(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 116457760 A
(43)申请公布日 2023.07.18

(21)申请号 202180075194.X
(22)申请日 2021.09.29
(30)优先权数据
17/091,635 2020.11.06 US
(85)PCT国际申请进入国家阶段日
2023.05.06
(86)PCT国际申请的申请数据
PCT/US2021/052600 2021.09.29
(87)PCT国际申请的公布数据
WO2022/098450 EN 2022.05.12
(71)申请人 甲骨文国际公司
地址 美国加利福尼亚
(72)发明人 V·H·玛里佩迪 T·J·波茨 J·O·M·拉尔森 M·泰雅
(74)专利代理机构 中国贸促会专利商标事务所有限公司 11038
专利代理师 刘玉洁
(51)Int.Cl.
*G06F 11/08*(2006.01)

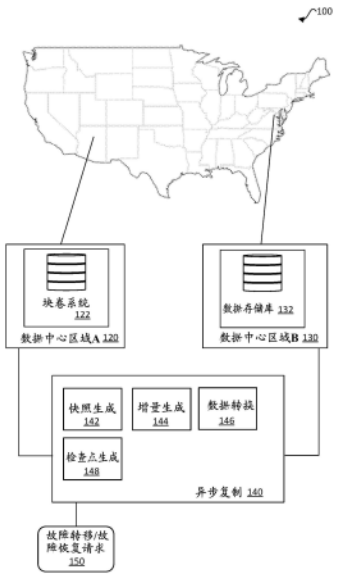
权利要求书3页 说明书31页 附图16页

(54)发明名称

异步跨区域块卷复制

(57)摘要

提供了用于块卷数据的跨区域复制的技术。这些技术包括一种方法,其中计算机系统实现操作,包括:在第一地理区域和第一逻辑时间创建块卷的第一快照,该块卷包括多个分区,将第一快照数据传输到在第二地理区域的对象存储系统,在第一地理区域和第二逻辑时间创建块卷的第二快照,生成多个增量,将与多个增量对应的多个增量数据集传输到在第二地理区域处的对象存储系统,至少部分地通过聚集与多个增量和第一快照相关联的对象元数据来生成检查点,接收生成还原卷的还原请求,以及从检查点生成还原卷。




CN 116457760 A

1.一种方法,包括:

由计算机系统在第一地理区域和第一逻辑时间创建块卷的第一快照,所述块卷包括多个分区;

由计算机系统将与第一快照对应的第一快照数据传输到第二地理区域处的对象存储系统;

由计算机系统在第一地理区域和第二逻辑时间创建块卷的第二快照;

由计算机系统生成多个增量,所述多个增量中的每个增量与所述多个分区中的分区对应;

由计算机系统将与所述多个增量对应的多个增量数据集传输到第二地理区域处的对象存储系统;

由计算机系统至少部分地通过聚集与所述多个增量和第一快照相关联的对象元数据来生成检查点;

由计算机系统接收生成还原卷的还原请求;以及

由计算机系统从检查点生成还原卷。

2.如权利要求1所述的方法,其中生成所述多个增量包括:

生成第二快照与第一快照之间的比较;

基于所述比较来确定与第一快照数据和第二快照所对应的第二快照数据之间的改变对应的经修改的数据;以及

生成所述多个增量,其中所述多个增量描述用于所述多个分区的经修改的数据。

3.如权利要求1所述的方法,其中创建第一快照包括:

与逻辑时间对应地暂停用于所述多个分区的输入/输出操作;

生成描述所述多个分区中的卷数据的多个块映像;以及

启用用于所述多个分区的输入/输出操作。

4.如权利要求1所述的方法,其中还原请求是故障转移请求,所述方法还包括:

使得能够在第二地理区域处生成还原卷;以及

使得能够在第二地理区域使用还原卷进行输入/输出操作。

5.如权利要求1所述的方法,其中还原请求是故障恢复请求,所述方法还包括:

在第二地理区域处生成还原卷;

使得能够在第一地理区域处生成故障恢复卷;以及

在第一地理区域处还原第一快照数据。

6.如权利要求1所述的方法,其中传输所述多个增量数据集包括:

从所述多个增量数据集生成多个分块对象;

传送所述多个增量;以及

将所述多个分块对象传送到对象存储系统。

7.如权利要求6所述的方法,其中:

检查点包括对象元数据的清单;以及

对象元数据包括与对象存储系统中的所述多个分块对象对应的分块指针。

8.如权利要求7所述的方法,其中聚集对象元数据包括更新清单以反映所述多个增量数据集与第一快照数据之间的多个差异。

9.一种计算机系统,包括

一个或多个处理器;

存储器,与所述一个或多个处理器通信,所述存储器被配置为存储计算机可执行指令,其中执行所述计算机可执行指令使得所述一个或多个处理器执行包括以下的步骤:

由计算机系统在第一地理区域和第一逻辑时间创建块卷的第一快照,所述块卷包括多个分区;

由计算机系统将与第一快照对应的第一快照数据传输到第二地理区域处的对象存储系统;

由计算机系统在第一地理区域和第二逻辑时间创建块卷的第二快照;

由计算机系统生成多个增量,所述多个增量中的每个增量与所述多个分区中的分区对应;

由计算机系统将与所述多个增量对应的多个增量数据集传输到第二地理区域处的对象存储系统;

由计算机系统至少部分地通过聚集与所述多个增量和第一快照相关联的对象元数据来生成检查点;

由计算机系统接收生成还原卷的还原请求;以及

由计算机系统从检查点生成还原卷。

10.如权利要求9所述的计算机系统,其中生成所述多个增量包括:

生成第二快照与第一快照之间的比较;

基于所述比较来确定与第一快照数据和第二快照所对应的第二快照数据之间的改变对应的经修改的数据;以及

生成所述多个增量,其中所述多个增量描述用于所述多个分区的经修改的数据。

11.如权利要求9所述的计算机系统,其中创建第一快照包括:

与逻辑时间对应地暂停用于所述多个分区的输入/输出操作;

生成描述所述多个分区中的卷数据的多个块映像;以及

启用用于所述多个分区的输入/输出操作。

12.如权利要求9所述的计算机系统,其中还原请求是故障转移请求,所述方法还包括:

使得能够在第二地理区域处生成还原卷;以及

使得能够在第二地理区域使用还原卷进行输入/输出操作。

13.如权利要求9所述的计算机系统,其中还原请求是故障恢复请求,该方法还包括:

在第二地理区域处生成还原卷;

使得能够在第一地理区域处生成故障恢复卷;以及

在第一地理区域处还原第一快照数据。

14.如权利要求9所述的计算机系统,其中传输所述多个增量数据集包括:

从所述多个增量数据集生成多个分块对象;

传送所述多个增量;以及

将所述多个分块对象传送到对象存储系统。

15.如权利要求14所述的计算机系统,其中检查点包括对象元数据的清单;以及

对象元数据包括与对象存储系统中的所述多个分块对象对应的分块指针。

16.如权利要求15所述的计算机系统,其中聚集对象元数据包括更新清单以反映所述多个增量数据集与第一快照数据之间的多个差异。

17.一种存储计算机可执行指令的计算机可读存储介质,所述计算机可执行指令在被执行时使得计算机系统的一个或多个处理器执行包括以下的步骤:

由计算机系统在第一地理区域和第一逻辑时间创建块卷的第一快照,所述块卷包括多个分区;

由计算机系统将与第一快照对应的第一快照数据传输到第二地理区域处的对象存储系统;

由计算机系统在第一地理区域和第二逻辑时间创建块卷的第二快照;

由计算机系统生成多个增量,所述多个增量中的每个增量与所述多个分区中的分区对应;

由计算机系统将与所述多个增量对应的多个增量数据集传输到第二地理区域处的对象存储系统;

由计算机系统至少部分地通过聚集与所述多个增量和第一快照相关联的对象元数据来生成检查点;

由计算机系统接收生成还原卷的还原请求;以及

由计算机系统从检查点生成还原卷。

18.如权利要求17所述的计算机可读存储介质,其中生成所述多个增量包括:

生成第二快照与第一快照之间的比较;

基于所述比较来确定与第一快照数据和第二快照所对应的第二快照数据之间的改变对应的经修改的数据;以及

生成所述多个增量,其中所述多个增量描述用于所述多个分区的经修改的数据。

19.如权利要求17所述的计算机可读存储介质,其中传输所述多个增量数据集包括:

从所述多个增量数据集生成多个分块对象;

传送所述多个增量;以及

将所述多个分块对象传送到对象存储系统。

20.如权利要求19所述的计算机可读存储介质,其中:

检查点包括对象元数据的清单;以及

对象元数据包括与对象存储系统中的所述多个分块对象对应的分块指针。

# 异步跨区域块卷复制

[0001] 相关申请的交叉引用

[0002] 本申请要求于2020年11月6日提交的标题为“ASYNCHRONOUS CROSS-REGION BLOCK VOLUME REPLICATION”的美国非临时申请No.17/091,635的权益和优先权。美国非临时申请No.17/091,635的内容出于所有目的通过引用整体并入本文。

## 背景技术

[0003] 基于云的平台为用户提供可扩展且灵活的计算资源。此类基于云的平台，也称为基础设施即服务(IaaS)，可以围绕客户的数据提供整套云解决方案，例如，用于创作变换、加载数据和呈现数据的解决方案。在一些情况下，客户数据可以存储在块卷存储装置和/或分布式存储系统(例如，云存储装置)中的对象存储装置中。

## 发明内容

[0004] 提供了用于异步跨区域块卷复制的技术(例如，方法、系统、存储可由一个或多个处理器执行的代码或指令的非暂态计算机可读介质)。

[0005] 在实施例中，一种方法包括由计算机系统在第一地理区域和第一逻辑时间创建块卷的第一快照，该块卷包括多个分区。该方法包括由计算机系统将与第一快照对应的第一快照数据传输到第二地理区域处的对象存储系统。该方法包括由计算机系统在第一地理区域和第二逻辑时间创建块卷的第二快照。该方法包括由计算机系统生成多个增量，多个增量中的每个增量与多个分区中的分区对应。该方法包括由计算机系统将与多个增量对应的多个增量数据集传输到第二地理区域处的对象存储系统。该方法包括由计算机系统至少部分地通过聚集与多个增量和第一快照相关联的对象元数据来生成检查点。该方法包括由计算机系统接收生成还原卷的还原请求。该方法还包括由计算机系统从检查点生成恢复卷。

[0006] 在变体中，生成多个增量包括生成第二快照与第一快照之间的比较。生成多个增量可以包括基于比较确定与第一快照数据和第二快照所对应的第二快照数据之间的改变对应的经修改的数据。多个增量可以描述用于多个分区的经修改的数据。创建第一快照可以包括与逻辑时间对应地暂停用于多个分区的输入/输出操作，生成描述多个分区中的卷数据的多个块映像，以及启用用于多个分区的输入/输出操作。还原请求可以是或包括故障转移(failover)请求。该方法还可以包括使得还原卷能够在第二地理区域处生成并且使得能够在第二地理区域使用还原卷进行输入/输出操作。还原请求可以是或包括故障恢复(failback)请求，并且该方法还可以包括在第二地理区域处生成还原卷，从而使得能够至少部分地通过克隆还原卷而在第一地理区域生成故障恢复卷，以及在第一地理区域处还原第一快照数据。传输多个增量数据集可以包括从多个增量数据集生成多个分块(chunk)对象，传送多个增量，以及将多个分块对象传送到对象存储系统。检查点可以包括对象元数据的清单(manifest)。对象元数据可以包括与对象存储系统中的多个分块对象对应的分块指针。聚集对象元数据可以包括更新清单以反映多个增量数据集与第一快照数据之间的多个差异。

[0007] 在某些实施例中，计算机系统包括一个或多个处理器和与该一个或多个处理器通信的存储器，该存储器被配置为存储计算机可执行指令，其中执行计算机可执行指令使得一个或多个处理器执行如上所述的方法或其变体的步骤中的一个或多个步骤。

[0008] 在某些实施例中，计算机可读存储介质存储计算机可执行指令，这些指令在被执行时使计算机系统的一个或多个处理器执行如上所述的方法或其变体的一个或多个步骤。

## 附图说明

[0009] 图1图示了根据一个或多个实施例的用于异步跨区域块卷复制的示例系统。

[0010] 图2图示了根据一个或多个实施例的用于异步块卷复制的示例技术。

[0011] 图3图示了根据一个或多个实施例的用于通过异步复制来还原块卷系统的示例技术。

[0012] 图4图示了根据一个或多个实施例的用于聚集块卷复制元数据的示例技术。

[0013] 图5图示了根据一个或多个实施例的用于从备用卷生成故障转移卷的示例技术。

[0014] 图6图示了根据一个或多个实施例的用于调整备用卷的大小的示例技术。

[0015] 图7图示了根据一个或多个实施例的用于生成还原卷的示例流程。

[0016] 图8图示了根据一个或多个实施例的用于生成故障转移卷的示例流程。

[0017] 图9图示了根据一个或多个实施例的用于生成故障恢复卷的示例流程。

[0018] 图10图示了根据一个或多个实施例的用于使用备用系统还原块卷系统的示例流程。

[0019] 图11图示了根据一个或多个实施例的用于调整块卷系统和备用系统的大小的示例流程。

[0020] 图12是图示根据至少一个实施例的用于将云基础设施实现为服务系统的一种模式的框图。

[0021] 图13是图示根据至少一个实施例的用于将云基础设施实现为服务系统的另一种模式的框图。

[0022] 图14是图示根据至少一个实施例的用于将云基础设施实现为服务系统的另一种模式的框图。

[0023] 图15是图示根据至少一个实施例的用于将云基础设施实现为服务系统的另一种模式的框图。

[0024] 图16是图示根据至少一个实施例的示例计算机系统的框图。

[0025] 在附图中，相似的组件和/或特征可以具有相同的附图标记。另外，相同类型的各种组件可以通过在附图标记后面加上破折号和区分相似组件的第二标记来区分。如果在说明书中仅使用第一附图标记，那么该描述适用于具有相同的第一附图标记的相似组件中的任何一个，而不管第二附图标记如何。

## 具体实施方式

[0026] 在下面的描述中，将描述各种实施例。出于解释的目的，阐述具体配置和细节以便提供对实施例的透彻理解。但是，对于本领域技术人员来说也将明显的是，可以在没有具体细节的情况下实践这些实施例。此外，可以省略或简化众所周知的特征以便不混淆所描述

的实施例。

[0027] 基于云的平台为用户提供可扩展且灵活的计算资源。此类基于云的平台，也称为基础设施即服务(IaaS)，可以围绕客户的数据提供整套云解决方案，诸如用于创作变换、加载数据和呈现数据的解决方案。在一些情况下，客户数据可以存储在块卷存储装置和/或分布式存储系统(例如，云存储装置)中的对象存储装置中。客户数据可以存储在位于一个地理区域的(例如，作为全球分布式存储系统的一部分的)数据中心中。可以至少部分地基于一个或多个性能度量来选择数据中心，性能度量包括时延、每秒输入-输出操作(IOPS)、吞吐量、成本、稳定性等。在一些情况下，数据中心的最优位置可以与客户的位置对应(例如，当客户生成大量内部数据时)。在一些情况下，最优位置可以与客户的客户端和/或用户的位置对应(例如，当客户运营内容交付网络时)。

[0028] 可以将客户数据复制到不同地理区域或数据中心(也称为可用性域，或“AD”)中的备份系统，以通过将备份系统指定为主系统在故障之后恢复(称为“故障转移”)。故障转移系统可以通过多个度量来表征，包括恢复点目标(RPO)、恢复时间目标(RTO)和灾难恢复(DR)。一般而言，RPO描述了在故障转移期间可能丢失多少数据，使得备份系统的目标可能是最小化RPO。通常，生成备份副本涉及在一个区域进行备份，等待备份上传到云系统，然后发起跨区域复制操作并等待目的地区域中的复制完成。大量的元数据开销可以与每个备份相关联，这可能以高重复率增加计算需求和资源开支。出于这个原因，RPO通常对于备份系统以小时的数量级可用(其中时间指示根据平均数据传送率考虑时的数据量)。

[0029] 类似地，RTO可以是故障转移系统的重要度量，因为它指示在故障转移之后还原数据所涉及的时间的长度。对于从不同地理区域还原的备份，输入-输出操作可能在典型条件下表现出更高的时延，直到卷完全还原。更重要的是，还原时间和按需数据操作的时延可能响应于系统负载中的尖峰而显著增加，例如，系统负载中的尖峰与响应于区域范围的故障而引起的数据流量转移对应。当分布式数据存储系统还实现跨区域备份副本的备份还原时，时延可能进一步提高。

[0030] 客户数据可以是有价值的，使得客户可以预期数据受到一种或多种方法的保护，该一种或多种方法包括例如数据冗余性和被实现以增加存储稳定性的基础设施(例如，诸如电源加固、浪涌保护等)。数据冗余性可以采取多种形式，包括但不限于在本地存储介质(例如，物理存储介质)中、在分布式存储系统中和/或在不同的分布式存储系统中复制客户数据。在一些情况下，可以一次复制所有客户数据，称为备份。在进行备份时，这种方法可以在系统输入-输出操作中包括显著的时延。例如，存储系统可以在进行备份时冻结所有的读和写操作，以确保保留客户数据的完整副本。在此类情况下，可以采用附加的缓冲容量来临时将新数据存储在分布式存储系统中和/或可以延迟客户数据请求。

[0031] 与整个卷的备份形成对比，客户数据可以被异步复制，如下面参考附图更详细描述的。例如，至少部分地通过生成描述数据存储系统的子单元(例如，块卷系统的分区)中的增量改变的对象，可以持续地(例如，以分钟为间隔，而不是对于备份系统可能典型的以小时为间隔)复制客户数据。在以下段落中，描述增量改变的对象也被称为“增量”。如参考图1更详细描述的，增量可以描述两个异步记录之间对客户数据的一个或多个改变。分布式存储系统中用户数据的状态的记录也被称为“快照”，并且可以表示单个逻辑时间的数据的状态，而不是单个时序时间(chronological time)的数据的状态，如下面参考附图更详细描

述的。在一些情况下，可以生成增量并且可以在生成增量时存储和/或传送/复制增量。因此，异步复制可以允许通过传送增量而不是备份来复制客户数据。

[0032] 至少出于这些原因，本文描述的技术相对于依赖于备份客户数据的数据冗余性方法具有一个或多个优点。例如，生成多个增量以复制快照后对数据的改变可以允许数据复制过程按时序时间细分，并且可以减少数据复制对时延、IOPS和/或其它性能度量的影响。

[0033] 如果系统故障发生在可用备份未反映对客户数据的有价值的改变时(例如，在进行现有备份很久之后发生故障，或自上次备份以来客户数据已发生重大改变)，那么依赖于备份整个卷的方法可能面临数据丢失的风险。实现异步复制还可以提高整体系统性能，以及可能改善客户数据的保存。因此，相对于备份系统，异步复制方法可以提供改进的RPO、RTO和DR。

[0034] 作为说明性示例，块卷系统可以在分布式存储系统中实现，该分布式存储系统在世界各地的多个地理区域中具有多个数据中心。存储在其中一个数据中心的块卷系统中的客户数据可以复制到第二数据中心，以防止在第一数据中心遇到灾难性故障时数据丢失。可以通过生成快照和增量来复制客户数据，其中快照可以提供客户数据跨多个分区的整体复制，并且增量可以允许对在前一个快照之后对单个分区的客户数据的改变的增量跟踪。一旦生成增量，就可以将它与它描述的客户数据一起传送到第二数据中心，在第一数据中心发生故障的情况下，第二数据中心的存储系统可以使用其来还原客户数据。在为一个分区生成增量时，其它分区可以保持可用于输入-输出操作并且可以允许第一数据中心处的存储系统维持期望的性能水平。在第一数据中心处出现故障的情况下，第二数据中心可以承担第一数据中心的角色(称为“故障转移”)和/或客户数据可以还原到第一数据中心，这可以恢复操作(称为“故障恢复”)。

[0035] 图1图示了根据一个或多个实施例的用于异步跨区域块卷复制的示例系统100。如上所述，系统100可以通过将客户数据从源系统复制到目的地系统来促进数据冗余性，以便在故障恢复期间提供客户数据。在一些实施例中，第一数据中心120(例如，源系统)可以将数据存储在块卷系统122中。在一些实施例中，第一数据中心120可以将数据存储在对象存储系统中，而不是块卷系统中。数据可以由第一数据中心120的用户生成和/或提供。第一数据中心120可以位于第一地理区域(例如，区域A)，其可以接近用户的位置和/或可以与一个或多个操作准则(诸如例如性能度量(例如，时延、IOPS、吞吐量、成本、稳定性等))对应。在一些情况下，第一数据中心120的常规操作的可能中断(例如，电力故障、数据损坏、分布式拒绝服务攻击(DDOS)、自然灾害等)可能会带来数据丢失的风险。为了潜在地降低数据丢失的风险，可以在第二数据中心130(例如，目的地系统)中复制数据。在一些实施例中，第二数据中心130可以位于与第一数据中心120的位置不同的地理区域，这可以降低由自然灾害或基础设施故障造成的数据丢失的风险。类似地，将数据存储在与第一数据中心120分开的第二数据中心130中可以减少针对可能通过以第一数据中心为目标而中断输入-输出操作和/或导致数据丢失的恶意动作(例如，DDOS、数据损坏等)的潜在漏洞。如下面更详细描述的，第二数据中心130可以将数据存储在数据存储库132中作为备用卷、对象存储装置和/或其它存储格式(例如，基于客户配置和/或偏好)。

[0036] 在一些实施例中，存储在第一数据中心120中的数据可以经由异步复制140系统在第二数据中心130中复制。如上所述，异步复制可以提供优于周期性备份复制的技术优势，

因为它可以表现出在恢复操作(例如,RTO)时减少正常输入-输出处理的中断,并且可以减少由第一数据中心120处的中断造成的数据丢失的程度(例如,RPO)。在一些实施例中,异步复制140可以包括生成复制数据并将复制数据从第一数据中心120递增地传送到第二数据中心130,而不是作为在特定时序时间生成的连贯备份映像,其可以被新的备份映像周期性地替换。代替地,如下面参考图2更详细描述的,异步复制140可以例如,通过生成和传送存储在第一数据中心120中的数据和/或数据的改变的记录,动态地更新存储在第二数据中心130中的数据。在一些实施例中,块卷系统122的客户和/或用户可以在创建组成卷时和/或作为在块卷系统122已经作为分布式数据存储系统操作之后可用的后续选项来配置异步复制140。在一些实施例中,配置异步复制140可以包括指定目的地区域(例如,第二地理区域,“区域B”)。在一些实施例中,例如,当复制的数据要存储在备用块卷系统中时,配置异步复制140可以包括指定目的地区域内的目的地AD。

[0037] 在一些实施例中,异步复制140可以包括快照生成142子系统,其可以生成存储在块卷系统122中的数据的快照。在一些情况下,快照可以描述在特定逻辑时间存储在块卷系统122中的用户数据的即时状态,如下面更详细描述的。与备份复制方法形成对比,异步复制140可以包括增量生成144子系统,通过该子系统可以确定在快照之后对块卷系统122中的数据所做的改变。这些改变进而可以被表示为增量,该增量可以被用于更新快照,而不是完全替换快照。快照生成142可以周期性地生成新的快照,例如以分钟的数量级(例如,在1-10分钟、10-20分钟、20-30分钟等的范围内)和/或以小时的数量级(例如,1-10小时)。

[0038] 在一些实施例中,快照生成142可以经由涉及块卷系统122的一个或多个分区的两阶段提交协议来进行。两阶段提交可以包括在快照生成142已经从块卷系统接收到分区上的输入-输出操作已经暂停的承诺(commitment)之后生成分区映像。在分区映像完成之后,可以释放分区以恢复输入-输出操作。在一些实施例中,两阶段提交可以作为整体应用于块卷系统122,使得在生成快照的时间期间可以阻塞整个块卷系统的所有读和写操作。在一些实施例中,可以在毫秒级(例如,1-15毫秒、5-10毫秒等)生成快照。

[0039] 在一些实施例中,快照生成142包括以分区为基础生成存储在块卷系统122中的数据的映像。在块卷系统122将数据存储在一个或多个分区中的情况下,快照生成142可以调度分区映像的生成,使得它不干扰分区的输入-输出操作。快照生成142可以生成和编译包括在块卷系统中的每个分区的映像,同时潜在地避免中断除了被映像的分区以外的分区的正常输入-输出操作。

[0040] 快照可以记录数据在特定逻辑时间的映像,其中逻辑时间是指快照生成142的实现。例如,由快照生成142生成的第一快照可以用第一逻辑时间来识别,并且可以在快照生成142处理块卷系统122的一个或多个分区的一段时序时间期间被生成。类似地,快照生成142可以在时序时间段(例如,1-10分钟等)过去之后生成可以用第二逻辑时间识别的块卷系统122的第二快照,其与快照生成142子系统的第二次迭代对应。

[0041] 在一些实施例中,增量生成144可以至少部分地通过允许异步复制140将新的或经修改的数据传送到第二数据中心130,而不发送未改变的数据,来促进异步复制140。在一些情况下,增量可以描述存储在块卷系统122的分区中的数据在第一快照与第二快照之间的改变。例如,增量生成144可以将第二快照与第一快照进行比较,并且可以确定哪些块已经被添加、移除、修改等。以这种方式,增量还可以描述在两个连续快照之间已经被移除的数

据。

[0042] 在一些情况下,增量可以是包括唯一标识符以及存储器指针的列表的逻辑结构。存储器指针可以描述在两个连续快照之间已经改变的数据在第一数据中心120和/或第二数据中心130中的存储器位置。如下面参考图2-图6更详细描述的,实现异步复制140可以包括将用于分区的增量传送到第二数据中心130,而不是传送整个快照。

[0043] 在一些实施例中,第二数据中心130处的数据存储库132可以将复制的数据存储为对象存储。因此,异步复制140可以包括数据转换146子系统以将来自第一数据中心120的数据转换成分块对象。在一些实施例中,来自块卷系统122的数据可以被结合到4MB分块对象中。在一些实施例中,分块对象可以实现类似于用于系统备份操作的格式,这可以允许异步复制140集成到采用备份和/或分层上传技术的现有分布式存储系统中。

[0044] 在一些实施例中,异步复制140可以在已经生成与块卷系统122的分区对应的增量之后生成检查点。在一些实施例中,检查点生成148子系统可以通过聚集增量并将聚集的改变应用到先前生成的检查点来生成检查点,如下面参考图4更详细描述的。

[0045] 当第一数据中心经历故障(例如,灾难性故障)并且例如块卷系统122的用户和/或客户期望系统的还原时,故障转移/故障恢复请求150可以提供给异步复制140。故障转移/故障恢复请求150可以是使异步复制140使用存储在第二地理区域中的复制的数据生成还原卷的还原请求的形式。在一些实施例中,异步复制140可以响应于接收到故障转移/故障恢复请求150而实现一个或多个还原操作。在一些实施例中,故障转移/故障恢复请求150可以包括分布式存储系统的外部服务的请求和/或分布式存储系统的用户的请求。在一些情况下,如当数据存储库132是块卷系统122的复制品(例如,备用卷)时,故障转移请求可以指示数据存储库132应当被配置为承担块卷系统122的角色。例如,数据存储库可以处置涉及用户数据的输入-输出操作。在一些实施例中,故障转移/故障恢复请求150可以指示第一数据中心120应当被配置为接收复制的数据,如上所述,并且恢复导致故障转移请求的故障之前的操作(例如,故障恢复还原)。

[0046] 图2图示了根据一个或多个实施例的用于异步块卷复制的示例技术200。如上所述,异步复制可以允许用户数据以增量单元而不是作为单个备份映像从源系统传送到目的地系统。在一些实施例中,块卷系统122可以存储可以从用户接收并经由输入-输出(I/O)操作210提供给用户的用户数据。这样接收到的数据可以存储在块卷系统122的一个或多个组成分区220中。异步复制系统(例如,图1的异步复制140)可以通过一个或多个处理将用户数据从块卷系统122复制和传送到数据存储库132,如下所述。

[0047] 在一些实施例中,可以通过创建第一快照230来复制用户数据(例如,操作250)。如参考图1更详细描述的,快照可以是存储在块卷系统122中的数据的记录。第一快照130的快照生成(例如,图1的快照生成142)可以发生在第一逻辑时间260。如参考图1更详细描述的,逻辑时间可以描述快照生成操作(例如,操作250)的迭代,例如,当快照生成以分区为基础进行时,从而实现与块卷系统122的每个组成分区220的两阶段提交协议。在此类情况下,第一逻辑时间260可以与创建快照的时间段对应。

[0048] 在一些情况下,第一快照230是存储在块卷系统122中的用户数据的异步复制的最先实现。在此类情况下,由第一快照230表示的整个数据集可以被传送到数据存储库132。以这种方式,第一快照130可以类似于备份,至少是因为用户数据和元数据的整个记录(例如,

如下所述的清单)可以从源系统传送到目的地系统(例如,操作252)。

[0049] 在一些实施例中,第一快照可以是块卷系统122和/或组成分区220的盘映像。在一些实施例中,第一快照230可以包括多个块映像232,使得由第一快照230描述的数据可以至少部分地通过单独地和/或成组地传送块映像232而被异步地传送到数据存储库132。如参考图3更详细描述的,例如当数据存储库132是对象存储系统时,块映像232可以被转换成要存储在数据存储库132中的分块对象。

[0050] 在一些实施例中,异步复制可以包括在第二逻辑时间262创建第二快照240(例如,操作254)。在一些实施例中,第二逻辑时间262与第一逻辑时间260之后的时间段对应,如上文参考图1更详细描述的。例如,第二快照240可以在创建第一快照230之后几分钟(例如,1-15分钟、5-10分钟等)通过上述两阶段提交协议创建。在一些实施例中,第二快照240可以包括相对于第一快照230对存储在块卷系统122中的数据的一个或多个改变。

[0051] 不是直接传送第二快照240,技术200而是可以包括生成与块卷系统122的组成分区对应的增量242(例如,操作256)。如上面参考图1更详细描述的,增量生成(例如,图1的增量生成144)可以包括将第一快照230中反映的数据的状态与第二快照240中反映的数据的状态进行比较以生成与目的地系统(例如,数据存储库132)的引用块标识符配对的(由指针识别的)经修改的块的列表,其可以构成用于每个分区220的增量242中包括的数据的至少一部分。在一些实施例中,增量还可以包括唯一增量标识符。在一些实施例中,唯一增量标识符可以与第二快照240的标识符对应。以这种方式,增量242可以反映可以追溯到第二逻辑时间262的对数据的改变,例如,作为提供I/O历史的方法。

[0052] 一旦生成,增量242就可以被传送到目的地系统,例如数据存储库132(例如,操作258)。在一些实施例中,增量242可以与添加到块卷系统122的任何新数据一起被传送。在第一逻辑时间260和第二逻辑时间262之间从块卷系统122移除的数据可以由包括在增量242中的元数据反映,而无需将任何数据(例如,分块对象)从源系统传送到目的地系统。

[0053] 图3图示了根据一个或多个实施例的用于通过异步复制来还原块卷系统的示例技术300。作为上面参考图1-图2描述的异步复制系统的一部分,目的地系统(例如,第二数据中心130)可以接收和维护存储在源系统(例如,图1的第一数据中心120)中的数据的复制品。在一些实施例中,第二数据中心130可以位于与源系统的地理区域不同的地理区域(例如,与区域A相对的区域B)中。第二数据中心130可以包括数据存储库132,其可以将复制的数据存储在对象存储装置中,作为一个或多个数据对象310。

[0054] 在一些实施例中,数据存储库132可以至少部分地基于异步复制系统的配置作为对象存储系统操作,使得异步复制可以被配置为将复制的数据存储为分块对象,而不是作为块。即,在一些实施例中,第二数据中心130可以包括备用卷,从而允许将增量直接应用到备用卷。在数据存储库132包括备用卷的实施例中,异步复制可以在不将块转换成对象(例如,数据转换146)的情况下完成。

[0055] 在一些实施例中,异步复制可以包括生成第一检查点320(例如,操作350)。检查点可以与快照不同,至少因为它可以包括描述存储在数据存储库132中的分块对象的位置和标识符的清单(例如,元数据),如参考图4更详细描述的。例如,在第一检查点320对应于作为异步复制的一部分生成了检查点的第一时间的情况下,第一检查点可以包括用于构成复制的数据的每个分块对象的分块指针的列表。

[0056] 作为异步复制的一部分，第二数据中心130将从源系统接收增量330和对应数据(例如，操作352)。如上所述，增量可以包括描述由第一快照与第二快照之间的输入-输出操作(图2的I/O操作210)引起的在源系统中存储的数据的改变的元数据。对应的数据可以作为分块对象(例如，4MB分块对象)被接收。

[0057] 可以作为生成第二检查点340的一部分将增量330应用于第一检查点320(例如，操作354)。生成第二检查点可以包括识别在增量330中识别出的一个或多个分块对象，以及应用由增量指示的所指示的修改。例如，到分块数据的一个或多个变换可以由增量330中的第一增量330-1指示，其中第一增量330-1(例如，经由分块指针)指示针对给定分区的所引用数据在数据存储库132的对象存储装置中的位置。在一些实施例中，作为生成第二检查点340的一部分，异步复制可以包括更新第一检查点320以反映由第一增量330-1指示的改变，而不是直接修改所引用的数据，如下面参考图4更详细描述的。

[0058] 在一些实施例中，第二数据中心130(例如，目的地系统)可以是故障转移/故障恢复请求(例如，图1的故障转移/故障恢复请求150)的主体，该请求可以从异步复制系统的用户(例如，图1的异步复制系统140，经由操作356)接收。故障转移/故障恢复请求可以响应于影响源系统(例如，图1的第一数据中心120)的故障(例如，自然灾害)。故障转移/故障恢复请求可以包括指导将如何从复制的数据和第二检查点还原块卷系统的参数，如下面参考图7-图10更详细描述的。

[0059] 实现故障转移/故障恢复请求可以包括生成故障转移/故障恢复卷，也称为还原卷(例如，操作358)。如上所述，故障转移系统可以托管在目的地系统处，而故障恢复系统可以托管在源系统处。在一些实施例中，如当接收到故障恢复请求时，第二检查点340可以被用于将存储在对象310中的数据映射到源系统(例如，块卷系统122)中的块。下面参考图5更详细地描述生成故障恢复卷的处理。类似地，故障转移可以包括在第二数据中心130处生成块卷系统以承担第一数据中心120关于输入-输出操作的角色，例如，直到第一数据中心120从故障中恢复。在第二数据中心130处生成故障转移卷可以包括，例如，通过使用第二检查点340，将存储在对象310中的数据映射到块，如下面参考图4更详细描述的。在一些实施例中，还原块卷数据可以包括一个或多个操作，包括但不限于在目的地区域中创建新的块卷、在目的地区域中实现故障转移、使得能够跨区域复制到源区域中的新卷上，以及在源区域执行故障转移。

[0060] 在一些实施例中，如当第二数据中心130维护备用卷以存储复制的数据时，备用卷可以已经包括来自源卷的所有改变，除了来自故障时尚未从源系统传送的任何增量的改变。因此，故障转移可以立即可用(例如，指示低RTO)，并且丢失的数据相对少(例如，快照生成时间为1-15分钟的数量级的RPO)，如下面参考图5更详细描述的。相比之下，实现整个备份映像的同步复制的备份系统可能导致几个小时的数量级的RPO。

[0061] 在一些实施例中，异步复制可以包括反向复制，而不是生成故障转移/故障恢复卷。在此类情况下，目的地系统可以被指定为源系统，而先前的源系统可以被重新指定为目的地系统。以这种方式，输入-输出操作可以发生在第二数据中心130处，其中数据在数据存储库132处被写入和读取，并且快照创建和增量生成也在那里发生。相应地，复制的数据可以被传送到第一数据中心120。

[0062] 图4图示了根据一个或多个实施例的用于聚集块卷复制元数据的示例技术400。异

步复制可以包括一个或多个组成操作(例如,快照创建、增量生成等)在可以复制数据的时间段内的多次迭代。在此类情况下,更新目的地系统(例如,图1-图3的第二数据中心130)可以包括维护与复制的数据对应的存储器位置的更新的记录,从而反映由每次后续迭代引入的改变。

[0063] 在一些实施例中,异步复制可以包括经由对增量(例如,图3的增量330)和第一检查点320的一个或多个操作来生成第二检查点340。例如,当在生成第一检查点320之后在目的地系统处从源系统接收到多个增量时,存储器位置的准确记录可以由如应用于第一检查点320的增量指示的改变更好地表示,而不是单独参考第一检查点。

[0064] 在一些实施例中,第一检查点320可以包括清单420。清单420可以与源系统的分区对应(例如,用于指导故障转移/故障恢复操作),因此,当源系统包括多个分区时,多个清单可以描述来自源系统的数据的异步复制。清单420可以包括分块指针422的列表(例如,分块指针1-N,其中"N"是引用包括在清单420中的分块指针422的数量的整数)。分块指针422可以描述与复制的数据相关联的存储器位置(例如,作为分块对象存储在数据存储库132中)。在生成第一检查点320之后,目的地系统可以接收多个增量,例如,如由快照创建和增量生成的迭代生成的增量。以这种方式,生成第二检查点340可以包括聚集增量(例如,操作450)。例如,聚集增量可以包括引用第一检查点320的标识符以确定从中生成增量的快照。以这种方式,检查点引用可以允许确定接收到的增量(还包括引用快照的标识符)是否已经反映在清单420中。

[0065] 在一些实施例中,增量可以与第一检查点组合以提供更新的清单440。组合增量可以包括修改如增量所指示的清单中的第一分块指针422-1(例如,指向数据的位置的指针可以响应于由增量描述的复制的数据的变换而改变)。以这种方式,通过将所有接收到的增量中指示的所有变换应用到清单,更新的清单440可以允许生成故障转移/故障恢复卷,其中与如果使用更早的列表(例如,清单420)相比可能减少RPO。类似地,对于异步复制的每次迭代,新生成的增量可以被生成并被传送到目的地系统,在那里它们可以与最新近的清单(例如,更新的清单)聚集。

[0066] 图5图示了根据一个或多个实施例的用于从备用卷生成故障恢复卷的示例技术500。如上所述,故障恢复请求(例如,图1的故障转移/故障恢复请求150)可以作为源系统(例如,第一数据中心120)处的故障恢复的一部分被接收。与托管在目的地系统或反向复制的故障转移卷相反,故障恢复卷可以是优选的,原因与选择第一数据中心120来托管用户数据的原因相同(例如,性能度量、稳定性等)。

[0067] 在一些实施例中,目的地系统(例如,第二数据中心130)可以实现备用卷560以复制源卷522处的数据。备用卷560可以包括从源卷接收的所有增量,使得备用卷反映源卷522的当前状态。例如,可以(例如,通过操作512)从源卷522生成第一增量540,并且第一增量540可以(例如,通过操作514)直接应用于备用卷560。在一些实施例中,将增量应用于备用卷560可以包括从源卷复制被引用的数据(例如,修改的块)并且以类似于输入-输出操作(例如,用户写入操作)的方式在备用卷中复制被引用的数据。此外,应用增量可以包括更新清单,如参考图4所描述的。虽然图4描述了关于分块指针的列表,但清单可以类似地描述备用卷中可以存储复制数据的存储器位置。

[0068] 在一些实施例中,更新备用卷560可以包括在离散操作中应用增量,使得增量可能

不会部分地应用于备用卷560。例如，可以生成备用卷560的第一快照530以提供回退位置，以防在应用增量的操作(例如，操作514)期间接收到故障转移/故障恢复请求。以这种方式，不是等待增量被应用，而是系统恢复可以实现第一快照530，这可以提供改进的RTO，代价是对RPO的递增增加。在一些实施例中，一旦第一增量540被应用于备用卷560，就可以删除第一快照530。进而，当生成第二增量550(或多个增量)时(例如，操作516)，可以在应用第二增量(例如，操作520)之前在备用卷560处生成第二快照532(例如，操作518)。在完全应用第二增量之后，也可以删除第二快照550。

[0069] 在一些实施例中，备用卷560可以被用于响应于接收到故障转移/故障恢复请求(例如，操作522)而生成故障转移卷或故障恢复卷。在一些实施例中，数据还原可以包括通过在故障转移卷上启用跨区域复制来将数据传送回源区域。在一些实施例中，当请求是故障恢复请求时，备用卷560可以在第一数据中心120处被克隆和还原(例如，操作524)。克隆备用卷560可以包括在第一数据中心120处创建故障恢复卷562并还原数据(例如，有时也称为“为克隆注水”)。在一些实施例中，如果在实现故障恢复请求时第二增量已经被完全应用，那么还原的数据可以包括由第二增量550指示的修改。使用克隆实现故障转移还原可以潜在地允许跨区域复制在现有备用卷上继续进行。有利地，克隆卷可以在几乎没有时延的情况下可用，从而允许对RTO的影响可忽略不计。

[0070] 图6图示了根据一个或多个实施例的用于调整备用卷的大小的示例技术600。源系统(诸如第一数据中心120处的源卷522)的正常操作可以包括调整源卷522的大小(例如，添加分区、移除分区、调整源卷522的一个或多个分区的大小等)。在一些实施例中，由于增量是为分区生成的，因此调整源卷522的大小可能影响源卷增量522到备用卷560的映射。因此，在实现一种或多种方法来限制由于调整源卷522的大小而引入的增量应用中的潜在错误之后，可以调整备用卷560的大小以将调整源卷522的大小考虑在内。

[0071] 在一些实施例中，可以从源卷522生成一个或多个增量620，如上面参考图1-图2更详细描述的(例如，操作610)。如上面参考图5所描述的，增量620可以应用于备用卷560(例如，操作612)。在一些情况下，调整源卷522的大小可以包括从用户或另一个系统接收调整大小请求(例如，操作614)。为了潜在地限制调整大小对RPO的影响，可以在调整源卷522的大小之前在接收到调整大小请求之后生成最后的增量630(例如，操作616)。在一些实施例中，生成最后的增量630可以包括创建源卷522的新快照，至少部分地基于新快照生成增量，以及应用由此生成的增量。以这种方式，备用卷560可以描述紧接在通过调整源卷522的大小(例如，操作618)来实现调整大小请求之前的源卷的状态。应用最后的增量(例如，操作620)可以与调整源卷522的大小解耦，使得一旦生成最后的增量630，就可以在最后的增量可以应用于备用卷560之前调整源卷的大小。在一些实施例中，可以在应用最后的增量之后以对应的方式调整备用卷560的大小(例如，操作622)。

[0072] 图7图示了根据一个或多个实施例的用于生成还原卷的示例流程700。流程700的操作可以被实现为硬件电路系统和/或存储为计算机系统(诸如图1的异步复制系统140)的非暂态计算机可读介质上的计算机可读指令。如实现的那样，指令表示包括电路系统或可由计算机系统的(一个或多个)处理器执行的代码的模块。此类指令的执行将计算机系统配置为执行本文描述的特定操作。每个电路系统或代码与处理器结合执行(一个或多个)相应的操作。虽然操作是以特定次序说明的，但应当理解的是，特定次序不是必需的并且可以省

略、跳过和/或重新排序一个或多个操作。

[0073] 在示例中，流程700包括操作702，其中计算机系统在第一地理区域创建块卷的快照。如参考图1-图2更详细描述的，块卷(例如，图1的块卷系统122)可以托管在第一地理区域中的第一数据中心(例如，图1的第一数据中心120)处，其可以至少部分地基于与输入/输出操作、稳定性、成本等相关联的一个或多个性能度量来确定。创建第一快照(例如，图2的操作250)可以包括生成构成块卷的一个或多个块的一个或多个块映像(例如，图2的块映像232)。创建快照可以包括实现两阶段提交协议，由此可以针对块卷的分区暂停输入-输出操作(例如，图2的I/O操作210)，在此期间创建用于那个分区的块映像。在为分区或块卷系统创建快照之后，计算机系统可以恢复块卷上的输入-输出操作。

[0074] 在示例中，流程700包括操作704，其中计算机系统将与第一快照对应的快照数据传输到第二地理区域处的第二存储系统。在一些实施例中，如当第一快照是块卷的第一个创建的快照时或者如果先前创建的快照不可用，那么快照数据可以包括块映像。因此，块映像可以被传输到系统(例如，图1的数据存储库132)。在一些实施例中，块映像可以被转换成分块对象并且作为快照数据被传输到第二存储系统(例如，图2的操作252)。在一些情况下，分块对象连同描述快照数据的元数据一起被传输，元数据包括例如列出识别快照数据在存储器中的位置的分块指针的清单。

[0075] 在示例中，流程700包括操作706，其中计算机系统在第一地理区域创建块卷的第二快照。如参考图2更详细描述的，第二快照(例如，图2的第二快照240)可以描述在创建了第一快照的第一逻辑时间(例如，图2的第一逻辑时间260)之后块卷在第二逻辑时间(例如，图2的第二逻辑时间262)的状态。第二快照可以包括关于在创建第一快照之后发生的对存储在块卷系统中的数据的修改(例如，对数据的读-写操作)的信息。

[0076] 在示例中，流程700包括操作708，其中计算机系统生成多个增量。不是将第二快照直接传输到第二存储系统，而是多个增量可以描述在创建第二快照(例如，第二逻辑时间)与第一快照(例如，第一逻辑时间)之间对存储在块卷中的数据的一个或多个改变。生成增量可以包括将第二快照中的分区的块映像与第一快照中的分区的对应映像进行比较，并且识别对分区中的数据的一个或多个修改。如参考图1-图2更详细描述的，增量可以包括元数据，元数据包括但不限于与快照对应的增量标识符、描述块卷中被修改的数据的一个或多个块标识符，和/或描述由块标识符识别的数据的位置的对应数量的对象标识符。

[0077] 在示例中，流程700包括操作710，其中计算机系统传输用于多个增量的数据。如参考图2更详细描述的，传输数据可以包括创建从块卷系统到第二存储系统的数据的复制品。这可以包括生成分块对象并将分块对象复制到第二存储系统(例如，目的地系统)，如当复制的数据存储在对象存储装置中时。数据可以伴随有多个增量，这些增量可以提供描述以分区为基础的分块对象与块卷系统的块之间的对应关系的元数据。

[0078] 在示例中，流程700包括操作712，其中计算机系统在第二地理区域生成检查点。如参考图3-图4更详细描述的，检查点(例如，图3的第一检查点320)可以包括描述多个分块指针的清单(例如，图4的清单420)。在一些实施例中，生成检查点可以包括将增量与先前生成的检查点聚集(例如，图4的操作450)以生成更新的检查点(例如，图3的第二检查点340)，更新的检查点可以包括描述反映应用于检查点的多个增量中包括的修改的第二多个分块指针的更新的清单(例如，图4的更新的清单420)。

[0079] 在示例中，流程700包括操作714，其中计算机系统接收还原请求。还原请求(例如，图1的故障转移/故障恢复请求150)可以包括在第二地理位置生成还原卷(也称为故障转移卷)的用户请求。在一些实施例中，还原请求可以包括在第一地理位置生成还原卷(也称为故障恢复卷)的用户请求。在一些实施例中，还原请求可以包括反转异步复制的请求，从而将第二存储系统(例如，图1的数据存储库132)指定为源系统并将块卷系统(例如，图1的块卷系统122)指定为异步复制的目的地系统。在一些实施例中，还原请求可以作为异步复制的配置的一部分自动生成(例如，无需用户交互)。例如，异步复制系统可以被配置为响应于在源系统处发生的故障而自动生成还原请求。

[0080] 在示例中，流程700包括操作716，其中计算机系统生成还原卷。如参考图3和图5所描述的，生成还原卷可以包括生成块卷，例如，这是通过参考更新的清单(例如，图4的更新的清单440)将复制品数据从存储在对象存储装置中的分块对象映射到块卷中的块。以这种方式，可以生成还原卷，其RPO可以低于备份系统所能达到的RPO。类似地，可以反映接收还原请求与恢复正常输入/输出操作之间的时间的系统还原操作的RTO可以取决于还原请求是故障转移还是故障恢复请求，这在下面参考图8和图9更详细地描述。

[0081] 图8图示了根据一个或多个实施例的用于生成故障转移卷的示例流程800。流程800的操作可以被实现为硬件电路系统和/或存储为计算机系统(诸如图1的异步复制系统140)的非暂态计算机可读介质上的计算机可读指令。如实现的那样，指令表示包括电路系统或可由计算机系统的(一个或多个)处理器执行的代码的模块。此类指令的执行将计算机系统配置为执行本文描述的特定操作。每个电路系统或代码与处理器结合执行(一个或多个)相应的操作。虽然操作是以特定次序说明的，但应当理解的是，特定次序不是必需的并且可以省略、跳过和/或重新排序一个或多个操作。

[0082] 在示例中，流程800包括操作802，其中计算机系统接收作为故障转移请求的还原请求。如上面参考图1-图6更详细描述的，在影响可以形成分布式存储系统(例如，云存储)的一部分的块卷系统(例如，图1的块卷系统122)的灾难或其它故障之后，可以由异步复制(例如，图1的异步复制140)接收故障转移请求。在一些实施例中，故障转移请求可以包括生成在可能不受影响块卷系统的故障影响的第二地理区域再现块卷系统的故障转移卷的请求。

[0083] 在示例中，流程800包括操作804，其中计算机系统在第二地理区域生成故障转移卷。在一些情况下，故障转移卷可以允许在影响块卷系统的故障被解决之前恢复输入-输出操作(例如，图2的I/O操作210)。在此类情况下，与故障恢复卷相比，可以通过生成故障转移卷来改进RTO。如上面参考图7所述，使用存储在对象存储装置中的复制品数据生成故障转移卷可以包括参考描述复制品数据在存储器中的位置的更新的清单(例如，图4的更新的清单440)来生成故障转移卷，该更新的清单是使用异步复制期间生成的最近聚集的增量来更新的。

[0084] 在示例中，流程800包括操作806，其中计算机系统为故障转移卷注水。如上面更详细描述的，为故障转移卷注水可以包括将更新的清单中描述的数据从对象存储装置(例如，如上文参考图1描述的分块对象中的数据)恢复到可通过块卷系统的读/写操作(例如，图2的I/O操作210)寻址的块。在一些实施例中，块卷系统的分区可以保存在故障转移卷中。

[0085] 在示例中，流程800包括操作808，其中计算机系统开始输入-输出操作。在其中来

自块卷系统的复制的数据已被还原并映射到故障转移卷的“注水”之后，故障转移卷可以开始输入-输出操作并且可以承担块卷系统的角色，同时故障可以得到解决。

[0086] 图9图示了根据一个或多个实施例的用于生成故障恢复卷的示例流程900。流程900的操作可以被实现为硬件电路系统和/或存储为计算机系统(诸如图1的异步复制系统140)的非暂态计算机可读介质上的计算机可读指令。如实现的那样，指令表示包括电路系统或可由计算机系统的(一个或多个)处理器执行的代码的模块。此类指令的执行将计算机系统配置为执行本文描述的特定操作。每个电路系统或代码与处理器结合执行(一个或多个)相应的操作。虽然操作是以特定次序说明的，但应当理解的是，特定次序不是必需的并且可以省略、跳过和/或重新排序一个或多个操作。

[0087] 在示例中，流程900包括操作902，其中计算机系统接收作为故障恢复请求的还原请求。在一些实施例中，造成还原请求的故障可能不是无限期的。例如，故障可能可以在可预测的时间长度内得到解决，如当故障是由可以可靠地克服的临时电力或网络中断引起的时。因此，故障恢复请求可以表示相对于生成故障转移卷的优选替代方案(例如，从分布式存储性能的角度)，以及相对于故障转移卷在RTO方面的潜在有限差异。

[0088] 在示例中，流程900包括操作904，其中计算机系统在第二地理区域生成备用卷。在一些实施例中，备用卷可以包括类似于故障转移卷的块卷系统，除了它不是被设计为在第二地理区域(例如，图1的第二数据中心130)执行输入-输出操作之外。例如，备用卷对于分布式存储系统的用户可以是不可见的和/或可以被配置为不够资格进行输入-输出操作。在一些实施例中，备用卷可以(例如，在分区、大小等方面)再现块卷系统的结构以及块到存储为分块对象的复制的数据的映射。

[0089] 在示例中，流程900包括操作906，其中计算机系统在第一地理区域克隆备用卷。如上面参考图5更详细描述的，克隆备用卷(例如，图5的操作524)可以包括通过至少部分地通过使用备用卷的结构生成故障恢复卷来将块卷系统还原到第一地理区域。例如，备用卷可以(例如，参考更新的清单)描述复制的数据的结构和映射，使得故障恢复卷可以被配置为一旦复制的数据从第二地理区域复制回第一地理区域就用新的映射被还原。

[0090] 在示例中，流程900包括操作908，其中计算机系统在第一地理区域为故障恢复卷注水。在一些实施例中，为故障恢复卷注水可以包括将复制的数据还原到第一地理区域(例如，图1的第一数据中心120)，使得故障恢复卷可以恢复输入-输出操作。还原复制的数据可以包括将数据从第二地理区域复制到第一地理区域。还原复制的数据还可以包括将来自异步复制迭代的映射(例如，由增量应用生成的那些)的数据重新映射到与故障恢复块卷系统中的新数据位置对应的新映射。

[0091] 在示例中，流程900包括操作910，其中计算机系统在第一地理区域开始输入-输出操作。输入-输出操作描述使用故障恢复卷访问、存储和修改数据。作为系统还原的一部分，低RTO和低RPO可以提高系统性能。至少出于这个原因，在一些实施例中，输入-输出操作可以在操作908的数据还原正在进行时开始。例如，读-写操作可以在复制的数据被完全从第二地理区域复制回到第一地理区域之前开始。在一些实施例中，复制和重新映射可以在故障恢复卷上开始输入-输出操作之前全部完成。

[0092] 图10图示了根据一个或多个实施例的用于生成故障恢复卷的示例流程1000。流程1000的操作可以被实现为硬件电路系统和/或存储为计算机系统(诸如图1的异步复制系统

140)的非暂态计算机可读介质上的计算机可读指令。如实现的那样,指令表示包括电路系统或可由计算机系统的(一个或多个)处理器执行的代码的模块。此类指令的执行将计算机系统配置为执行本文描述的特定操作。每个电路系统或代码与处理器结合执行(一个或多个)相应的操作。虽然操作是以特定次序说明的,但应当理解的是,特定次序不是必需的并且可以省略、跳过和/或重新排序一个或多个操作。

[0093] 如参考图6更详细描述的,在一些实施例中,异步复制可以包括第二地理位置处的备用卷,作为复制的数据的对象存储装置的附加或替代。不是将由增量描述的数据转换成(例如,如参考图1的数据转换146所描述的)分块对象,而是备用卷可以包括块数据从块卷系统到备用卷的直接传送。在一些实施例中,备用卷对于块卷系统的用户可以是不可见的(例如,不可用于输入-输出操作),但可以维护更新的块数据以在中断影响块卷系统的情况下潜在地最小化RTO和RPO。在一些实施例中,备用卷可以与目的地区域中的对象存储装置交互,例如,通过高速缓存来自增量的数据,同时还对于对象存储数据实现检查点操作。

[0094] 在示例中,流程1000包括操作1002,其中计算机系统从第一地理区域处的块卷系统生成增量。如参考图2更详细描述的,增量可以表示在第一逻辑时间的第一快照与第二逻辑时间的第二快照之间对存储在块卷系统中的数据的修改。

[0095] 在示例中,流程1000包括操作1004,其中计算机系统在第二地理区域创建备用卷系统的快照。在一些实施例中,更新备用卷之前可以创建备用卷的快照。如参考图1更详细描述的,创建快照可以包括以分区为基础创建构成备用卷的块的多个块映像。在组装后,这些映像可以描述在与快照的创建对应的逻辑时间备用卷中的数据的状态(例如,如上文参考两阶段提交协议所描述的)。快照可以为异步复制系统提供回退位置以在还原块卷系统时使用。例如,如果在用新增量更新备用卷时接收到还原请求,那么异步复制可以使用快照生成还原卷。

[0096] 在示例中,流程1000包括操作1006,其中计算机系统将增量应用于备用卷系统。与参考使用对象存储装置的数据复制所描述的方法形成对比,可以通过直接应用增量来更新备用卷。例如,如参考图2更详细描述的,增量可以包括描述块卷系统的存储器中用于作为修改的结果被复制的数据的位置的元数据(例如,块标识符)。以这种方式,可以通过应用来自增量的所指示的修改来对备用卷进行修改。

[0097] 在示例中,流程1000包括操作1008,其中计算机系统接收故障恢复请求。如上所述,可以在块卷系统处的故障或其它操作中断之后接收故障恢复请求,使得备用系统可以被用于在已经解决中断之后生成故障恢复系统。

[0098] 在示例中,流程1000包括操作1010,其中计算机系统克隆备用卷。如参考图5更详细描述的,克隆备用卷(例如,图5的操作524)可以包括在第一地理区域(例如,图1的第一数据中心120)创建可以(例如,在大小、分区等方面)再现备用卷的结构的故障恢复卷(例如,图5的故障恢复卷562)。

[0099] 在示例中,流程1000包括操作1012,其中计算机系统将复制的数据还原到第一地理区域处的故障恢复卷。如上所述,还原复制的数据可以包括将块数据从第二地理区域复制到第一地理区域(也称为“注水”)并将块数据映射到故障恢复卷。在一些实施例中,一旦复制的数据被还原和/或在还原复制的数据时,输入-输出操作可以在故障恢复卷上开始。

[0100] 图11图示了根据一个或多个实施例的用于生成故障恢复卷的示例流程1100。流程

1100的操作可以被实现为硬件电路系统和/或存储为计算机系统(诸如图1的异步复制系统140)的非暂态计算机可读介质上的计算机可读指令。如实现的那样,指令表示包括电路系统或可由计算机系统的(一个或多个)处理器执行的代码的模块。此类指令的执行将计算机系统配置为执行本文描述的特定操作。每个电路系统或代码与处理器结合执行(一个或多个)相应的操作。虽然操作是以特定次序说明的,但应当理解的是,特定次序不是必需的并且可以省略、跳过和/或重新排序一个或多个操作。

[0101] 在示例中,流程1100包括操作1102,其中计算机系统在第一地理区域接收调整大小请求。在异步复制系统的操作期间,例如,在快照创建的迭代之间,或在增量生成期间,第一地理区域处的块卷系统可能经历调整大小请求。由于调整大小请求可以包括添加和/或移除一个或多个分区,或者改变块卷系统的一个或多个分区的大小,因此调整大小之前生成的增量可能在备用卷调整大小之后与备用卷不兼容。

[0102] 在示例中,流程1100包括操作1104,其中计算机系统在第一地理区域生成最后的增量。在一些实施例中,异步复制系统可以响应于块卷系统接收到调整大小请求而实现经修改的复制协议。例如,不是在接收到调整大小请求后立即调整块卷系统的大小,而是可以创建另一个快照并生成一个或多个最后的增量。这种方法可以允许复制的数据反映在调整大小之前块卷系统的更新的状态,这进而可以减少从复制的数据生成的还原卷的RPO。

[0103] 在示例中,流程1100包括操作1106,其中计算机系统将最后的增量应用于第二地理区域处的备用卷。如上所述,应用最后的增量可以包括修改备用卷以反映由在操作1104中生成的一个或多个最后的增量所描述的对块卷系统的修改。

[0104] 在示例中,流程1100包括操作1108,其中计算机系统调整第一地理区域处的块卷的大小。在操作1104之后,可以将调整大小请求应用于块卷系统。

[0105] 在示例中,流程1100包括操作1110,其中计算机系统调整与调整大小请求对应的备用卷的大小。类似地,在操作1106之后,也可以以与调整大小请求对应的方式调整备用卷的大小。由于异步复制可以生成特定于分区的增量,因此可以对备用卷进行结构化以再现块卷系统的结构。例如,备用卷可以包括与块卷系统相同数量的相同大小的分区。

[0106] 如上所述,基础设施即服务(IaaS)是一种特定类型的云计算。IaaS可以被配置为通过公共网络(例如,互联网)提供虚拟化计算资源。在IaaS模型中,云计算提供商可以托管基础设施组件(例如,服务器、存储设备、网络节点(例如,硬件)、部署软件、平台虚拟化(例如,管理程序层)等)。在一些情况下,IaaS提供商还可以提供各种服务来伴随这些基础设施组件(例如,计费、监视、日志记录、安全性、负载平衡和聚类等)。因此,由于这些服务可能是策略驱动的,因此IaaS用户可能能够实现策略来驱动负载平衡,以维持应用的可用性和性能。

[0107] 在一些情况下,IaaS客户可以通过诸如互联网之类的广域网(WAN)访问资源和服务,并且可以使用云提供商的服务来安装应用栈的剩余元件。例如,用户可以登录到IaaS平台以创建虚拟机(VM)、在每个VM上安装操作系统(OS)、部署诸如数据库之类的中间件、为工作负载和备份创建存储桶,甚至将企业软件安装到该VM中。然后,客户可以使用提供商的服务来执行各种功能,包括平衡网络流量、解决应用问题、监视性能、管理灾难恢复等。

[0108] 在大多数情况下,云计算模型将需要云提供商的参与。云提供商可以但不一定是专门提供(例如,供应、出租、销售)IaaS的第三方服务。实体也可能选择部署私有云,从而成

为其自己的基础设施服务提供商。

[0109] 在一些示例中，IaaS部署是将新应用或应用的新版本放置到准备好的应用服务器等上的处理。它还可以包括准备服务器(例如，安装库、守护进程等)的处理。这通常由云提供商在管理程序层之下管理(例如，服务器、存储装置、网络硬件和虚拟化)。因此，客户可以负责处理(OS)、中间件和/或(例如，在(例如，可以按需启动的)自助服务虚拟机等上的)应用部署。

[0110] 在一些示例中，IaaS供应可以指获取计算机或虚拟主机以供使用，以及甚至在它们上安装所需的库或服务。大多数情况下，部署不包括供应，并且供应可能需要被首先执行。

[0111] 在一些情况下，IaaS供应存在两个不同的问题。首先，在任何东西运行之前供应初始基础设施集存在最初的挑战。其次，一旦所有东西已被供应，就存在演进现有基础设施(例如，添加新服务、更改服务、移除服务等)的挑战。在一些情况下，可以通过启用以声明方式定义基础设施的配置来解决这两个挑战。换句话说，基础设施(例如，需要哪些组件以及它们如何交互)可以由一个或多个配置文件来定义。因此，基础设施的总体拓扑(例如，哪些资源依赖于哪些资源，以及它们各自如何一起工作)可以以声明的方式描述。在一些情况下，一旦定义了拓扑，就可以生成创建和/或管理配置文件中描述的不同组件的工作流。

[0112] 在一些示例中，基础设施可以具有许多互连的元件。例如，可能存在一个或多个虚拟私有云(VPC)(例如，可配置和/或共享计算资源的潜在按需池)，也称为核心网络。在一些示例中，还可以供应一个或多个安全组规则以定义将如何设置网络的安全性以及一个或多个虚拟机(VM)。也可以供应其它基础设施元件，诸如负载平衡器、数据库等。随着期望和/或添加越来越多的基础设施元件，基础设施可以逐步演进。

[0113] 在一些情况下，可以采用连续部署技术来使得能够跨各种虚拟计算环境部署基础设施代码。此外，所描述的技术可以使得能够在这些环境内进行基础设施管理。在一些示例中，服务团队可以编写期望部署到一个或多个但通常是许多不同的(例如，跨各种不同的地理位置，有时跨越整个世界的)生产环境的代码。但是，在一些示例中，必须首先设置将在其上部署代码的基础设施。在一些情况下，供应可以手动完成，可以利用供应工具供应资源，和/或一旦供应基础设施就可以利用部署工具部署代码。

[0114] 图12是图示根据至少一个实施例的IaaS体系架构的示例模式的框图1200。服务运营商1202可以通信地耦合到可以包括虚拟云网络(VCN)1206和安全主机子网1208的安全主机租赁1204。在一些示例中，服务运营商1202可以使用一个或多个客户端计算设备，其可以是便携式手持设备(例如，**iPhone®**、蜂窝电话、**iPad®**、计算平板、个人数字助理(PDA))或可穿戴设备(例如，Google**Glass®**头戴式显示器)、运行软件(诸如Microsoft Windows**Mobile®**)和/或各种移动操作系统(诸如iOS、Windows Phone、Android、BlackBerry 8、Palm OS等)，并且支持互联网、电子邮件、短消息服务(SMS)、**Blackberry®**或其它通信协议。替代地，客户端计算设备可以是通用个人计算机，包括例如运行各种版本的Microsoft**Windows®** 、Apple**Macintosh®**和/或Linux操作系统的个人计算机和/或膝上型计算机。客户端计算设备可以是运行包括但不限于各种GNU/Linux操作系统(诸如例如Google Chrome OS)的各种商业上可获得的**UNIX®**或类UNIX

操作系统中的任何一种的工作站计算机。替代地或附加地,客户端计算设备可以是任何其它电子设备,诸如瘦客户端计算机、支持互联网的游戏系统(例如,具有或不具有Kinect®姿势输入设备的Microsoft Xbox游戏控制台),和/或能够通过可以访问VCN 1206和/或互联网的网络进行通信的个人消息传递设备。

[0115] VCN 1206可以包括本地对等网关(LPG)1210,其可以经由包含在SSH VCN 1212中的LPG 1210通信地耦合到安全壳(SSH)VCN 1212。SSH VCN 1212可以包括SSH子网1214,并且SSH VCN 1212可以经由包含在控制平面VCN 1216中的LPG 1210通信地耦合到控制平面VCN 1216。此外,SSH VCN 1212可以经由LPG 1210通信地耦合到数据平面VCN 1218。控制平面VCN 1216和数据平面VCN 1218可以包含在可以由IaaS提供商拥有和/或操作的服务租赁1219中。

[0116] 控制平面VCN 1216可以包括充当外围网络(例如,公司内部网和外部网络之间的公司网络的部分)的控制平面非军事区(DMZ)层1220。基于DMZ的服务器可以承担有限责任并有助于使安全漏洞受到控制。此外,DMZ层1220可以包括一个或多个负载平衡器(LB)子网1222、可以包括(一个或多个)应用子网1226的控制平面应用层1224、可以包括(一个或多个)数据库(DB)子网1230(例如,(一个或多个)前端DB子网和/或(一个或多个)后端DB子网)的控制平面数据层1228。包含在控制平面DMZ层1220中的(一个或多个)LB子网1222可以通信地耦合到包含在控制平面应用层1224中的(一个或多个)应用子网1226和可以包含在控制平面VCN 1216中的互联网网关1234,并且(一个或多个)应用子网1226可以通信地耦合到包含在控制平面数据层1228中的(一个或多个)DB子网1230以及服务网关1236和网络地址转换(NAT)网关1238。控制平面VCN 1216可以包括服务网关1236和NAT网关1238。

[0117] 控制平面VCN 1216可以包括数据平面镜像应用层1240,其可以包括(一个或多个)应用子网1226。包含在数据平面镜像应用层1240中的(一个或多个)应用子网1226可以包括可以执行计算实例1244的虚拟网络接口控制器(VNIC)1242。计算实例1244可以将数据平面镜像应用层1240的(一个或多个)应用子网1226通信地耦合到可以包含在数据平面应用层1246中的(一个或多个)应用子网1226。

[0118] 数据平面VCN 1218可以包括数据平面应用层1246、数据平面DMZ层1248和数据平面数据层1250。数据平面DMZ层1248可以包括(一个或多个)LB子网1222,其可以通信地耦合到数据平面应用层1246的(一个或多个)应用子网1226和数据平面VCN 1218的互联网网关1234。(一个或多个)应用子网1226可以通信地耦合到数据平面VCN 1218的服务网关1236和数据平面VCN 1218的NAT网关1238。数据平面数据层1250还可以包括可以通信地耦合到数据平面应用层1246的(一个或多个)应用子网1226的(一个或多个)DB子网1230。

[0119] 控制平面VCN 1216和数据平面VCN 1218的互联网网关1234可以通信地耦合到元数据管理服务1252,元数据管理服务1252可以通信地耦合到公共互联网1254。公共互联网1254可以通信地耦合到控制平面VCN 1216和数据平面VCN 1218的NAT网关1238。控制平面VCN 1216和数据平面VCN 1218的服务网关1236可以通信地耦合到云服务1256。

[0120] 在一些示例中,控制平面VCN 1216或数据平面VCN 1218的服务网关1236可以对云服务1256进行应用编程接口(API)调用,而无需通过公共互联网1254。从服务网关1236对云服务1256的API调用可以是单向的:服务网关1236可以对云服务1256进行API调用,并且云服务1256可以将请求的数据发送到服务网关1236。但是,云服务1256不可以发起对服务网

关1236的API调用。

[0121] 在一些示例中，安全主机租赁1204可以直接连接到服务租赁1219，服务租赁1219否则可以被隔离。安全主机子网1208可以通过LPG 1210与SSH子网1214通信，LPG 1210可以使得能够在否则被隔离的系统上进行双向通信。将安全主机子网1208连接到SSH子网1214可以使安全主机子网1208能够访问服务租赁1219内的其它实体。

[0122] 控制平面VCN 1216可以允许服务租赁1219的用户设置或以其它方式供应期望的资源。在控制平面VCN 1216中供应的期望资源可以在数据平面VCN 1218中部署或以其它方式使用。在一些示例中，控制平面VCN 1216可以与数据平面VCN 1218隔离，并且控制平面VCN 1216的数据平面镜像应用层1240可以经由VNIC 1242与数据平面VCN 1218的数据平面应用层1246通信，VNIC 1242可以包含在数据平面镜像应用层1240和数据平面应用层1246中。

[0123] 在一些示例中，系统的用户或客户可以通过可以将请求传送到元数据管理服务1252的公共互联网1254来做出请求，例如创建、读取、更新或删除(CRUD)操作。元数据管理服务1252可以通过互联网网关1234将请求传送到控制平面VCN 1216。请求可以被包含在控制平面DMZ层1220中的(一个或多个)LB子网1222接收。(一个或多个)LB子网1222可以确定请求是有效的，并且响应于该确定，(一个或多个)LB子网1222可以将请求传输到包含在控制平面应用层1224中的(一个或多个)应用子网1226。如果请求被验证并且需要对公共互联网1254的调用，那么对公共互联网1254的调用可以被传输到可以对公共互联网1254进行调用的NAT网关1238。请求可能期望存储的存储器可以存储在(一个或多个)DB子网1230中。

[0124] 在一些示例中，数据平面镜像应用层1240可以促进控制平面VCN 1216和数据平面VCN 1218之间的直接通信。例如，可能期望对包含在数据平面VCN 1218中的资源应用对配置的更改、更新或其它适当的修改。经由VNIC 1242，控制平面VCN 1216可以与包含在数据平面VCN 1218中的资源直接通信，并且从而可以执行对配置的更改、更新或其它适当的修改。

[0125] 在一些实施例中，控制平面VCN 1216和数据平面VCN 1218可以包含在服务租赁1219中。在这种情况下，系统的用户或客户可能不拥有或操作控制平面VCN 1216或数据平面VCN 1218。替代地，IaaS提供商可以拥有或操作控制平面VCN 1216和数据平面VCN 1218，这两者都可以包含在服务租赁1219中。该实施例可以使得能够隔离可能阻止用户或客户与其它用户或其它客户的资源交互的网络。此外，该实施例可以允许系统的用户或客户私有地存储数据库，而无需依赖可能不具有期望的安全级别的公共互联网1254进行存储。

[0126] 在其它实施例中，包含在控制平面VCN 1216中的(一个或多个)LB子网1222可以被配置为从服务网关1236接收信号。在这个实施例中，控制平面VCN 1216和数据平面VCN 1218可以被配置为由IaaS提供商的客户调用而无需调用公共互联网1254。IaaS提供商的客户可能期望这个实施例，因为客户使用的(一个或多个)数据库可以由IaaS提供商控制并且可以存储在服务租赁1219上，服务租赁1219可以与公共互联网1254隔离。

[0127] 图13是图示根据至少一个实施例的IaaS体系架构的另一个示例模式的框图1300。服务运营商1302(例如，图12的服务运营商1202)可以通信地耦合到安全主机租赁1304(例如，图12的安全主机租赁1204)，该安全主机租赁1304可以包括虚拟云网络(VCN)1306(例如，图12的VCN 1206)和安全主机子网1308(例如，图12的安全主机子网1208)。VCN 1306可

以包括本地对等网关(LPG)1310(例如,图12的LPG 1210),其可以经由包含在SSH VCN 1312中的LPG 1210通信地耦合到安全壳(SSH)VCN 1312(例如,图12的SSH VCN 1212)。SSH VCN 1312可以包括SSH子网1314(例如,图12的SSH子网1214),并且SSH VCN 1312可以经由包含在控制平面VCN 1316中的LPG 1310通信地耦合到控制平面VCN 1316(例如,图12的控制平面VCN 1216)。控制平面VCN 1316可以包含在服务租赁1319(例如,图12的服务租赁1219)中,并且数据平面VCN 1318(例如,图12的数据平面VCN 1218)可以包含在可能由系统的用户或客户拥有或操作的客户租赁1321中。

[0128] 控制平面VCN 1316可以包括可以包括(一个或多个)LB子网1322(例如,图12的(一个或多个)LB子网1222)的控制平面DMZ层1320(例如,图12的控制平面DMZ层1220)、可以包括(一个或多个)应用子网1326(例如,图12的(一个或多个)应用子网1226)的控制平面应用层1324(例如,图12的控制平面应用层1224)、可以包括(一个或多个)数据库(DB)子网1330(例如,类似于图12的(一个或多个)DB子网1230)的控制平面数据层1328(例如,图12的控制平面数据层1228)。包含在控制平面DMZ层1320中的(一个或多个)LB子网1322可以通信地耦合到包含在控制平面应用层1324中的(一个或多个)应用子网1326和可以包含在控制平面VCN 1316中的互联网网关1334(例如,图12的互联网网关1234),并且(一个或多个)应用子网1326可以通信地耦合到包含在控制平面数据层1328中的(一个或多个)DB子网1330以及服务网关1336(例如,图12的服务网关)和网络地址转换(NAT)网关1338(例如,图12的NAT网关1238)。控制平面VCN 1316可以包括服务网关1336和NAT网关1338。

[0129] 控制平面VCN 1316可以包括可以包括(一个或多个)应用子网1326的数据平面镜像应用层1340(例如,图12的数据平面镜像应用层1240)。包含在数据平面镜像应用层1340中的(一个或多个)应用子网1326可以包括可以执行计算实例1344(例如,类似于图12的计算实例1244)的虚拟网络接口控制器(VNIC)1342(例如,1242的VNIC)。计算实例1344可以促进数据平面镜像应用层1340的(一个或多个)应用子网1326和可以包含在数据平面应用层1346(例如,图12的数据平面应用层1246)中的(一个或多个)应用子网1326之间经由包含在数据平面镜像应用层1340中的VNIC 1342和包含在数据平面应用层1346中的VNIC 1342的通信。

[0130] 包含在控制平面VCN 1316中的互联网网关1334可以通信地耦合到元数据管理服务1352(例如,图12的元数据管理服务1252),该元数据管理服务1352可以通信地耦合到公共互联网1354(例如,图12的公共互联网1254)。公共互联网1354可以通信地耦合到包含在控制平面VCN 1316中的NAT网关1338。包含在控制平面VCN 1316中的服务网关1336可以通信地耦合到云服务1356(例如,图12的云服务1256)。

[0131] 在一些示例中,数据平面VCN 1318可以包含在客户租赁1321中。在这种情况下,IaaS提供商可以为每个客户提供控制平面VCN 1316,并且IaaS提供商可以为每个客户设置包含在服务租赁1319中的唯一计算实例1344。每个计算实例1344可以允许包含在服务租赁1319中的控制平面VCN 1316和包含在客户租赁1321中的数据平面VCN 1318之间的通信。计算实例1344可以允许在包含在服务租赁1319中的控制平面VCN 1316中供应的资源被部署或以其它方式在包含在客户租赁1321中的数据平面VCN 1318中使用。

[0132] 在其它示例中,IaaS提供商的客户可以具有居住在客户租赁1321中的数据库。在这个示例中,控制平面VCN 1316可以包括数据平面镜像应用层1340,其可以包括(一个或多

个)应用子网1326。数据平面镜像应用层1340可以驻留在数据平面VCN 1318中,但数据平面镜像应用层1340可能不居住在数据平面VCN 1318中。即,数据平面镜像应用层1340可以访问客户租赁1321,但是数据平面镜像应用层1340可能不存在于数据平面VCN 1318中或者可能不由IaaS提供商的客户拥有或操作。数据平面镜像应用层1340可以被配置为对数据平面VCN 1318进行调用,但可以不被配置为对包含在控制平面VCN 1316中的任何实体进行调用。客户可能期望在数据平面VCN 1318中部署或以其它方式使用在控制平面VCN 1316中供应的资源,并且数据平面镜像应用层1340可以促进客户的期望部署或资源的其它使用。

[0133] 在一些实施例中,IaaS提供商的客户可以将过滤器应用到数据平面VCN 1318。在这个实施例中,客户可以确定数据平面VCN 1318可以访问什么,并且客户可以限制从数据平面VCN 1318对公共互联网1354的访问。IaaS提供商可能无法应用过滤器或以其它方式控制数据平面VCN 1318对任何外部网络或数据库的访问。客户将过滤器和控制应用到包含在客户租赁1321中的数据平面VCN 1318上可以帮助将数据平面VCN 1318与其它客户和公共互联网1354隔离开。

[0134] 在一些实施例中,云服务1356可以由服务网关1336调用以访问公共互联网1354、控制平面VCN 1316或数据平面VCN 1318上可能不存在的服务。云服务1356与控制平面VCN 1316或数据平面VCN 1318之间的连接可以不是实时的或连续的。云服务1356可以存在于由IaaS提供商拥有或操作的不同网络上。云服务1356可以被配置为接收来自服务网关1336的调用并且可以被配置为不接收来自公共互联网1354的调用。一些云服务1356可以与其它云服务1356隔离,并且控制平面VCN 1316可以与可能与控制平面VCN 1316不在同一区域的云服务1356隔离。例如,控制平面VCN 1316可能位于“区域1”,并且云服务“部署12”可能位于区域1和“区域2”。如果包含在位于区域1中的控制平面VCN 1316中的服务网关1336对部署12进行调用,那么该调用可以被传输到区域1中的部署12。在这个示例中,控制平面VCN 1316或区域1中的部署12可能不与区域2中的部署12通信地耦合或以其它方式通信。

[0135] 图14是图示根据至少一个实施例的IaaS体系架构的另一个示例模式的框图1400。服务运营商1402(例如,图12的服务运营商1202)可以通信地耦合到安全主机租赁1404(例如,图12的安全主机租赁1204),该安全主机租赁1404可以包括虚拟云网络(VCN)1406(例如,图12的VCN 1206)和安全主机子网1408(例如,图12的安全主机子网1208)。VCN 1406可以包括LPG 1410(例如,图12的LPG 1210),其可以经由包含在SSH VCN 1412中的LPG 1410通信地耦合到SSH VCN 1412(例如,图12的SSH VCN 1212)。SSH VCN 1412可以包括SSH子网1414(例如,图12的SSH子网1214),并且SSH VCN 1412可以经由包含在控制平面VCN 1416中的LPG 1410通信地耦合到控制平面VCN 1416(例如,图12的控制平面VCN 1216)并且经由包含在数据平面VCN 1418中的LPG 1410通信地耦合到数据平面VCN 1418(例如,图12的数据平面1218)。控制平面VCN 1416和数据平面VCN 1418可以包含在服务租赁1419(例如,图12的服务租赁1219)中。

[0136] 控制平面VCN 1416可以包括可以包括(一个或多个)负载平衡器(LB)子网1422(例如,图12的(一个或多个)LB子网1222)的控制平面DMZ层1420(例如,图12的控制平面DMZ层1220)、可以包括(一个或多个)应用子网1426(例如,类似于图12的(一个或多个)应用子网1226)的控制平面应用层1424(例如,图12的控制平面应用层1224)、可以包括(一个或多个)DB子网1430的控制平面数据层1428(例如,图12的控制平面数据层1228)。包含在控制平面

DMZ层1420中的(一个或多个)LB子网1422可以通信地耦合到包含在控制平面应用层1424中的(一个或多个)应用子网1426和可以包含在控制平面VCN 1416中的互联网网关1434(例如,图12的互联网网关1234),并且(一个或多个)应用子网1426可以通信地耦合到包含在控制平面数据层1428中的(一个或多个)DB子网1430以及服务网关1436(例如,图12的服务网关)和网络地址转换(NAT)网关1438(例如,图12的NAT网关1238)。控制平面VCN 1416可以包括服务网关1436和NAT网关1438。

[0137] 数据平面VCN 1418可以包括数据平面应用层1446(例如,图12的数据平面应用层1246)、数据平面DMZ层1448(例如,图12的数据平面DMZ层1248),以及数据平面数据层1450(例如,图12的数据平面数据层1250)。数据平面DMZ层1448可以包括可以通信地耦合到数据平面应用层1446的(一个或多个)可信应用子网1460和(一个或多个)不可信应用子网1462以及包含在数据平面VCN 1418中的互联网网关1434的(一个或多个)LB子网1422。(一个或多个)可信应用子网1460可以通信地耦合到包含在数据平面VCN 1418中的服务网关1436、包含在数据平面VCN 1418中的NAT网关1438以及包含在数据平面数据层1450中的(一个或多个)DB子网1430。(一个或多个)不可信应用子网1462可以通信地耦合到包含在数据平面VCN 1418中的服务网关1436和包含在数据平面数据层1450中的(一个或多个)DB子网1430。数据平面数据层1450可以包括可以通信地耦合到包含在数据平面VCN 1418中的服务网关1436的(一个或多个)DB子网1430。

[0138] (一个或多个)不可信应用子网1462可以包括可以通信地耦合到租户虚拟机(VM)1466(1)-(N)的一个或多个主VNIC 1464(1)-(N)。每个租户VM 1466(1)-(N)可以通信地耦合到可以包含在相应容器出口VCN 1468(1)-(N)中的相应应用子网1467(1)-(N),相应容器出口VCN 1468(1)-(N)可以包含在相应客户租赁1470(1)-(N)中。相应的辅助VNIC 1472(1)-(N)可以促进数据平面VCN 1418中包含的(一个或多个)不可信应用子网1462与容器出口VCN 1468(1)-(N)中包含的应用子网之间的通信。每个容器出口VCN 1468(1)-(N)可以包括NAT网关1438,该NAT网关1438可以通信地耦合到公共互联网1454(例如,图12的公共互联网1254)。

[0139] 包含在控制平面VCN 1416中并且包含在数据平面VCN 1418中的互联网网关1434可以通信地耦合到元数据管理服务1452(例如,图12的元数据管理系统1252),该元数据管理服务1452可以通信地耦合到公共互联网1454。公共互联网1454可以通信地耦合到包含在控制平面VCN 1416中并且包含在数据平面VCN 1418中的NAT网关1438。包含在控制平面VCN 1416中和包含在数据平面VCN 1418中的服务网关1436可以通信地耦合到云服务1456。

[0140] 在一些实施例中,数据平面VCN 1418可以与客户租赁1470集成。在一些情况下,诸如在执行代码时可能期望支持的情况下,这种集成对于IaaS提供商的客户可能是有用的或期望的。客户可能提供可能具有破坏性、可能与其它客户资源通信或可能以其它方式导致非期望效果的代码来运行。作为对此的响应,IaaS提供商可以确定是否运行由客户给予IaaS提供商的代码。

[0141] 在一些示例中,IaaS提供商的客户可以向IaaS提供商授予临时网络访问,并请求附加到数据平面层应用1446的功能。运行该功能的代码可以在VM 1466(1)-(N)中执行,并且该代码可以不被配置为在数据平面VCN 1418上的其它任何地方运行。每个VM 1466(1)-(N)可以连接到一个客户租赁1470。包含在VM 1466(1)-(N)中的相应容器1471(1)-(N)可以

被配置为运行代码。在这种情况下,可以存在双重隔离(例如,容器1471(1)-(N)运行代码,其中容器1471(1)-(N)可能至少包含在(一个或多个)不可信应用子网1462中包含的VM 1466(1)-(N)中),这可以帮助防止不正确的或以其它方式非期望的代码损坏IaaS提供商的网络或损坏不同客户的网络。容器1471(1)-(N)可以通信地耦合到客户租赁1470并且可以被配置为传输或接收来自客户租赁1470的数据。容器1471(1)-(N)可以不被配置为从数据平面VCN 1418中的任何其它实体传输或接收数据。在运行代码完成后,IaaS提供商可以终止或以其它方式处理掉容器1471(1)-(N)。

[0142] 在一些实施例中,(一个或多个)可信应用子网1460可以运行可以由IaaS提供商拥有或操作的代码。在这个实施例中,(一个或多个)可信应用子网1460可以通信地耦合到(一个或多个)DB子网1430并且被配置为在(一个或多个)DB子网1430中执行CRUD操作。(一个或多个)不可信应用子网1462可以通信地耦合到(一个或多个)DB子网1430,但是在这个实施例中,(一个或多个)不可信应用子网可以被配置为在(一个或多个)DB子网1430中执行读取操作。可以包含在每个客户的VM 1466(1)-(N)中并且可以运行来自客户的代码的容器1471(1)-(N)可以不与(一个或多个)DB子网1430通信地耦合。

[0143] 在其它实施例中,控制平面VCN 1416和数据平面VCN 1418可以不直接通信地耦合。在这个实施例中,控制平面VCN 1416和数据平面VCN 1418之间可能没有直接通信。但是,通信可以通过至少一种方法间接发生。LPG 1410可以由IaaS提供商建立,其可以促进控制平面VCN 1416和数据平面VCN 1418之间的通信。在另一个示例中,控制平面VCN 1416或数据平面VCN 1418可以经由服务网关1436调用云服务1456。例如,从控制平面VCN 1416对云服务1456的调用可以包括对可以与数据平面VCN 1418通信的服务的请求。

[0144] 图15是图示根据至少一个实施例的IaaS体系架构的另一个示例模式的框图1500。服务运营商1502(例如,图12的服务运营商1202)可以通信地耦合到安全主机租赁1504(例如,图12的安全主机租赁1204),该安全主机租赁1504可以包括虚拟云网络(VCN)1506(例如,图12的VCN 1206)和安全主机子网1508(例如,图12的安全主机子网1208)。VCN 1506可以包括LPG 1510(例如,图12的LPG 1210),该LPG 1510可以经由包含在SSH VCN 1512(例如,图12的SSH VCN 1212)中的LPG 1510通信地耦合到SSH VCN 1512。SSH VCN 1512可以包括SSH子网1514(例如,图12的SSH子网1214),并且SSH VCN 1512可以经由包含在控制平面VCN 1516中的LPG 1510通信地耦合到控制平面VCN 1516(例如,图12的控制平面VCN 1216)并且经由包含在数据平面VCN 1518中的LPG 1510通信地耦合到数据平面VCN 1518(例如,图12的数据平面1218)。控制平面VCN 1516和数据平面VCN 1518可以包含在服务租赁1519(例如,图12的服务租赁1219)中。

[0145] 控制平面VCN 1516可以包括可以包括(一个或多个)LB子网1522(例如,图12的(一个或多个)LB子网1222)的控制平面DMZ层1520(例如,图12的控制平面DMZ层1220)、可以包括(一个或多个)应用子网1526(例如,图12的(一个或多个)应用子网1226)的控制平面应用层1524(例如,图12的控制平面应用层1224)、可以包括(一个或多个)DB子网1530(例如,图14的(一个或多个)DB子网1430)的控制平面数据层1528(例如,图12的控制平面数据层1228)。包含在控制平面DMZ层1520中的(一个或多个)LB子网1522可以通信地耦合到包含在控制平面应用层1524中的(一个或多个)应用子网1526和可以包含在控制平面VCN 1516中的互联网网关1534(例如,图12的互联网网关1234),并且(一个或多个)应用子网1526可以

通信地耦合到包含在控制平面数据层1528中的(一个或多个)DB子网1530以及服务网关1536(例如,图12的服务网关)和网络地址转换(NAT)网关1538(例如,图12的NAT网关1238)。控制平面VCN 1516可以包括服务网关1536和NAT网关1538。

[0146] 数据平面VCN 1518可以包括数据平面应用层1546(例如,图12的数据平面应用层1246)、数据平面DMZ层1548(例如,图12的数据平面DMZ层1248)),以及数据平面数据层1550(例如,图12的数据平面数据层1250)。数据平面DMZ层1548可以包括(一个或多个)LB子网1522,其可以通信地耦合到数据平面应用层1546的(一个或多个)可信应用子网1560(例如,图14的(一个或多个)可信应用子网1460)和(一个或多个)不可信应用子网1562(例如,图14的(一个或多个)不可信应用子网1462)以及包含在数据平面VCN 1518中的互联网网关1534。(一个或多个)可信应用子网1560可以通信地耦合到包含在数据平面VCN 1518中的服务网关1536、包含在数据平面VCN 1518中的NAT网关1538以及包含在数据平面数据层1550中的(一个或多个)DB子网1530。(一个或多个)不可信应用子网1562可以通信地耦合到包含在数据平面VCN 1518中的服务网关1536和包含在数据平面数据层1550中的(一个或多个)DB子网1530。数据平面数据层1550可以包括可以通信地耦合到包含在数据平面VCN 1518中的服务网关1536的(一个或多个)DB子网1530。

[0147] (一个或多个)不可信应用子网1562可以包括可以通信地耦合到驻留在(一个或多个)不可信应用子网1562内的租户虚拟机(VM)1566(1)-(N)的主VNIC 1564(1)-(N)。每个租户VM 1566(1)-(N)可以在相应的容器1567(1)-(N)中运行代码,并且可通信地耦合到可以包含在容器出口VCN 1568中可以包含的数据平面应用层1546中的应用子网1526。相应的辅助VNIC 1572(1)-(N)可以促进包含在数据平面VCN 1518中的(一个或多个)不可信应用子网1562和包含在容器出口VCN 1568中的应用子网之间的通信。容器出口VCN可以包括可以通信地耦合到公共互联网1554(例如,图12的公共互联网1254)的NAT网关1538。

[0148] 包含在控制平面VCN 1516中和包含在数据平面VCN 1518中的互联网网关1534可以通信地耦合到元数据管理服务1552(例如,图12的元数据管理系统1252),该元数据管理服务1552可以通信地耦合到公共互联网1554。公共互联网1554可以通信地耦合到包含在控制平面VCN 1516中并且包含在数据平面VCN 1518中的NAT网关1538。包含在控制平面VCN 1516中并且包含在数据平面VCN 1518中的服务网关1536可以通信地耦合到云服务1556。

[0149] 在一些示例中,图15的框图1500的体系架构所示的模式可以被认为是图14的框图1400的体系架构所示的模式的例外,并且如果IaaS提供商不能直接与客户(例如,断开连接的区域)通信,那么这种模式可能是IaaS提供商的客户所期望的。客户可以实时访问每个客户的VM 1566(1)-(N)中包含的相应容器1567(1)-(N)。容器1567(1)-(N)可以被配置为对包含在数据平面应用层1546的(一个或多个)应用子网1526中的相应的辅助VNIC 1572(1)-(N)进行调用,该数据平面应用层1546可以包含在容器出口VCN 1568中。辅助VNIC 1572(1)-(N)可以将调用传输到NAT网关1538,NAT网关1538可以将调用传输到公共互联网1554。在这个示例中,可以由客户实时访问的容器1567(1)-(N)可以与控制平面VCN 1516隔离,并且可以与数据平面VCN 1518中包含的其它实体隔离。容器1567(1)-(N)也可以与来自其它客户的资源隔离。

[0150] 在其它示例中,客户可以使用容器1567(1)-(N)来调用云服务1556。在这个示例中,客户可以运行容器1567(1)-(N)中从云服务1556请求服务的代码。容器1567(1)-(N)可

以将该请求传输到辅助VNIC 1572(1)-(N),辅助VNIC 1572(1)-(N)可以将请求传输到NAT网关,该NAT网关可以将请求传输到公共互联网1554。公共互联网1554可以经由互联网网关1534将请求传输到包含在控制平面VCN 1516中的(一个或多个)LB子网1522。响应于确定请求有效,(一个或多个)LB子网可以将请求传输到(一个或多个)应用子网1526,该(一个或多个)应用子网1526可以经由服务网关1536将请求传输到云服务1556。

[0151] 应当认识到的是,各图中描绘的IaaS体系架构1200、1300、1400、1500可以具有除所描绘的组件之外的其它组件。另外,各图中所示的实施例仅仅是可以结合本公开的实施例的云基础设施系统的一些示例。在一些其它实施例中,IaaS系统可以具有比各图中所示的更多或更少的组件、可以组合两个或更多个组件,或者可以具有不同的组件配置或布置。

[0152] 在某些实施例中,本文描述的IaaS系统可以包括以自助服务、基于订阅、弹性可扩展、可靠、高度可用和安全的方式交付给客户的应用套件、中间件和数据库服务产品。此类IaaS系统的示例是本受让人提供的Oracle云基础设施(OCI)。

[0153] 图16图示了其中可以实现本公开的各种实施例的示例计算机系统1600。系统1600可以用于实现上述任何计算机系统。如图所示,计算机系统1600包括经由总线子系统1602与多个外围子系统通信的处理单元1604。这些外围子系统可以包括处理加速单元1606、I/O子系统1608、存储子系统1618和通信子系统1624。存储子系统1618包括有形计算机可读存储介质1622和系统存储器1610。

[0154] 总线子系统1602提供用于让计算机系统1600的各种组件和子系统按意图彼此通信的机制。虽然总线子系统1602被示意性地示出为单条总线,但是总线子系统的替代实施例可以利用多条总线。总线子系统1602可以是若干种类型的总线结构中的任何一种,包括使用各种总线体系架构中的任何一种的存储器总线或存储器控制器、外围总线、以及局部总线。例如,这种体系架构可以包括工业标准体系架构(ISA)总线、微通道体系架构(MCA)总线、增强型ISA(EISA)总线、视频电子标准协会(VESA)局部总线和外围部件互连(PCI)总线,其可以被实现为按IEEE P1386.1标准制造的Mezzanine总线。

[0155] 可以被实现为一个或多个集成电路(例如,常规微处理器或微控制器)的处理单元1604控制计算机系统1600的操作。一个或多个处理器可以被包括在处理单元1604中。这些处理器可以包括单核或多核处理器。在某些实施例中,处理单元1604可以被实现为一个或多个独立的处理单元1632和/或1634,其中在每个处理单元中包括单核或多核处理器。在其它实施例中,处理单元1604也可以被实现为通过将两个双核处理器集成到单个芯片中而形成的四核处理单元。

[0156] 在各种实施例中,处理单元1604可以响应于程序代码执行各种程序并且可以维护多个并发执行的程序或进程。在任何给定的时间,要被执行的程序代码中的一些或全部程序代码可以驻留在(一个或多个)处理器1604中和/或存储子系统1618中。通过适当的编程,(一个或多个)处理器1604可以提供上述各种功能。计算机系统1600可以附加地包括处理加速单元1606,其可以包括数字信号处理器(DSP)、专用处理器,等等。

[0157] I/O子系统1608可以包括用户接口输入设备和用户接口输出设备。用户接口输入设备可以包括键盘、诸如鼠标或轨迹球之类的定点设备、结合到显示器中的触摸板或触摸屏、滚动轮、点击轮、拨盘、按钮、开关、键盘、具有语音命令识别系统的音频输入设备、麦克风以及其它类型的输入设备。用户接口输入设备可以包括,例如,运动感测和/或手势识别

设备，诸如Microsoft**Kinect®**运动传感器，其使得用户能够使用姿势和口头命令通过自然用户接口来控制诸如Microsoft **Xbox®** 360游戏控制器之类的输入设备并与该输入设备交互。用户接口输入设备也可以包括眼睛姿势识别设备，诸如从用户检测眼睛活动(例如，当拍摄照片和/或做出菜单选择时的“眨眼”)并且将眼睛姿势转换为到输入设备(例如，Google**Glass®**)中的输入的Google**Glass®**眨眼检测器。此外，用户接口输入设备可以包括使用户能够通过语音命令与语音识别系统(例如，**Siri®**导航器)交互的语音识别感测设备。

[0158] 用户接口输入设备也可以包括但不限于三维(3D)鼠标、操纵杆或指向棒、游戏面板和绘图板，以及音频/视觉设备，诸如扬声器、数码相机、数码摄像机、便携式媒体播放器、网络摄像头、图像扫描仪、指纹扫描仪、条形码阅读器3D扫描仪、3D打印机、激光测距仪和视线跟踪设备。此外，用户接口输入设备可以包括，例如，医学成像输入设备，诸如计算机断层扫描、磁共振成像、正电子发射断层扫描、医疗超声设备。用户接口输入设备也可以包括，例如，诸如MIDI键盘、数字乐器等之类的音频输入设备。

[0159] 用户接口输出设备可以包括显示子系统、指示灯，或者诸如音频输出设备之类的非可视显示器，等等。显示子系统可以是阴极射线管(CRT)、诸如使用液晶显示器(LCD)或等离子显示器的平板设备、投影设备、触摸屏，等等。一般而言，术语“输出设备”的使用意在包括用于从计算机系统1600向用户或其它计算机输出信息的所有可能类型的设备和机制。例如，用户接口输出设备可以包括，但不限于，可视地传达文本、图形和音频/视频信息的各种显示设备，诸如监视器、打印机、扬声器、耳机、汽车导航系统、绘图仪、语音输出设备，以及调制解调器。

[0160] 计算机系统1600可以包括包含被示为当前位于系统存储器1610内的软件元件的存储子系统1618。系统存储器1610可以存储可加载并且可在处理单元1604上执行的程序指令，以及在这些程序的执行期间所产生的数据。

[0161] 取决于计算机系统1600的配置和类型，系统存储器1610可以是易失性的(诸如随机存取存储器(RAM))和/或非易失性的(诸如只读存储器(ROM)、闪存存储器，等等)。RAM通常包含可被处理单元1604立即访问和/或目前正被处理单元1604操作和执行的数据和/或程序模块。在一些实施方式中，系统存储器1610可以包括多种不同类型的存储器，例如静态随机存取存储器(SRAM)或动态随机存取存储器(DRAM)。在一些实施方式中，包含有助于诸如在启动期间在计算机系统1600内的元件之间传送信息的基本例程的基本输入/输出系统(BIOS)通常可以被存储在ROM中。作为示例，但不是限制，系统存储器1610也示出了可以包括客户端应用、web浏览器、中间层应用、关系数据库管理系统(RDBMS)等的应用程序1612、程序数据1614以及操作系统1616。作为示例，操作系统1616可以包括各种版本的Microsoft **Windows®**、Apple **Macintosh®**和/或Linux操作系统、各种可商业获得的 **UNIX®** 或类UNIX操作系统(包括但不限于各种GNU/Linux操作系统、Google **Chrome®** OS等)和/或诸如iOS、**Windows®** Phone、**Android®** OS、**BlackBerry®** 10OS和**Palm®** OS操作系统之类的移动操作系统。

[0162] 存储子系统1618也可以提供用于存储提供一些实施例的功能的基本编程和数据结构的有形计算机可读存储介质。当被处理器执行时提供上述功能的软件(程序、代码模

块、指令)可以被存储在存储子系统1618中。这些软件模块或指令可以被处理单元1604执行。存储子系统1618也可以提供用于存储根据本公开被使用的数据的储存库。

[0163] 存储子系统1600也可以包括可被进一步连接到计算机可读存储介质1622的计算机可读存储介质读取器1620。与系统存储器1610一起,并且可选地与其相结合,计算机可读存储介质1622可以全面地表示用于临时和/或更持久地包含、存储、发送和检索计算机可读信息的远程、本地、固定和/或可移除存储设备加存储介质。

[0164] 包含代码或代码的部分的计算机可读存储介质1622也可以包括本领域已知或使用的任何适当的介质,包括存储介质和通信介质,诸如但不限于,以用于信息的存储和/或传输的任何方法或技术实现的易失性和非易失性、可移除和不可移除介质。这可以包括有形的计算机可读存储介质,诸如RAM、ROM、电可擦除可编程ROM(EEPROM)、闪存存储器或其它存储器技术、CD-ROM、数字多功能盘(DVD)或其它光学存储装置、磁带盒、磁带、磁盘存储装置或其它磁存储设备,或者其它有形的计算机可读介质。这也可以包括非有形的计算机可读介质,诸如数据信号、数据传输,或者可以被用来发送期望信息并且可以被计算系统1600访问的任何其它介质。

[0165] 举例来说,计算机可读存储介质1622可以包括从不可移除的非易失性磁介质读取或写到其的硬盘驱动器、从可移除的非易失性磁盘读取或写到其的磁盘驱动器、以及从可移除的非易失性光盘(诸如CD ROM、DVD和**Blu-Ray®**盘或其它光学介质)读取或写到其的光盘驱动器。计算机可读存储介质1622可以包括,但不限于,**Zip®**驱动器、闪存卡、通用串行总线(USB)闪存驱动器、安全数字(SD)卡、DVD盘、数字视频带,等等。计算机可读存储介质1622也可以包括基于非易失性存储器的固态驱动器(SSD)(诸如基于闪存存储器的SSD、企业闪存驱动器、固态ROM等)、基于易失性存储器的SSD(诸如固态RAM、动态RAM、静态RAM)、基于DRAM的SSD、磁阻RAM(MRAM)SSD,以及使用基于DRAM和闪存存储器的SSD的组合的混合SSD。盘驱动器及其关联的计算机可读介质可以为计算机系统1600提供计算机可读指令、数据结构、程序模块及其它数据的非易失性存储。

[0166] 通信子系统1624提供到其它计算机系统和网络的接口。通信子系统1624用作用于从其它系统接收数据和从计算机系统1600向其它系统发送数据的接口。例如,通信子系统1624可以使计算机系统1600能够经由互联网连接到一个或多个设备。在一些实施例中,通信子系统1624可以包括用于(例如,使用蜂窝电话技术,诸如3G、4G或EDGE(用于全球演进的增强型数据速率)之类的先进数据网络技术,WiFi(IEEE 802.11系列标准),或其它移动通信技术,或其任意组合)访问无线语音和/或数据网络的射频(RF)收发器组件、全球定位系统(GPS)接收器组件和/或其它组件。在一些实施例中,作为无线接口的附加或者替代,通信子系统1624可以提供有线网络连接(例如,以太网)。

[0167] 在一些实施例中,通信子系统1624也可以代表可以使用计算机系统1600的一个或多个用户接收结构化和/或非结构化数据馈送1626、事件流1628、事件更新1630等形式的输入通信。

[0168] 举例来说,通信子系统1624可以被配置为实时地从社交网络和/或其它通信服务的用户接收数据馈送1626,诸如**Twitter®**馈送、**Facebook®**更新、诸如丰富站点摘要(RSS)馈送之类的web馈送和/或来自一个或多个第三方信息源的实时更新。

[0169] 此外，通信子系统1624也可被配置为接收连续数据流形式的数据，这可以包括本质上可以是连续的或无界的没有明确终止的实时事件的事件流1628和/或事件更新1630。产生连续数据的应用的示例可以包括，例如，传感器数据应用、金融报价机、网络性能测量工具(例如，网络监视和流量管理应用)、点击流分析工具、汽车流量监视，等等。

[0170] 通信子系统1624也可被配置为向一个或多个数据库输出结构化和/或非结构化数据馈送1626、事件流1628、事件更新1630，等等，该一个或多个数据库可以与耦合到计算机系统1600的一个或多个流式数据源计算机通信。

[0171] 计算机系统1600可以是各种类型之一，包括手持便携式设备(例如，iPhone®蜂窝电话、iPad®计算平板电脑、PDA)、可穿戴设备(例如，Google®Glass头戴式显示器)、PC、工作站、大型机、信息站、服务器机架、或任何其它数据处理系统。

[0172] 由于计算机和网络的不断变化的本质，在图中绘出的计算机系统1600的描述仅仅作为具体的示例。具有比图中绘出的系统更多或更少组件的许多其它配置是可能的。例如，定制的硬件也可以被使用和/或特定的元素可以用硬件、固件、软件(包括小应用程序)或其组合来实现。另外，也可以采用到诸如网络输入/输出设备之类的其它计算设备的连接。基于本文提供的公开内容和教导，本领域普通技术人员将认识到实现各种实施例的其它方式和/或方法。

[0173] 虽然已经描述了本公开的具体实施例，但是各种修改、变更、替代构造和等效形式也包含在本公开的范围内。本公开的实施例不限于在某些特定数据处理环境内操作，而是可以在多个数据处理环境内自由操作。此外，虽然已经使用特定系列的事务和步骤描述了本公开的实施例，但是本领域技术人员应该清楚本公开的范围不限于所描述系列的事务和步骤。上述实施例的各种特征和方面可以单独或联合使用。

[0174] 另外，虽然已经使用硬件和软件的特定组合描述了本公开的实施例，但是应当认识到硬件和软件的其它组合也在本公开的范围内。本公开的实施例可以仅用硬件、或仅用软件、或使用它们的组合来实现。本文描述的各种处理可以以任何组合在相同的处理器或在不同的处理器上实现。相应地，在组件或模块被描述为被配置为执行某些操作的情况下，可以例如通过设计电子电路来执行操作、通过对可编程电子电路(诸如微处理器)进行编程来执行操作，或其任何组合来完成这样的配置。进程可以使用多种技术进行通信，这些技术包括但不限于用于进程间通信的常规技术，并且不同的进程对可以使用不同的技术，或者同一对进程可以在不同时间使用不同的技术。

[0175] 相应地，说明书和附图被认为是说明性的而不是限制性的。但是，显然可以对其进行添加、减少、删除和其它修改和改变而不背离权利要求中阐述的更广泛的精神和范围。因此，虽然已经描述了具体的公开实施例，但这些并不旨在进行限制。各种修改和等效形式都在以下权利要求的范围内。

[0176] 在描述所公开的实施例的上下文中(尤其在以下权利要求的上下文中)使用术语“一”和“一个”和“该”以及类似的所指对象要被解释为涵盖单数和复数，除非本文另有指示或与上下文明显矛盾。除非另有说明，否则术语“包括”、“具有”、“包含”和“含有”要被解释为开放式术语(即，意思是“包括但不限于”)。术语“连接”应被解释为部分或全部包含在内、附接到或连接在一起，即使中间存在一些东西。除非本文另有指示，否则本文中值范围的记载仅旨在用作单独引用落入该范围内的每个单独值的便捷方法，并且每个单独值被并入说

明书中，就好像它在本文中单独记载一样。除非本文另有指示或以其他方式与上下文明显矛盾，否则本文所述的所有方法都可以以任何合适的顺序执行。本文提供的任何和所有示例或示例性语言(例如，“诸如”)的使用仅旨在更好地阐明本公开的实施例并且不对本公开的范围构成限制，除非另有声明。说明书中的任何语言都不应被解释为指示任何未要求保护的元素对于本公开的实践是必不可少的。

[0177] 析取语言，诸如短语“X、Y或Z中的至少一个”，除非另有明确说明，否则意图在一般用于表示项目、术语等可以是X、Y或Z，或它们的任何组合(例如，X、Y和/或Z)的上下文中理解。因此，这种析取语言通常不旨在也不应暗示某些实施例需要X中的至少一个、Y中的至少一个或Z中的至少一个各自存在。

[0178] 本文描述了本公开的优选实施例，包括已知用于实施本公开的最佳模式。那些优选实施例的变型对于本领域普通技术人员在阅读上述描述后会变得明显。普通技术人员应该能够适当地采用这样的变型并且可以以不同于本文具体描述的方式来实践本公开。相应地，本公开包括在适用法律允许的情况下对所附权利要求中记载的主题的所有修改和等效形式。此外，除非在本文中另有指示，否则本公开涵盖在其所有可能的变型中的上述元素的任何组合。

[0179] 本文引用的所有参考文献，包括出版物、专利申请和专利，均以相同的程度通过引用并入本文，就好像每个参考文献单独且具体地指示为通过引用并入并在本文中整体阐述一样。

[0180] 在前述的说明书中，本公开的各方面参考其具体实施例进行了描述，但本领域技术人员将认识到的是，本公开不限于此。上述公开的各个特征和方面可以被单独或联合使用。此外，在不脱离本说明书的更广泛精神和范围的情况下，实施例可以在除本文所述的环境和应用之外的任何数量的环境和应用中被使用。相应地，本说明书和附图应当被认为是说明性而不是限制性的。

[0181] 以下条款描述了所公开的实施方式的实施例：

[0182] 条款1：一种方法，包括：

[0183] 由计算机系统在第一地理区域和第一逻辑时间创建块卷的第一快照，该块卷包括多个分区；

[0184] 由计算机系统将与第一快照对应的第一快照数据传输到第二地理区域处的对象存储系统；

[0185] 由计算机系统在第一地理区域和第二逻辑时间创建块卷的第二快照；

[0186] 由计算机系统生成多个增量，所述多个增量中的每个增量与所述多个分区中的分区对应；

[0187] 由计算机系统将与所述多个增量对应的多个增量数据集传输到第二地理区域处的对象存储系统；

[0188] 由计算机系统至少部分地通过聚集与所述多个增量和第一快照相关联的对象元数据来生成检查点；

[0189] 由计算机系统接收生成还原卷的还原请求；以及

[0190] 由计算机系统从检查点生成还原卷。

[0191] 条款2：条款1的方法，其中生成所述多个增量包括：

[0192] 生成第二快照与第一快照之间的比较；

[0193] 基于比较确定与第一快照数据和第二快照所对应的第二快照数据之间的改变对应的经修改的数据；以及

[0194] 生成所述多个增量，其中所述多个增量描述用于所述多个分区的经修改的数据。

[0195] 条款3：条款1的方法，其中创建第一快照包括：

[0196] 与逻辑时间对应地暂停用于所述多个分区的输入/输出操作；

[0197] 生成描述所述多个分区中的卷数据的多个块映像；以及

[0198] 启用用于所述多个分区的输入/输出操作。

[0199] 条款4：条款1的方法，其中还原请求是故障转移请求，该方法还包括：

[0200] 使得能够在第二地理区域处生成还原卷；以及

[0201] 使得能够在第二地理区域使用还原卷进行输入/输出操作。

[0202] 条款5：条款1的方法，其中还原请求是故障恢复请求，该方法还包括：

[0203] 在第二地理区域处生成还原卷；

[0204] 使得能够在第一地理区域处生成故障恢复卷；以及

[0205] 在第一地理区域处还原第一快照数据。

[0206] 条款6：条款1的方法，其中传输所述多个增量数据集包括：

[0207] 从所述多个增量数据集生成多个分块对象；

[0208] 传送所述多个增量；以及

[0209] 将所述多个分块对象传送到对象存储系统。

[0210] 条款7：条款6的方法，其中：

[0211] 检查点包括对象元数据的清单；以及

[0212] 对象元数据包括与对象存储系统中的所述多个分块对象对应的分块指针。

[0213] 条款18：条款7的方法，其中聚集对象元数据包括更新清单以反映所述多个增量数据集与第一快照数据之间的多个差异。

[0214] 条款9：一种计算机系统，包括

[0215] 一个或多个处理器；

[0216] 存储器，与所述一个或多个处理器通信，该存储器被配置为存储计算机可执行指令，其中执行计算机可执行指令使得所述一个或多个处理器执行包括以下的步骤：

[0217] 由计算机系统在第一地理区域和第一逻辑时间创建块卷的第一快照，该块卷包括多个分区；

[0218] 由计算机系统将与第一快照对应的第一快照数据传输到第二地理区域处的对象存储系统；

[0219] 由计算机系统在第一地理区域和第二逻辑时间创建块卷的第二快照；

[0220] 由计算机系统生成多个增量，所述多个增量中的每个增量与所述多个分区中的分区对应；

[0221] 由计算机系统将与所述多个增量对应的多个增量数据集传输到第二地理区域处的对象存储系统；

[0222] 由计算机系统至少部分地通过聚集与所述多个增量和第一快照相关联的对象元数据来生成检查点；

[0223] 由计算机系统接收生成还原卷的还原请求;以及

[0224] 由计算机系统从检查点生成还原卷。

[0225] 条款10:条款9的计算机系统,其中生成所述多个增量包括:

[0226] 生成第二快照与第一快照之间的比较;

[0227] 基于比较确定与第一快照数据和第二快照所对应的第二快照数据之间的改变对应的经修改的数据;以及

[0228] 生成所述多个增量,其中所述多个增量描述用于所述多个分区的经修改的数据。

[0229] 条款11:条款9的计算机系统,其中创建第一快照包括:

[0230] 与逻辑时间对应地暂停用于所述多个分区的输入/输出操作;

[0231] 生成描述所述多个分区中的卷数据的多个块映像;以及

[0232] 启用用于所述多个分区的输入/输出操作。

[0233] 条款12:条款9的计算机系统,其中还原请求是故障转移请求,该方法还包括:

[0234] 使得能够在第二地理区域处生成还原卷;以及

[0235] 使得能够在第二地理区域使用还原卷进行输入/输出操作。

[0236] 条款13:条款9的计算机系统,其中还原请求是故障恢复请求,该方法还包括:

[0237] 在第二地理区域处生成还原卷;

[0238] 使得能够在第一地理区域处生成故障恢复卷;以及

[0239] 在第一地理区域处还原第一快照数据。

[0240] 条款14:条款9的计算机系统,其中传输所述多个增量数据集包括:

[0241] 从所述多个增量数据集生成多个分块对象;

[0242] 传送所述多个增量;以及

[0243] 将所述多个分块对象传送到对象存储系统。

[0244] 条款15:条款14的计算机系统,其中检查点包括对象元数据的清单;以及

[0245] 对象元数据包括与对象存储系统中的所述多个分块对象对应的分块指针。

[0246] 条款16:条款15的计算机系统,其中聚集对象元数据包括更新清单以反映所述多个增量数据集与第一快照数据之间的多个差异。

[0247] 条款17:一种存储计算机可执行指令的计算机可读存储介质,指令在被执行时使得计算机系统的一个或多个处理器执行包括以下的步骤:

[0248] 由计算机系统在第一地理区域和第一逻辑时间创建块卷的第一快照,该块卷包括多个分区;

[0249] 由计算机系统将与第一快照对应的第一快照数据传输到第二地理区域处的对象存储系统;

[0250] 由计算机系统在第一地理区域和第二逻辑时间创建块卷的第二快照;

[0251] 由计算机系统生成多个增量,所述多个增量中的每个增量与所述多个分区中的分区对应;

[0252] 由计算机系统将与所述多个增量对应的多个增量数据集传输到第二地理区域处的对象存储系统;

[0253] 由计算机系统至少部分地通过聚集与所述多个增量和第一快照相关联的对象元数据来生成检查点;

[0254] 由计算机系统接收生成还原卷的还原请求;以及

[0255] 由计算机系统从检查点生成还原卷。

[0256] 条款18:条款17的计算机可读存储介质,其中生成所述多个增量包括:

[0257] 生成第二快照与第一快照之间的比较;

[0258] 基于比较确定与第一快照数据和第二快照所对应的第二快照数据之间的改变对应的经修改的数据;以及

[0259] 生成所述多个增量,其中所述多个增量描述用于所述多个分区的经修改的数据。

[0260] 条款19:条款17的计算机可读存储介质,其中传输所述多个增量数据集包括:

[0261] 从所述多个增量数据集生成多个分块对象;

[0262] 传送所述多个增量;以及

[0263] 将所述多个分块对象传送到对象存储系统。

[0264] 条款20:条款19的计算机可读存储介质,其中:

[0265] 检查点包括对象元数据的清单;以及

[0266] 对象元数据包括与对象存储系统中的所述多个分块对象对应的分块指针。

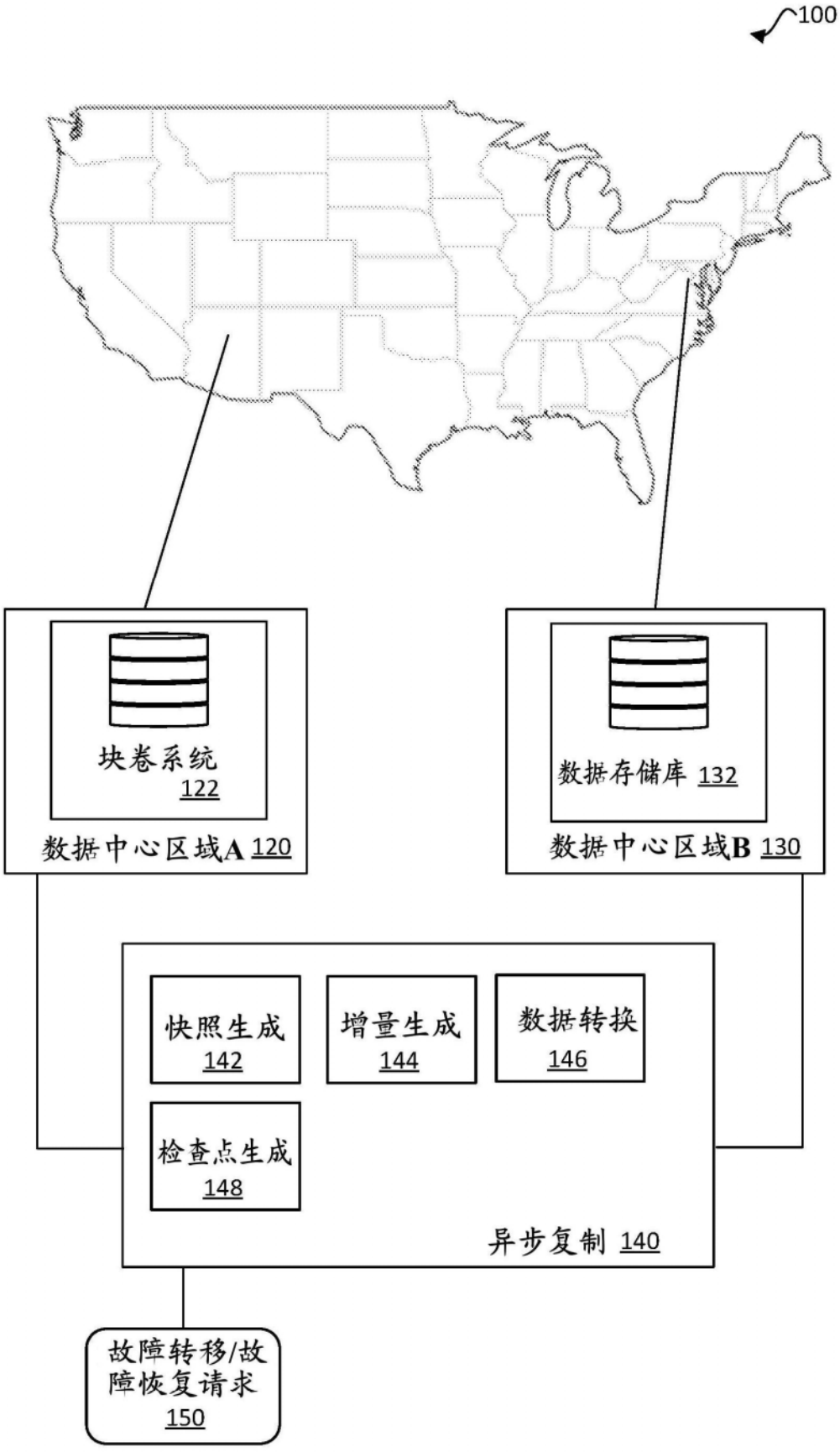

100
块卷系统
122
数据中心区域A 120
数据存储库 132
数据中心区域B 130
快照生成
142
增量生成
144
数据转换
146
检查点生成
148
异步复制 140
故障转移/故障恢复请求
150

图1

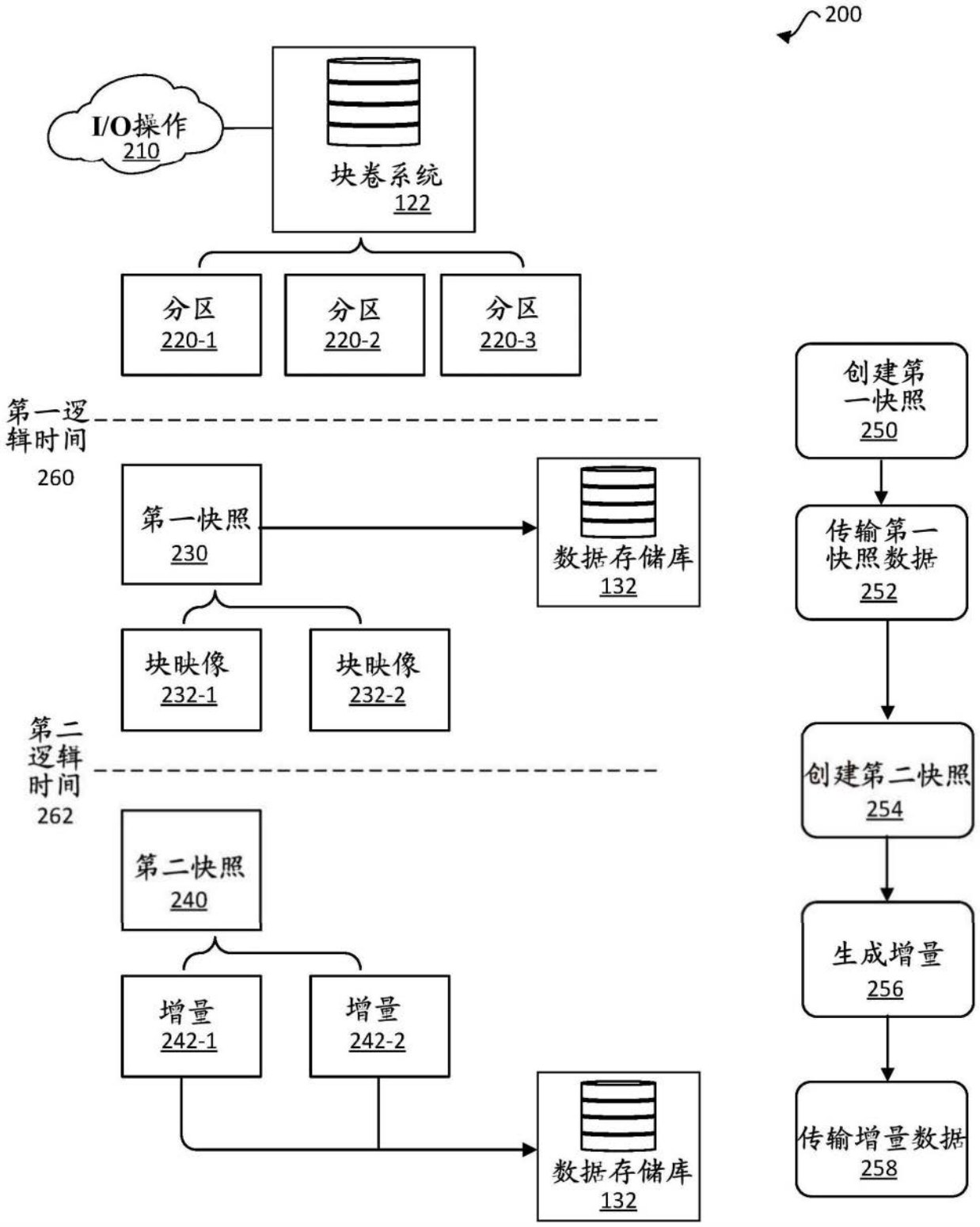

200
I/O操作
210
块卷系统
122
分区
220-1
分区
220-2
分区
220-3
第一逻辑时间
260
第一快照
230
数据存储库
132
块映像
232-1
块映像
232-2
第二逻辑时间
262
第二快照
240
增量
242-1
增量
242-2
数据存储库
132
创建第一快照
250
传输第一快照数据
252
创建第二快照
254
生成增量
256
传输增量数据
258

图2

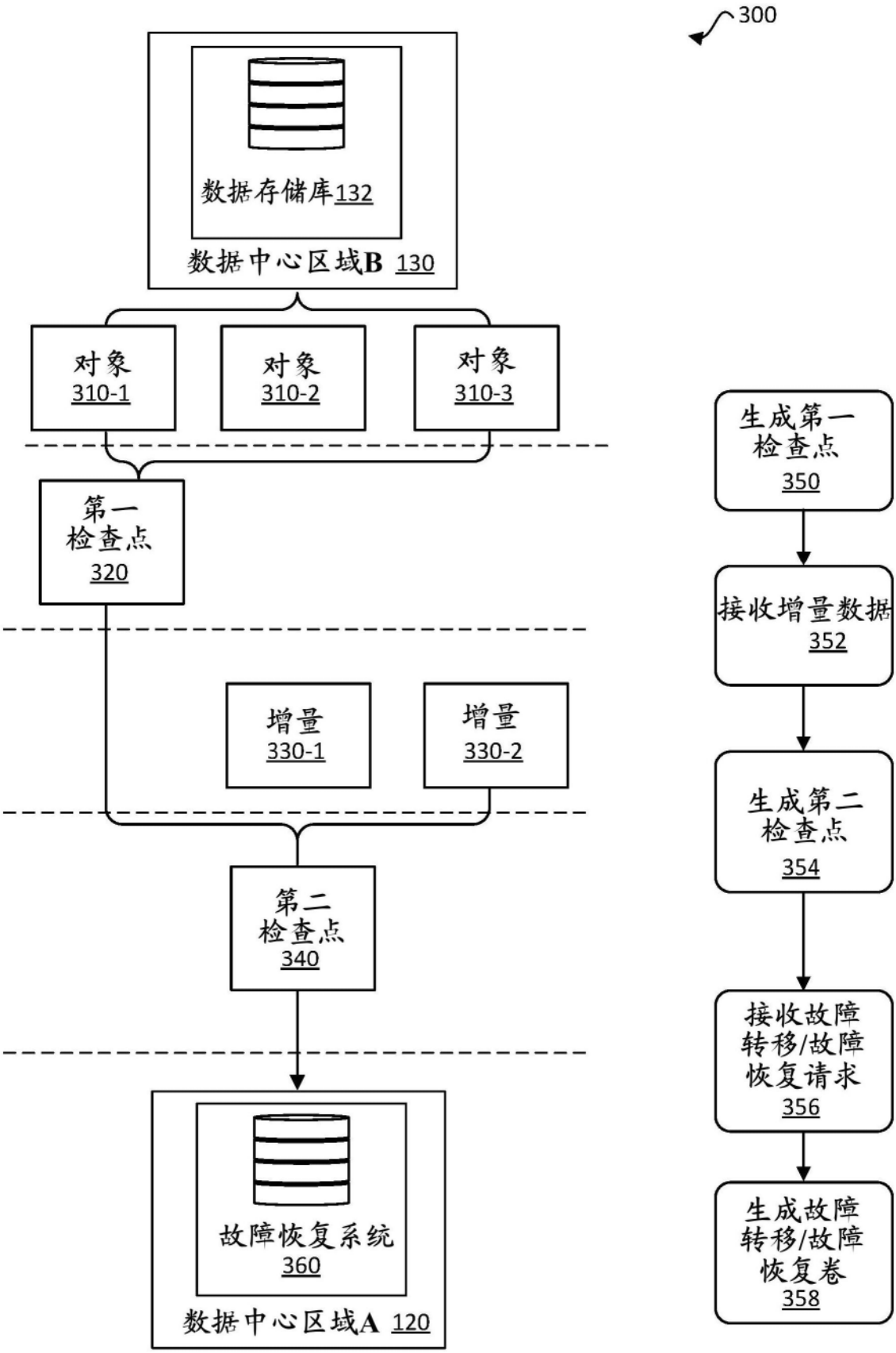

300
数据存储库132
数据中心区域B 130
对象 310-1
对象 310-2
对象 310-3
第一检查点 320
增量 330-1
增量 330-2
第二检查点 340
故障恢复系统 360
数据中心区域A 120
生成第一检查点 350
接收增量数据 352
生成第二检查点 354
接收故障转移/故障恢复请求 356
生成故障转移/故障恢复卷 358

图3

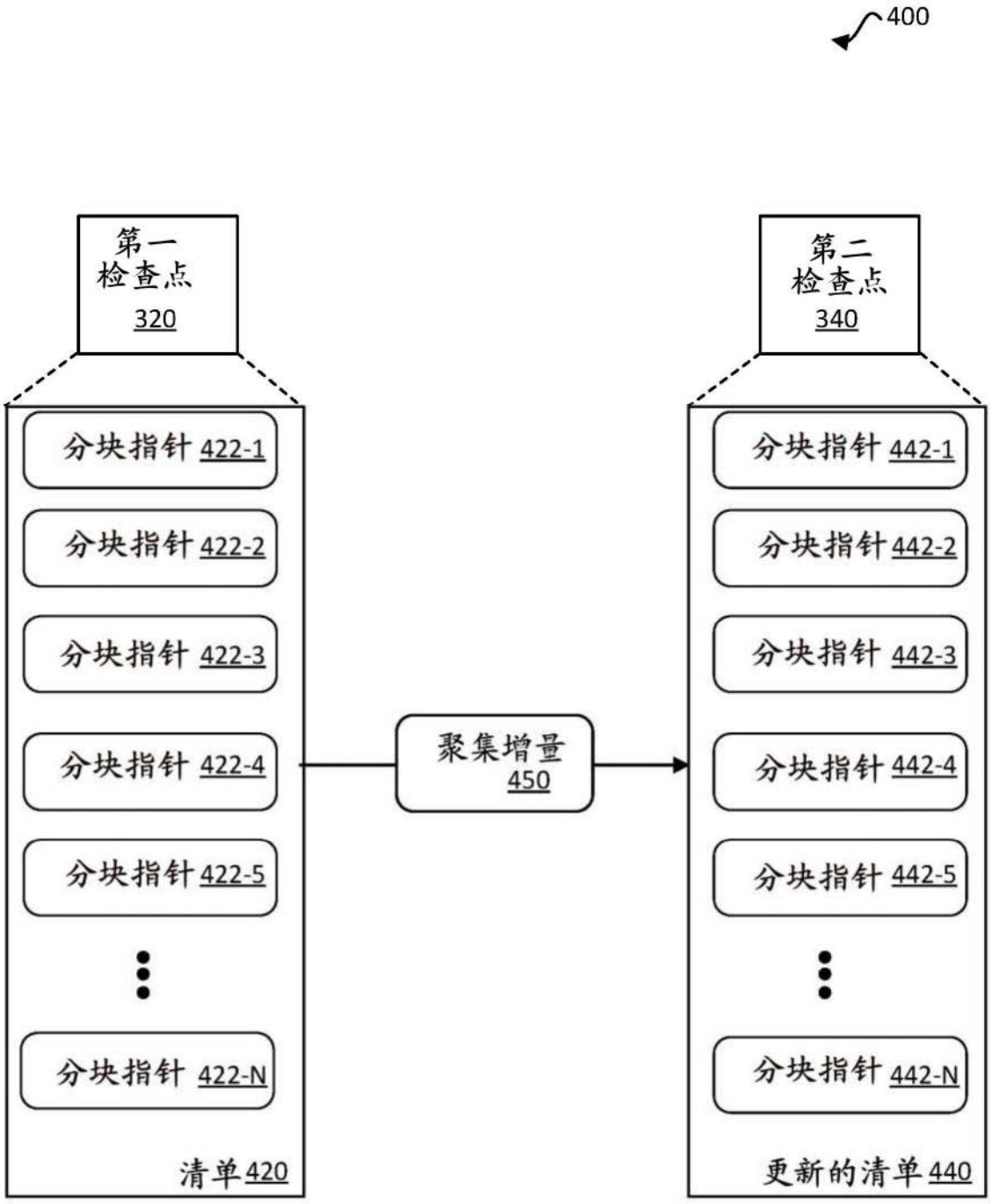

400
第一
检查点
320
第二
检查点
340
分块指针 422-1
分块指针 422-2
分块指针 422-3
分块指针 422-4
分块指针 422-5
分块指针 422-N
清单 420
聚集增量
450
分块指针 442-1
分块指针 442-2
分块指针 442-3
分块指针 442-4
分块指针 442-5
分块指针 442-N
更新的清单 440

图4

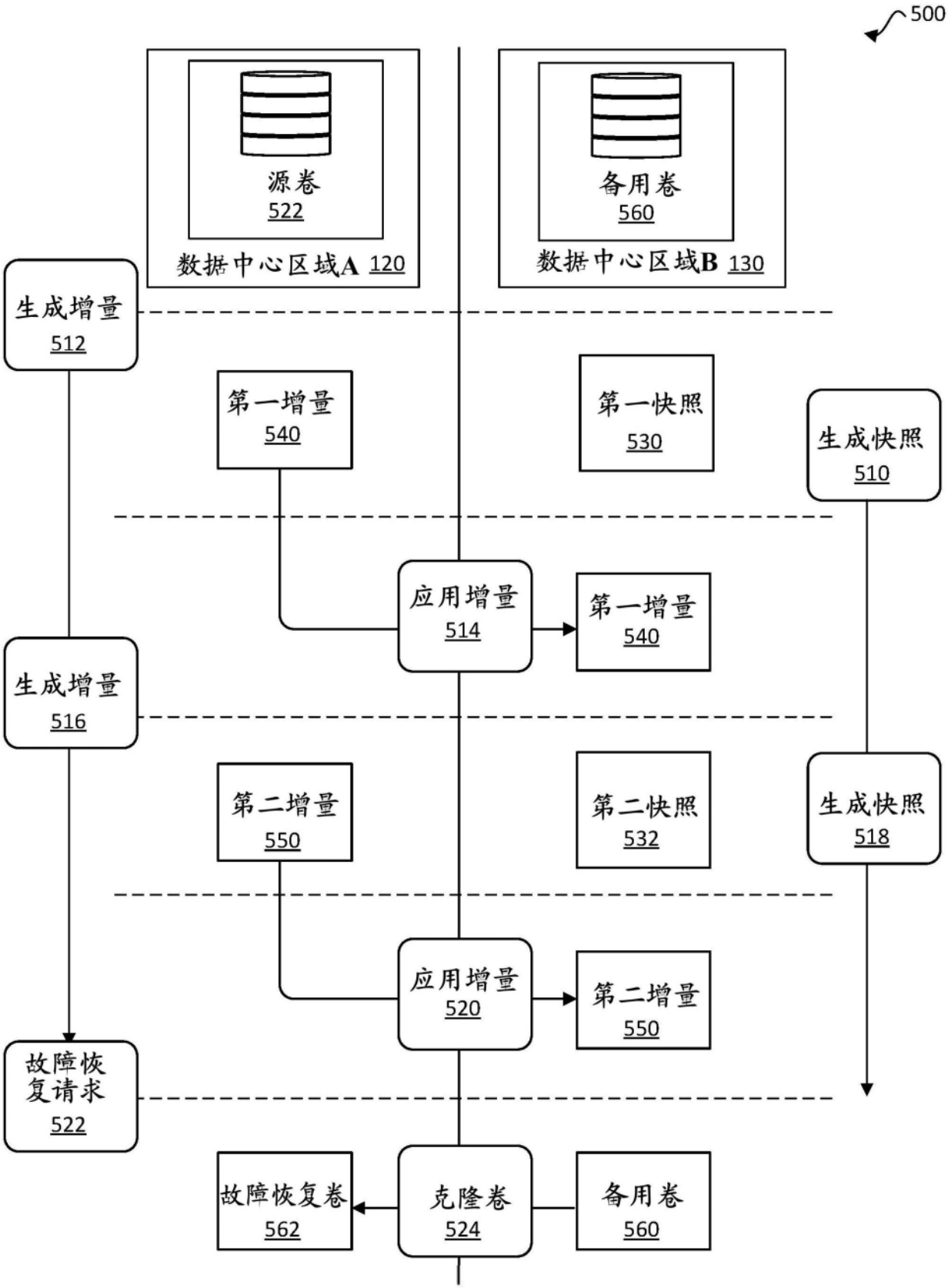

500
源卷
522
数据中心区域A 120
备用卷
560
数据中心区域B 130
生成增量
512
第一增量
540
第一快照
530
生成快照
510
应用增量
514
第一增量
540
生成增量
516
第二增量
550
第二快照
532
生成快照
518
应用增量
520
第二增量
550
故障恢复请求
522
故障恢复卷
562
克隆卷
524
备用卷
560

图5

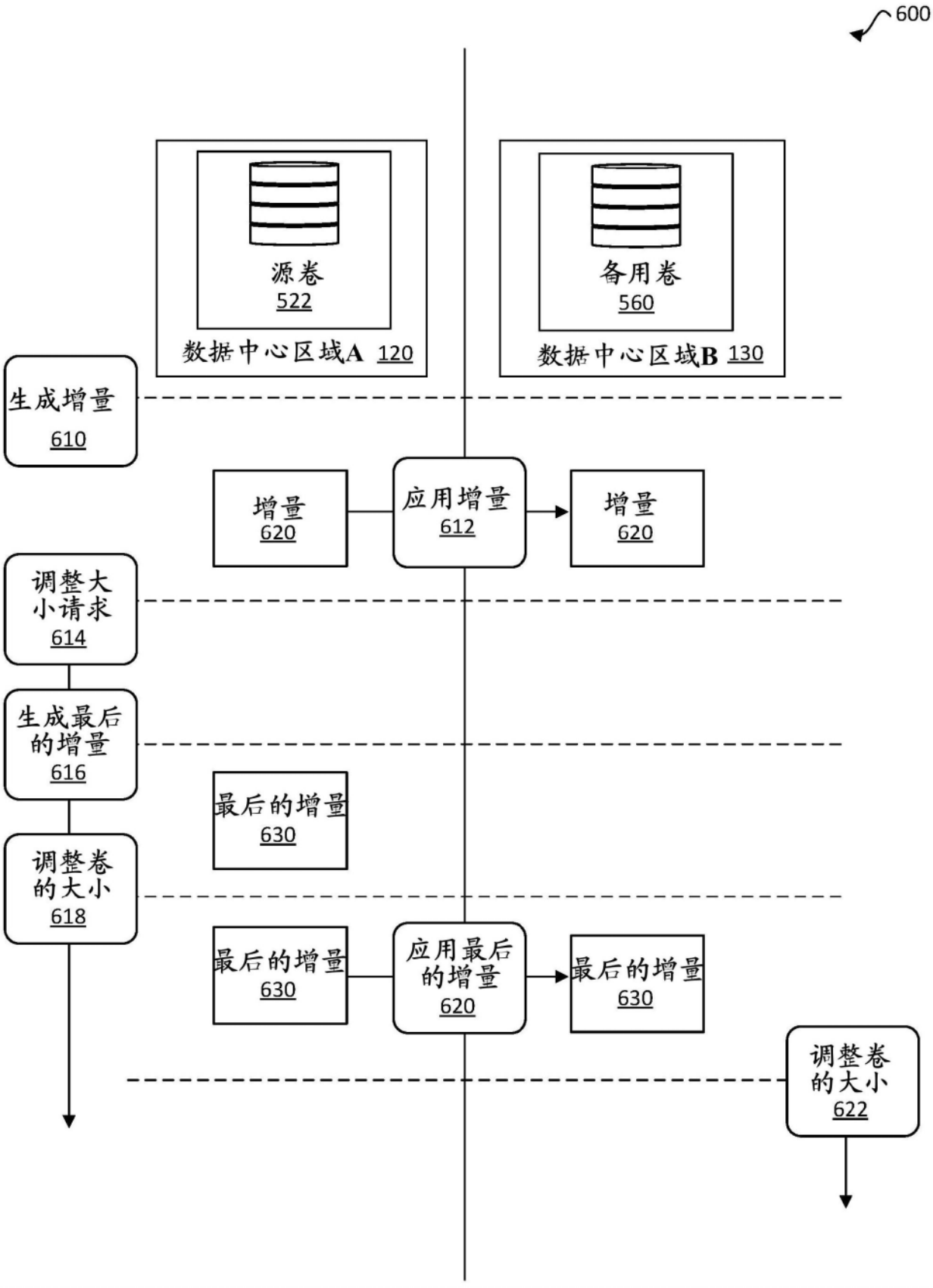

600
源卷
522
数据中心区域A 120
备用卷
560
数据中心区域B 130
生成增量
610
增量
620
应用增量
612
增量
620
调整大小请求
614
生成最后的增量
616
最后的增量
630
调整卷的大小
618
最后的增量
630
应用最后的增量
620
最后的增量
630
调整卷的大小
622

图6

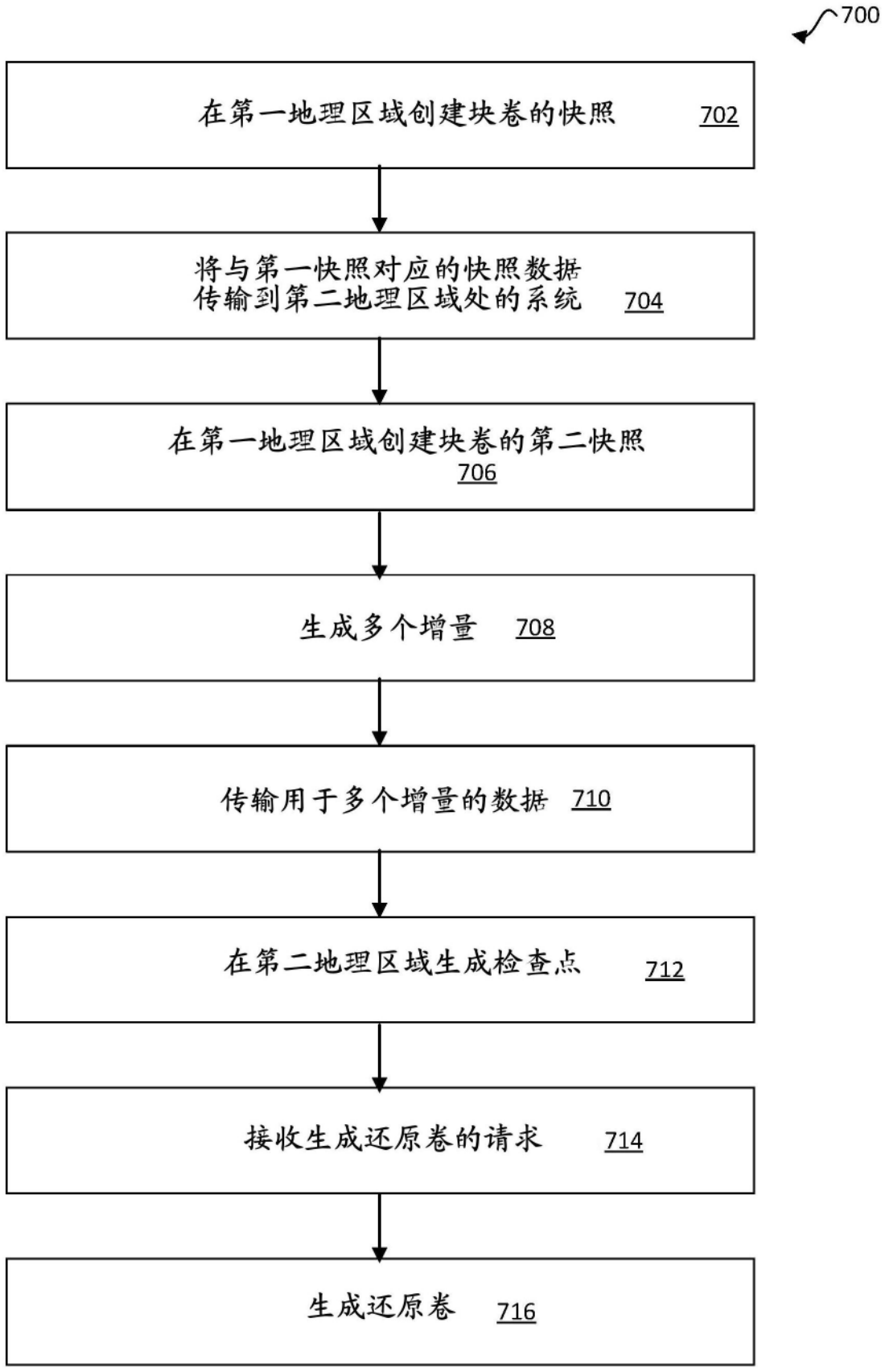

700
在第一地理区域创建块卷的快照 702
将与第一快照对应的快照数据传输到第二地理区域处的系统 704
在第一地理区域创建块卷的第二快照 706
生成多个增量 708
传输用于多个增量的数据 710
在第二地理区域生成检查点 712
接收生成还原卷的请求 714
生成还原卷 716

图7

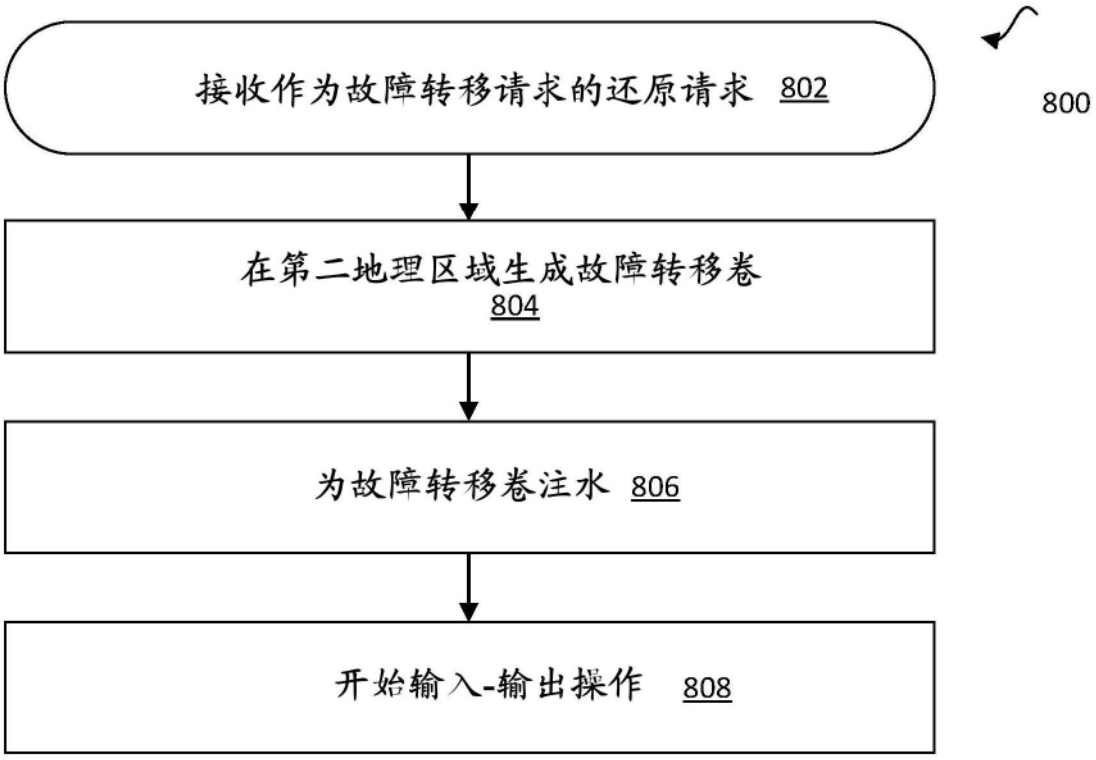

接收作为故障转移请求的还原请求 802
800
在第二地理区域生成故障转移卷 804
为故障转移卷注水 806
开始输入-输出操作 808

图8

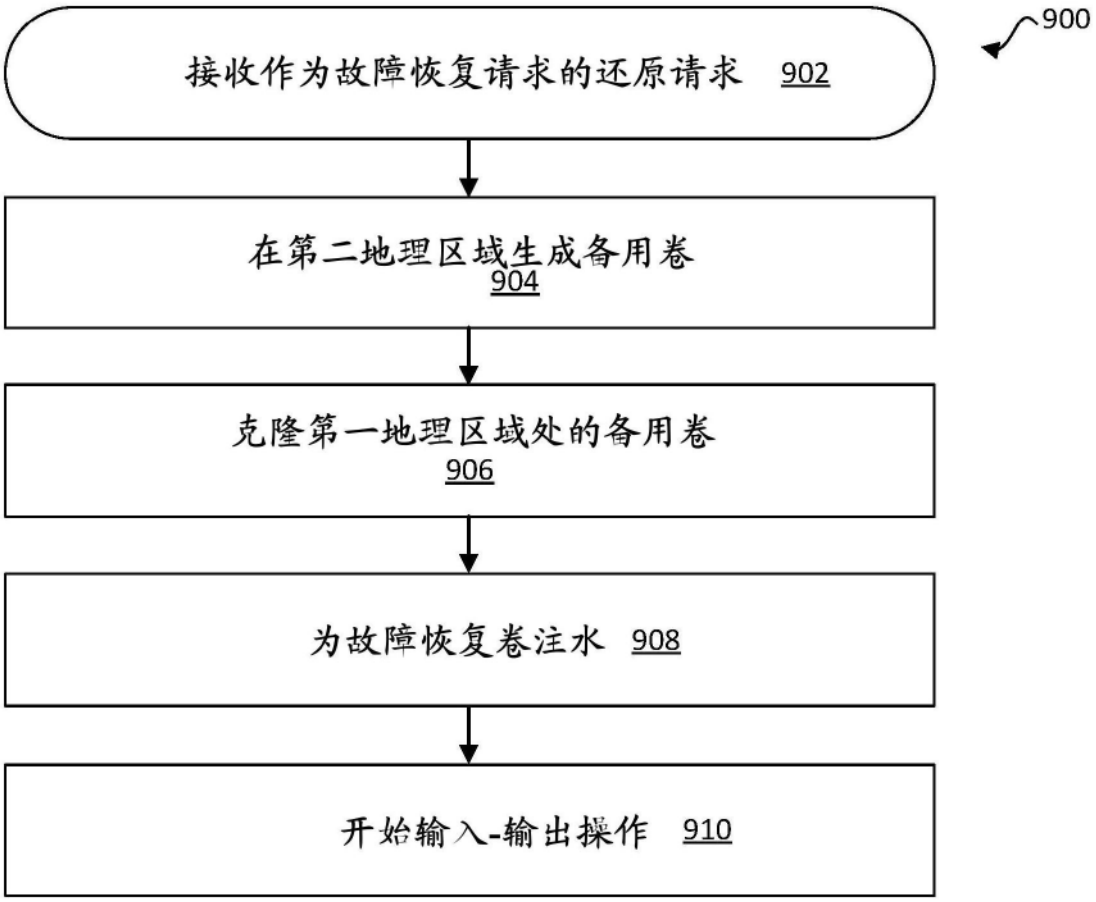

900
接收作为故障恢复请求的还原请求 902
在第二地理区域生成备用卷 904
克隆第一地理区域处的备用卷 906
为故障恢复卷注水 908
开始输入-输出操作 910

图9

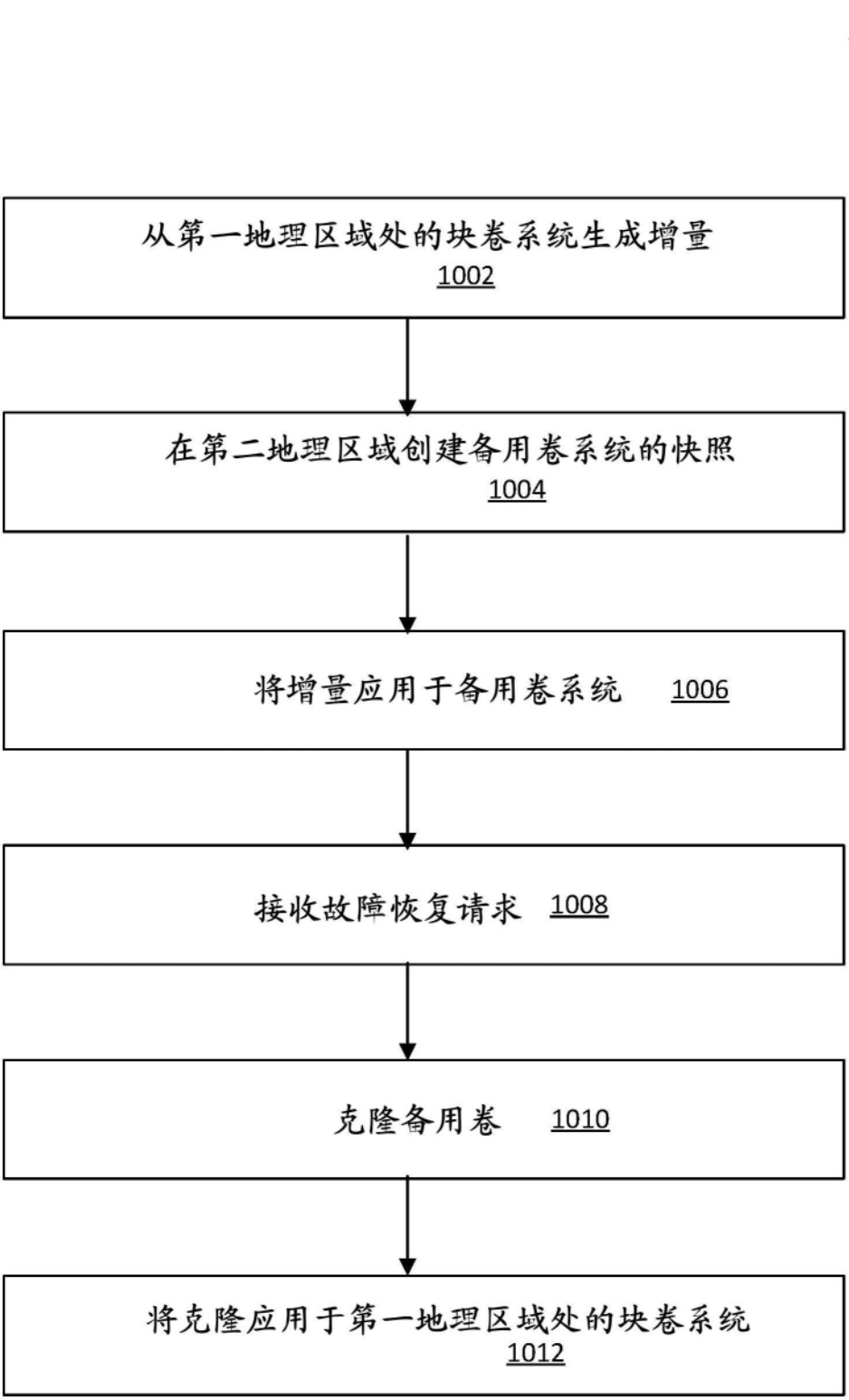

1000
从第一地理区域处的块卷系统生成增量
1002
在第二地理区域创建备用卷系统的快照
1004
将增量应用于备用卷系统 1006
接收故障恢复请求 1008
克隆备用卷 1010
将克隆应用于第一地理区域处的块卷系统
1012

图10

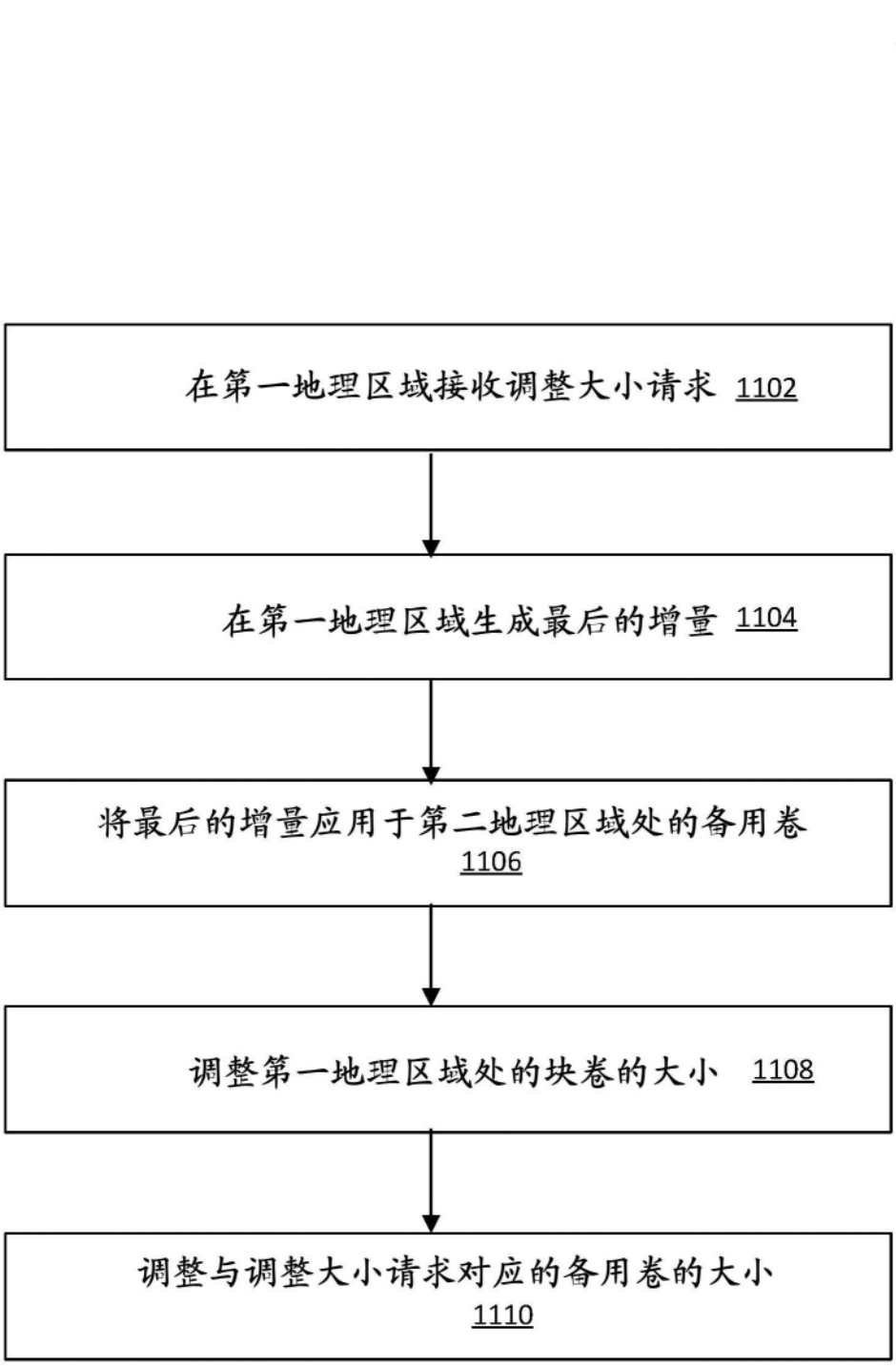

1100
在第一地理区域接收调整大小请求 1102
在第一地理区域生成最后的增量 1104
将最后的增量应用于第二地理区域处的备用卷 1106
调整第一地理区域处的块卷的大小 1108
调整与调整大小请求对应的备用卷的大小 1110

图11

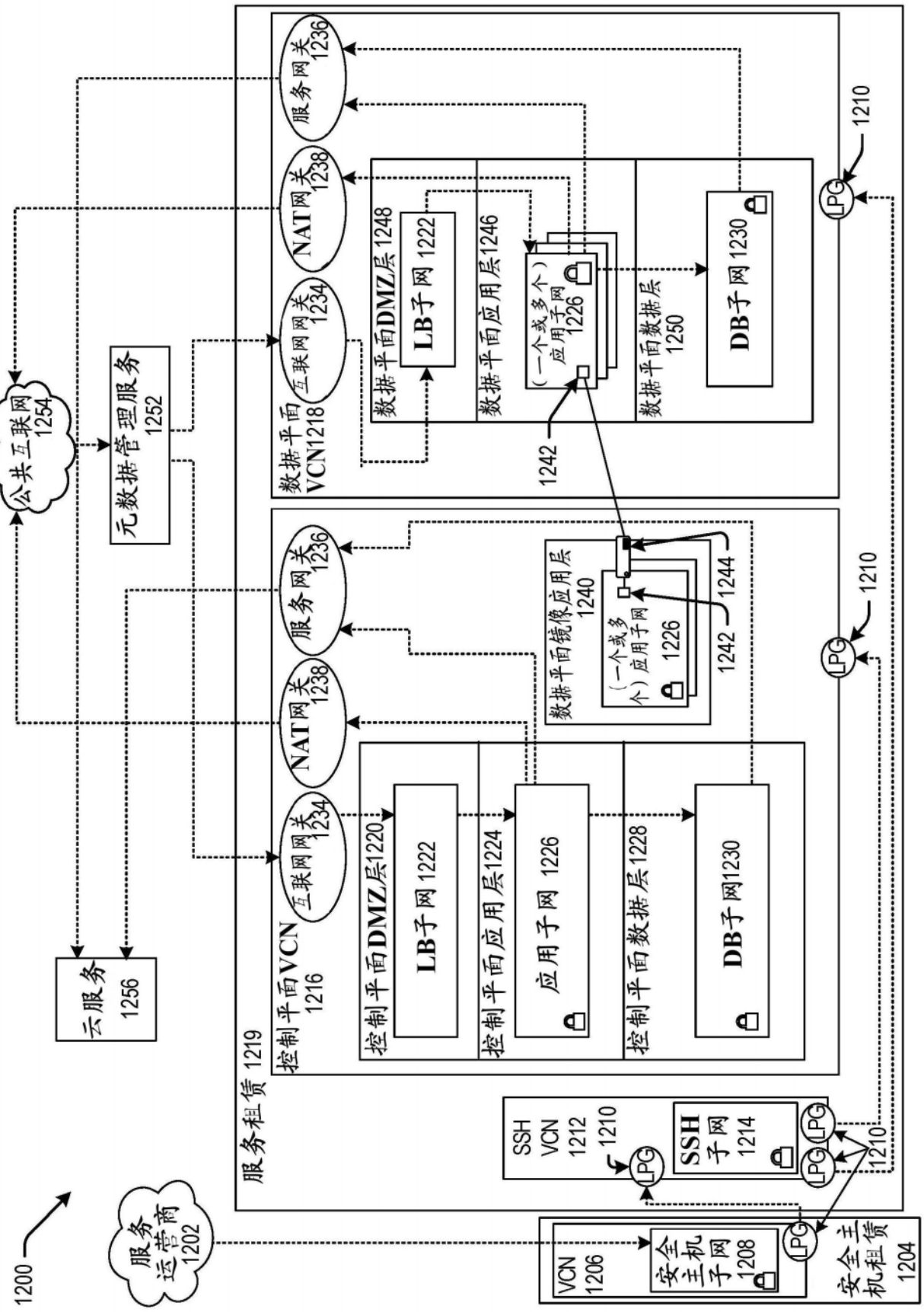

1200
服务运营商 1202
公共互联网 1254
元数据管理服务 1252
云服务 1256
服务租赁 1219
控制平面VCN 1216
互联网网关 1234
NAT网关 1238
服务网关 1236
控制平面DMZ层1220
LB子网 1222
控制平面应用层1224
应用子网 1226
控制平面数据层1228
DB子网1230
数据平面镜像应用层 1240
（一个或多个）应用子网 1226
1242
1244
1210
LPG
SSH VCN 1212
SSH 子网 1214
VCN 1206
安全主机子网 1208
安全主机租赁 1204
数据平面VCN1218
数据平面DMZ层1248
数据平面应用层1246
数据平面数据层 1250
（一个或多个）应用子网 1226
DB子网 1230

图12

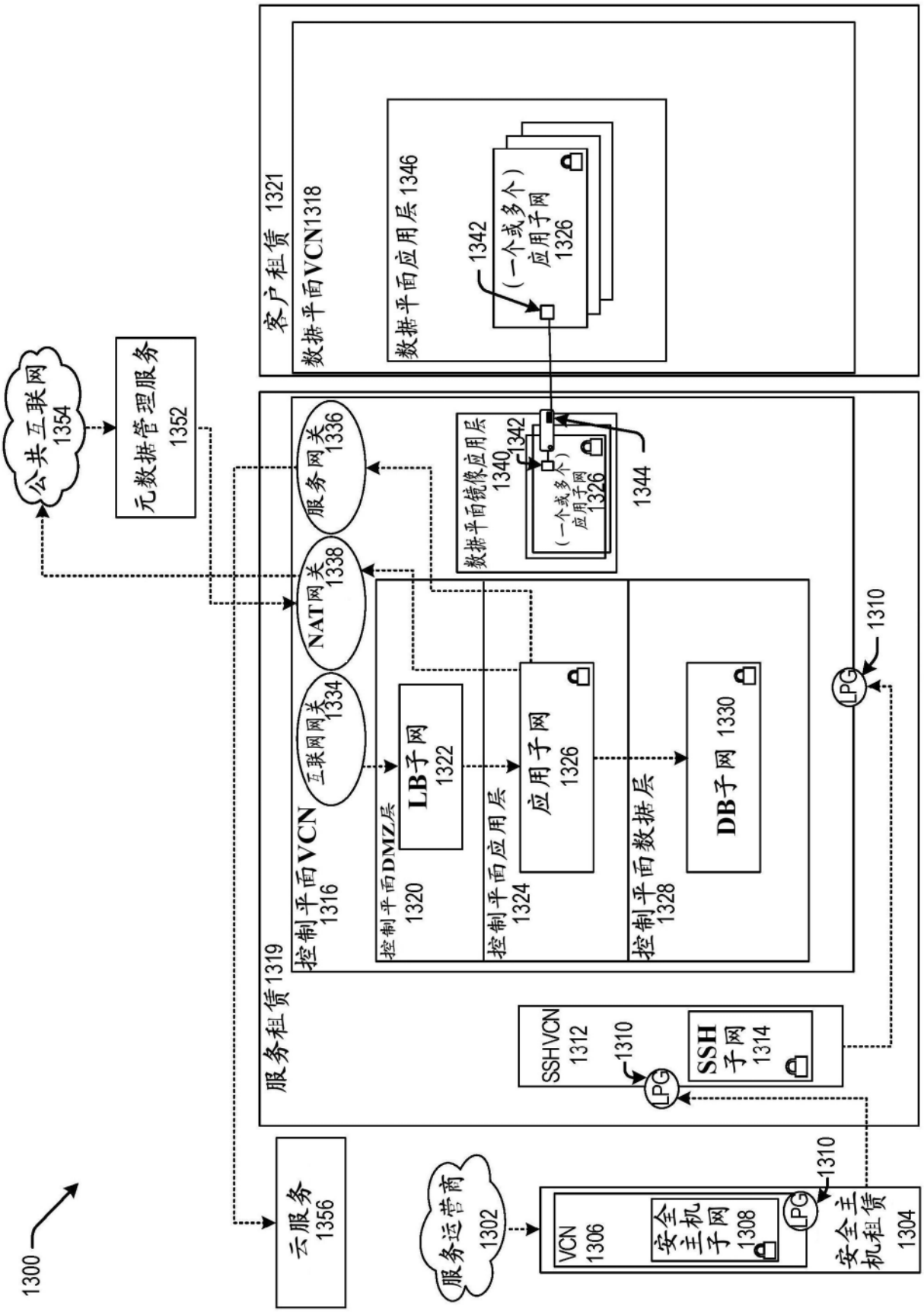

1300
公共互联网 1354
元数据管理服务 1352
客户租赁 1321
数据平面VCN1318
数据平面应用层 1346
1342
(一个或多个) 应用子网 1326
数据平面镜像应用层 1340
1342
(一个或多个) 应用子网 1326
1344
服务网关 1336
NAT网关 1338
互联网网关 1334
服务租赁 1319
控制平面VCN 1316
控制平面DMZ层 1320
LB子网 1322
控制平面应用层 1324
应用子网 1326
控制平面数据层 1328
DB子网 1330
1310
LPG
SSHVCN 1312
1310
SSH 子网 1314
云服务 1356
服务运营商 1302
VCN 1306
安全主机子网 1308
1310
安全主机租赁 1304

图13

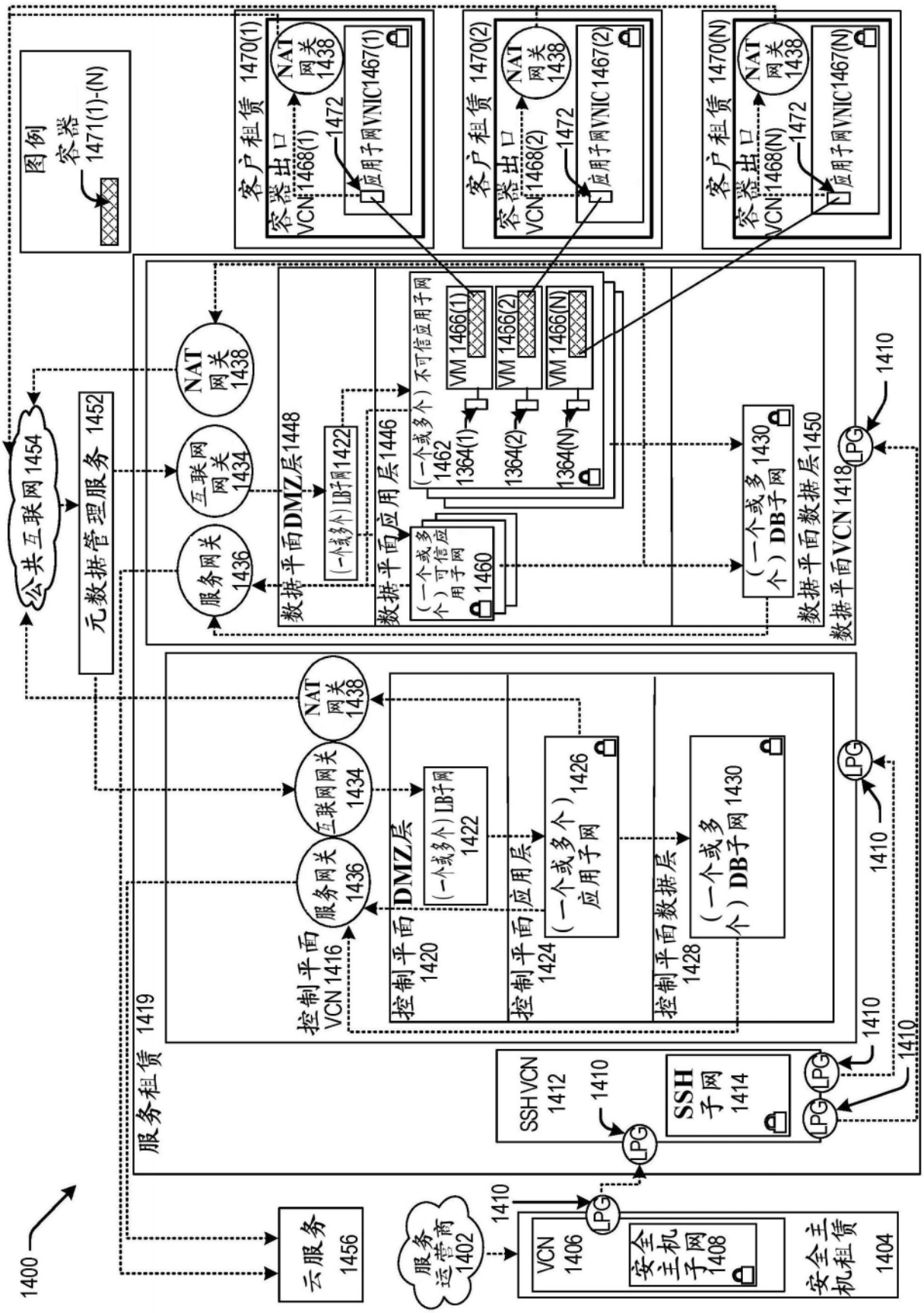

图例
容器 1471(1)-(N)
客户租赁 1470(1)
容器出口 VCN 1468(1)
NAT 网关 1438
1472
应用子网VNIC 1467(1)
客户租赁 1470(2)
容器出口 VCN 1468(2)
NAT 网关 1438
1472
应用子网VNIC 1467(2)
客户租赁 1470(N)
容器出口 VCN 1468(N)
NAT 网关 1438
1472
应用子网VNIC 1467(N)
公共互联网 1454
元数据管理服务 1452
NAT 网关 1438
互联网网关 1434
服务网关 1436
数据平面DMZ层 1448
(一个或多个) LB子网 1422
数据平面应用层 1446
(一个或多个) 不可信应用子网 1462
VM 1466(1)
VM 1466(2)
VM 1466(N)
1364(1)
1364(2)
1364(N)
(一个或多个) 可信应用子网 1460
(一个或多个) DB子网 1430
数据平面数据层 1450
数据平面VCN 1418
1410
服务租赁 1419
控制平面 VCN 1416
NAT 网关 1438
互联网网关 1434
服务网关 1436
控制平面 DMZ层 1420
(一个或多个) LB子网 1422
控制平面应用层 1424
(一个或多个) 应用子网 1426
控制平面数据层 1428
(一个或多个) DB子网 1430
1410
SSH VCN 1412
SSH 子网 1414
LPG
1400
云服务 1456
服务运营商 1402
VCN 1406
安全主机子网 1408
安全主机租赁 1404

图14

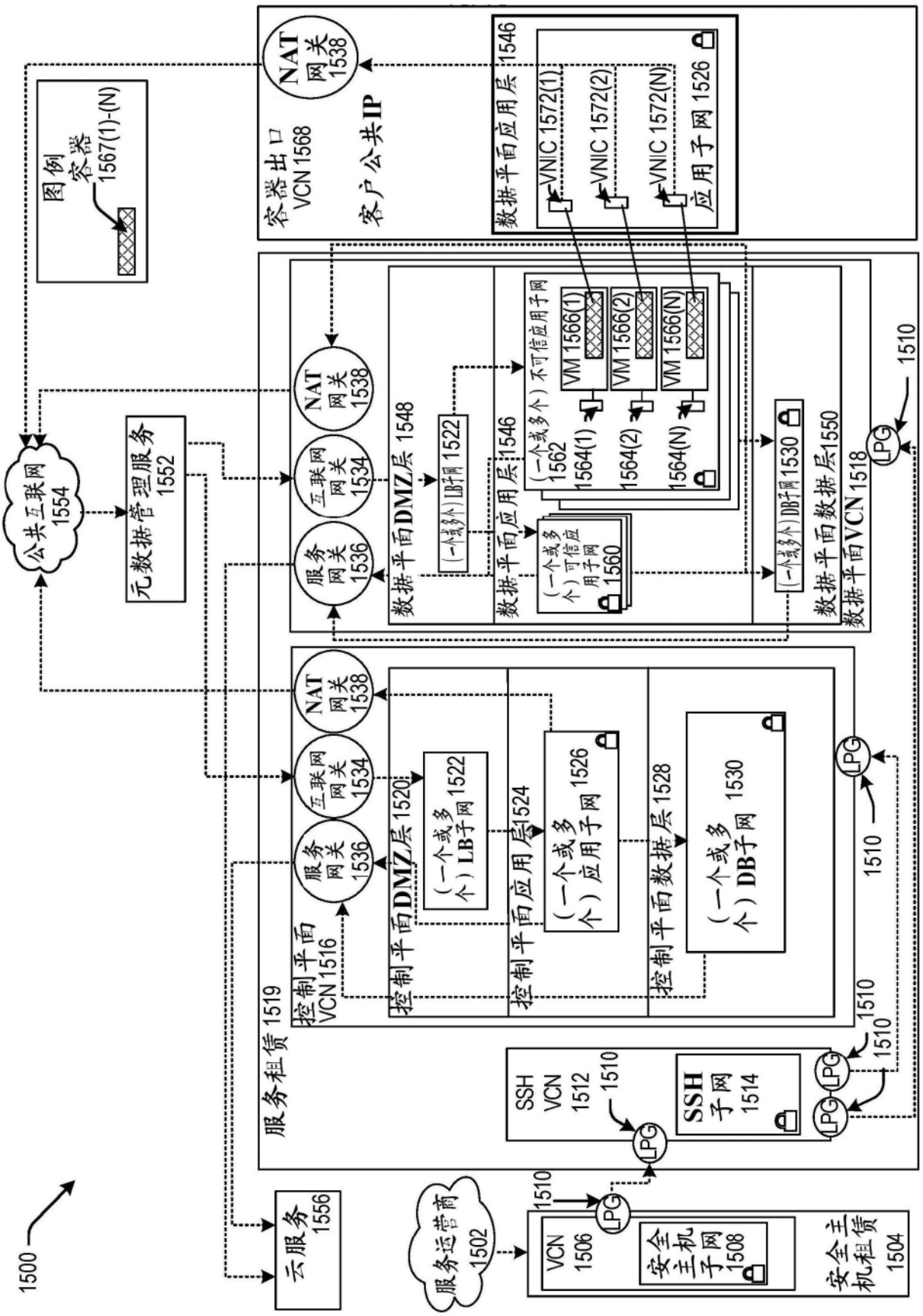

1500
图例
容器
1567(1)-(N)
容器出口
VCN 1568
客户公共IP
NAT
网关
1538
数据平面应用层 1546
VNIC 1572(1)
VNIC 1572(2)
VNIC 1572(N)
应用子网 1526
公共互联网
1554
元数据管理服务
1552
互联网
网关
1534
服务
网关
1536
数据平面DMZ层 1548
(一个或多个) LB子网 1522
(一个或多个) 不可信应用子网
1562
VM 1566(1)
VM 1566(2)
VM 1566(N)
1564(1)
1564(2)
1564(N)
(一个或多个) 可信应用子网
1560
(一个或多个) DB子网 1530
数据平面VCN 1518
数据平面数据层1550
LPG
1510
服务租赁 1519
控制平面
VCN1516
控制平面DMZ层 1520
(一个或多个) LB子网 1522
控制平面应用层1524
(一个或多个) 应用子网 1526
控制平面数据层 1528
(一个或多个) DB子网 1530
SSH
VCN
1512
SSH
子网
1514
云服务
1556
服务运营商
1502
VCN
1506
安全主机子网
1508
安全主机租赁
1504

图15

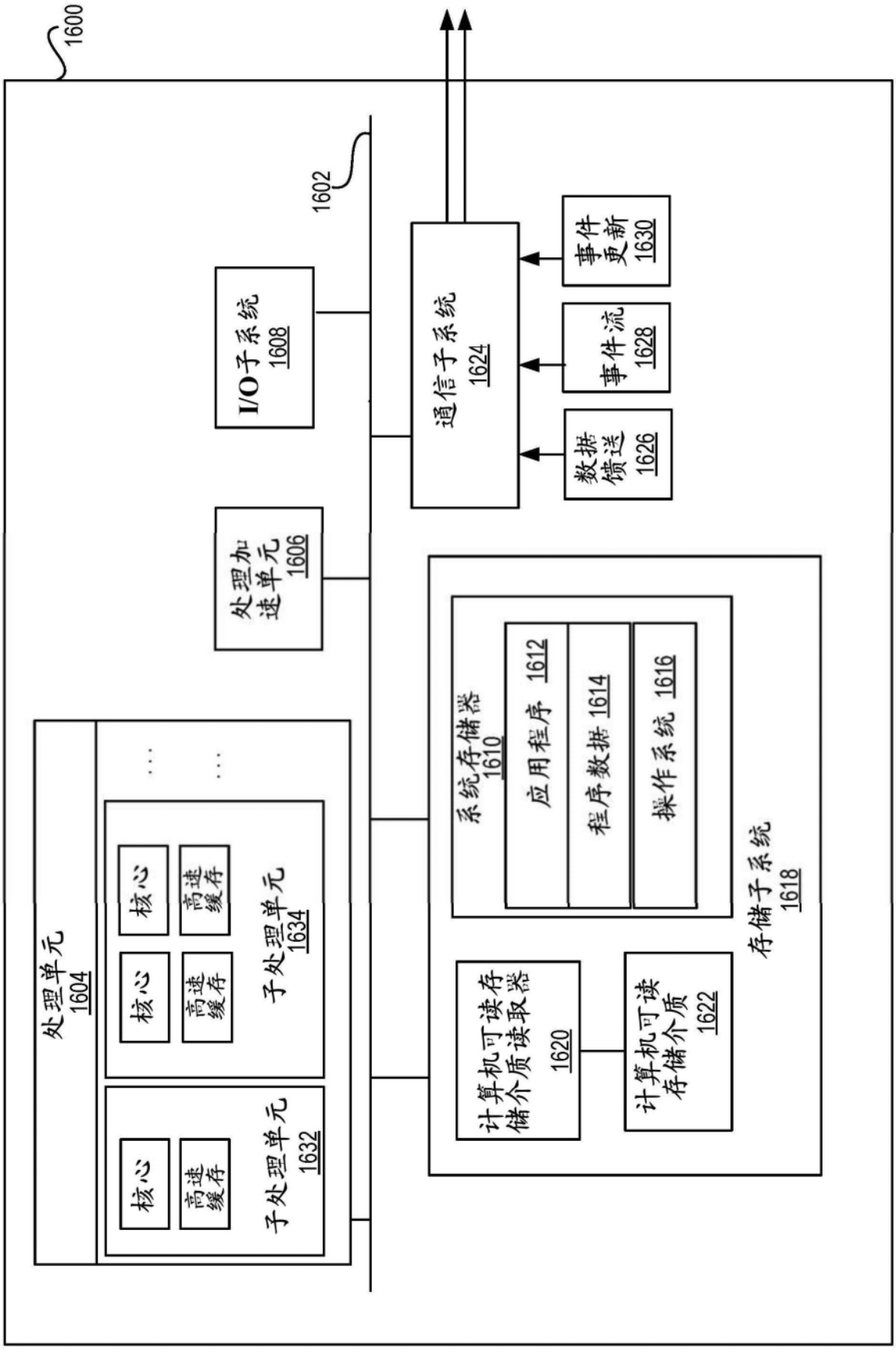

1600
1602
处理单元
1604
核心
高速缓存
子处理单元
1632
核心
高速缓存
核心
高速缓存
子处理单元
1634
I/O子系统
1608
处理加速单元
1606
通信子系统
1624
事件更新
1630
事件流
1628
数据馈送
1626
系统存储器
1610
应用程序 1612
程序数据 1614
操作系统 1616
计算机可读存储介质读取器
1620
计算机可读存储介质
1622
存储子系统
1618

图16