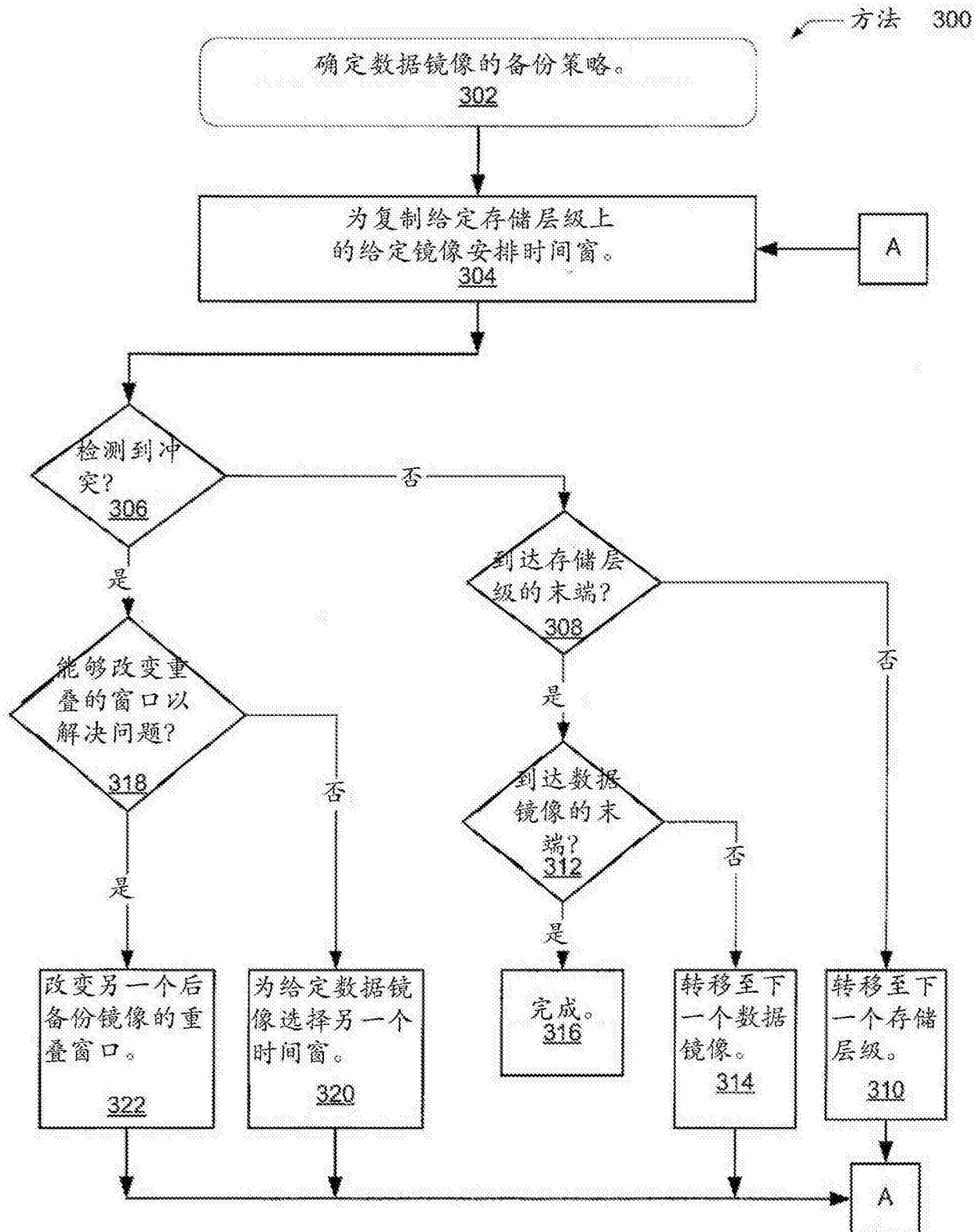


图 3

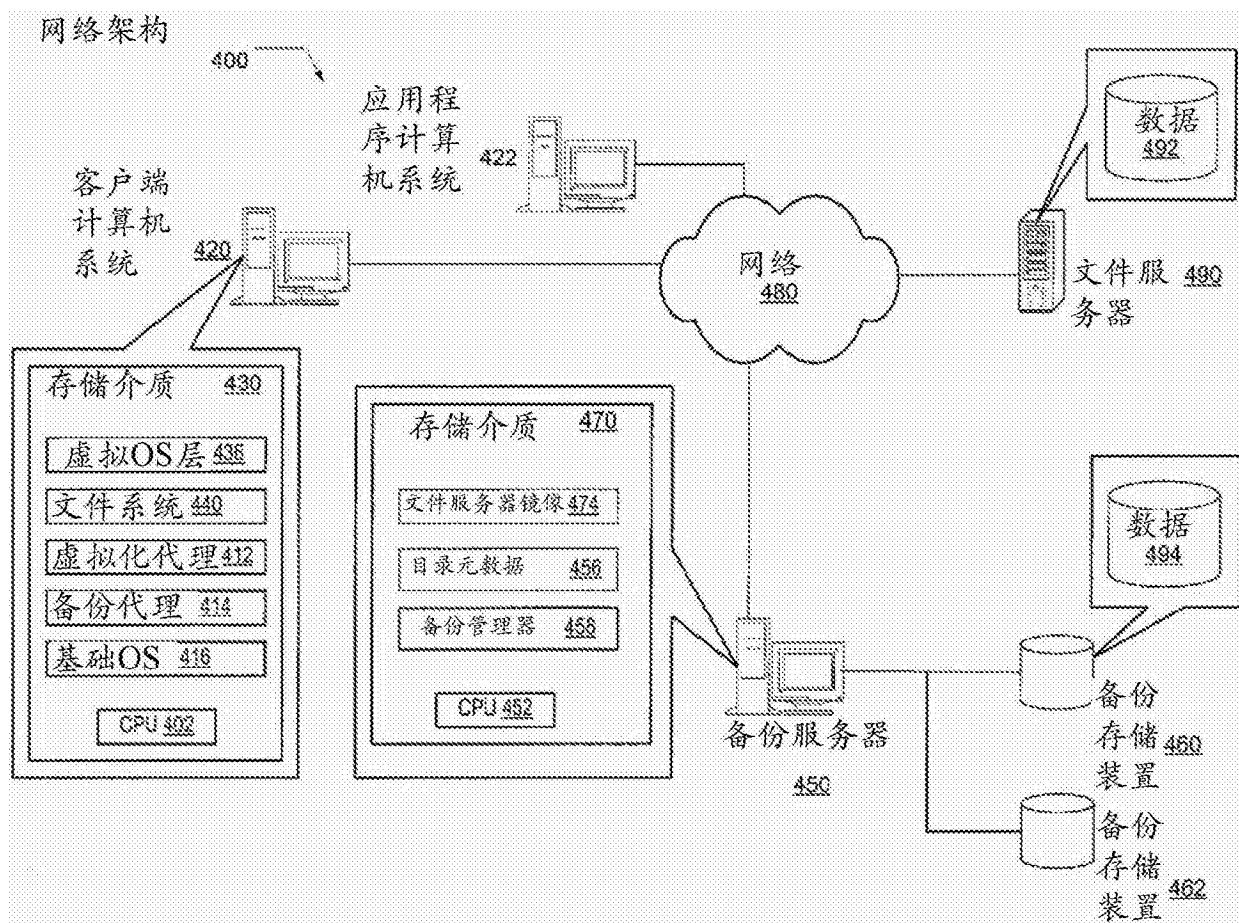


图 4