



(12) 发明专利

(10) 授权公告号 CN 113190382 B

(45) 授权公告日 2024. 12. 13

(21) 申请号 202110493813.9

(22) 申请日 2015.07.29

(65) 同一申请的已公布的文献号
申请公布号 CN 113190382 A

(43) 申请公布日 2021.07.30

(30) 优先权数据
14/454,415 2014.08.07 US(62) 分案原申请数据
201580052237.7 2015.07.29(73) 专利权人 微软技术许可有限责任公司
地址 美国华盛顿州(72) 发明人 D·克鲁泽 V·彼得
L·S·科波鲁 D·戴恩 M·乔治(74) 专利代理机构 北京世辉律师事务所 16093
专利代理师 李峥宇(51) Int.Cl.
G06F 11/14 (2006.01)(56) 对比文件
CN 103636165 A, 2014.03.12

审查员 高沛沛

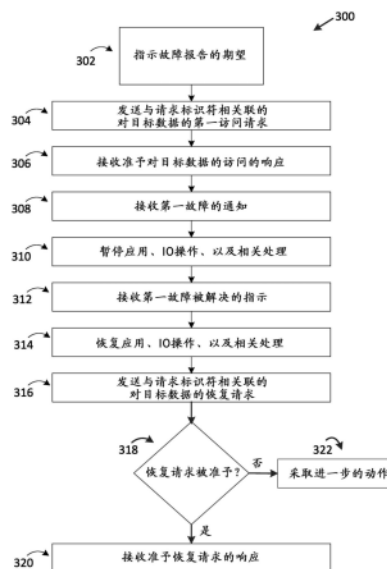
权利要求书3页 说明书14页 附图8页

(54) 发明名称

存储故障之后的安全数据访问

(57) 摘要

涉及存储故障之后的安全数据访问。请求标识符与各种系统和方法一起被采用,以便使得应用能够在故障期间暂停执行并且随后一旦故障被解决则恢复执行以及与存储系统的通信。驻留在节点上的应用可以尝试访问存储系统上的目标数据。应用发送访问目标数据的请求。该请求可以与识别应用的请求标识符相关联。当存储系统经历故障时,应用接收故障通知。作为该通知的结果,应用暂停执行。当节点重建通向存储系统的连接时,应用使用相同的请求标识符来恢复执行。如果恢复请求被准予,则应用被保证目标数据在故障期间未被修改。



1. 一种用于安全数据访问的系统,包括:
至少一个处理器;以及
存储器,存储指令,所述指令在由所述至少一个处理器执行时执行一组操作,所述操作包括:
由应用生成用于启用与存储系统的通信会话期间的故障的故障报告的指示;
向所述存储系统发送针对目标数据的第一访问请求,所述第一访问请求与第一请求标识符相关联,所述第一请求标识符用于识别所述应用并与所述目标数据相关联;
由所述系统存储所述第一请求标识符;
接收准予对所述目标数据的访问的第一访问响应;
接收第一故障的通知;
基于所述第一故障的所述通知,暂停所述应用;
恢复所述应用;
向所述存储系统发送针对所述目标数据的恢复请求,所述恢复请求与所述第一请求标识符相关联;以及
接收准予对所述目标数据的访问的恢复响应。
2. 根据权利要求1所述的系统,其中所述故障包括以下至少一项:
从所述存储系统断开;
与所述存储系统相关联的文件系统的卸除;以及
所述存储系统的重启。
3. 根据权利要求1所述的系统,其中用于启用故障报告的所述指示被提供给输入/输出I/O管理器,并且其中由于所述通信会话期间的故障而被延迟的访问请求被报告给所述应用。
4. 根据权利要求3所述的系统,其中所述第一故障通知从所述I/O管理器被接收,并且其中恢复所述应用发生在恢复通知从所述I/O管理器被接收时。
5. 根据权利要求1所述的系统,其中所述第一故障的通知包括对所述第一故障和所述第一故障的预期持续期中的至少一项的描述的附加故障信息。
6. 根据权利要求1所述的系统,其中暂停所述应用包括暂停所述应用的I/O操作和相关处理。
7. 根据权利要求1所述的系统,其中接收所述第一故障的通知包括接收第二访问请求尚未被完成的通知。
8. 根据权利要求1所述的系统,其中所述恢复请求包括指示所述恢复请求是用于恢复所述通信会话的请求的标记。
9. 根据权利要求1所述的系统,还包括完成在所述第一故障之前未被完成的被缓存的写入请求。
10. 一种计算机实现的方法,所述方法包括:
由输入/输出I/O管理器启用与存储系统的通信会话期间的故障的故障报告;
向所述存储系统发送针对目标数据的第一访问请求,所述第一访问请求与第一请求标识符相关联,所述第一请求标识符用于识别应用并与所述目标数据相关联;
存储所述第一请求标识符;

接收准予对所述目标数据的访问的第一访问响应；
从所述I/O管理器接收第一故障的通知；
基于所述第一故障的所述通知，暂停所述应用；
恢复所述应用；

向所述存储系统发送针对所述目标数据的恢复请求，所述恢复请求与所述第一请求标识符相关联；以及

接收准予对所述目标数据的访问的恢复响应。

11. 根据权利要求10所述的计算机实现的方法，其中所述故障包括以下至少一项：

从所述存储系统断开；
与所述存储系统相关联的文件系统的卸除；以及
所述存储系统的重启。

12. 根据权利要求10所述的计算机实现的方法，其中所述第一故障的通知包括对所述第一故障和所述第一故障的预期持续期中的至少一项的描述的附加故障信息。

13. 根据权利要求10所述的计算机实现的方法，其中恢复所述应用发生在恢复通知从所述I/O管理器被接收时。

14. 根据权利要求10所述的计算机实现的方法，其中接收所述第一故障的通知包括接收第二访问请求尚未被完成的请求。

15. 根据权利要求10所述的计算机实现的方法，其中所述恢复请求包括指示所述恢复请求是用于恢复所述通信会话的请求的标记。

16. 一种计算机实现的方法，所述方法包括：

由应用生成用于启用与存储系统的通信会话期间的故障的故障报告的指示；

向所述存储系统发送针对目标数据的第一访问请求，所述第一访问请求与第一请求标识符相关联，所述第一请求标识符用于识别所述应用并与所述目标数据相关联；

存储所述第一请求标识符；

接收准予对所述目标数据的访问的第一访问响应；

接收第一故障的通知；

基于所述第一故障的所述通知，暂停所述应用；

接收所述第一故障的解决的通知；

基于所述解决的通知，恢复所述应用；

向所述存储系统发送针对所述目标数据的恢复请求，所述恢复请求与所述第一请求标识符相关联；以及

接收准予对所述目标数据的访问的恢复响应。

17. 根据权利要求16所述的计算机实现的方法，其中所述故障包括以下至少一项：

从所述存储系统断开；
与所述存储系统相关联的文件系统的卸除；以及
所述存储系统的重启。

18. 根据权利要求16所述的计算机实现的方法，其中用于启用故障报告的所述指示被提供给输入/输出I/O管理器，并且其中由于所述通信会话期间的故障而被延迟的访问请求被报告给所述应用。

19. 根据权利要求16所述的计算机实现的方法, 其中所述第一故障通知从I/O管理器被接收, 并且其中所述解决的通知从所述I/O管理器被接收。

20. 根据权利要求16所述的计算机实现的方法, 其中所述第一故障的通知包括对所述第一故障和所述第一故障的预期持续期中的至少一项的描述的附加故障信息。

存储故障之后的安全数据访问

[0001] 本申请是国际申请号为PCT/US2015/042555、国际申请日为2015年07月29日、于2017年03月27日进入中国国家阶段、国家申请号为201580052237.7、发明名称为“存储故障之后的安全数据访问”的中国发明专利申请的分案申请。

技术领域

[0002] 本申请总体而言涉及计算机领域,具体而言涉及在存储故障之后进行安全数据访问的系统和方法。

背景技术

[0003] 集群化环境(例如,工作负载跨多个机器分布的环境)常用于向客户机提供故障转移和高信息可用性。集群化环境允许客户机经由作为环境一部分的一个或多个节点来访问资源。集群化环境可以充当客户机、服务器、或这两者。在客户机集群服务器中,应用可以驻留在组成集群的节点中的任何节点上。应用可以发出对存储系统所存储的目标数据的访问请求。如果在访问请求期间发生错误,则应用可能经历影响执行的延迟。

[0004] 关于这些和其他考虑而做出示例。此外,虽然讨论了相对具体的问题,但是应当理解,这些示例不应当限于解决背景中所识别的具体问题。

发明内容

[0005] 这一概述被提供以简化形式介绍概念的选择,这些概念下面在详细描述章节中进一步被描述。这一概述不意图识别所要求保护的主题的关键特征或必要特征,也不意图用作确定所要求保护的主题的范围的帮助。

[0006] 系统和方法在本文中被公开,它们使得应用能够在故障期间暂停执行并且随后一旦故障被解决则恢复执行以及与存储系统的通信。根据一些方面,驻留在客户机集群中的节点上的应用向服务器集群中的节点发送访问请求以访问目标数据。在示例中,请求标识符被用于识别请求资源的应用和/或相关工作负载。请求标识符可以与请求一起被提供。当客户机访问目标数据时,请求标识符被存储在持久存储装置中并且与所请求的目标数据相关联。

[0007] 在应用完成它对目标数据的操作之前,故障可能发生,其使得应用丢失与存储系统的通信。例如,存储系统中的服务器可能重启,相关文件系统可能被移除,通信链路可能出故障,等等。在这样的情况下,由应用对存储系统所存储的目标数据而发出的访问请求将经历延迟或完全出故障。然而,即使应用对目标数据的锁定在故障期间到期,所存储的标识符也可能未从持久存储装置删除。在故障期间,应用可以接收它已经丢失与存储系统的通信的通知并且可以暂停执行。在重建与存储系统的通信时,应用可以接收解决通知并且尝试恢复执行。进一步地,应用可以向存储系统发送恢复请求以重新得到对目标数据的访问。恢复请求可以包括或参考请求标识符。请求标识符准许管理请求的存储系统确定恢复请求属于之前已经访问资源的相同应用。这样做允许服务器准予应用的访问目标数据的恢复请

求。在目标数据在恢复请求之前被修改的情况下,存储系统(例如,通过从持久存储装置删除所存储的标识符)将请求标识符从目标数据解除关联,以使得应用可能无法恢复它与存储系统的先前通信。作为结果,应用确信如果它能够使用恢复请求和原始请求标识符重建通信,则目标数据自从通信故障以来尚未被修改。否则,应用被警告目标数据已经改变,并且应用可以做出是否采取其他动作的决定。

[0008] 示例可以被实施为计算机过程、计算系统,或者被实施为制品,诸如计算机程序产品或计算机可读介质。计算机程序产品可以是计算机存储介质,其由计算机系统可读取并且编码指令的计算机程序以用于执行计算机过程。

附图说明

[0009] 参考以下附图来描述非限制性且非穷尽性示例。

[0010] 图1图示了可以用于实施本文所描述的示例的系统。

[0011] 图2图示了可以用于实施本文所描述的示例的系统。

[0012] 图3是与存储系统通信的应用的操作流程。

[0013] 图4是与应用通信的存储系统的操作流程。

[0014] 图5是图示了本发明的方面可以利用其而被实践的计算机设备的示例的框图。

[0015] 图6A和图6B是本发明的方面可以利用其而被实践的移动计算设备的简化框图。

[0016] 图7是本发明的方面可以被实践在其中的分布式计算系统的简化框图。

具体实施方式

[0017] 下面参考附图更完全地描述各种方面,附图形成其一部分,并且示出了具体的示例性方面。然而,示例可以按许多不同形式来实施并且不应当被解释为限于本文阐述的示例。因此,示例可以采取硬件实施、或整体地软件实施、或组合软件和硬件方面的实施的形式。以下详细描述因此不在限制性意义上被采取。

[0018] 本公开的示例涉及向应用提供可靠地访问存储系统上的目标数据并处置通信故障的机制。在示例中,存储系统可以是本地设备、网络附接存储设备、分布式文件服务器、或计算环境中的任何其他类型的存储系统。目标数据可以驻留在独立服务器上,或者它可以驻留在集群化环境中。在本文所公开的示例中,集群化环境可以包括一个或多个节点(例如,客户机和/或服务器设备)。

[0019] 在一种示例中,应用可以请求对目标数据的访问。目标数据可以存储在本地(例如,在客户机节点上)、在远程设备(例如,客户机集群化环境中的不同节点或远程服务器)中、或在与客户机集群化环境不同的集群化环境(例如,包含多个节点的环境)中。例如,在集群化环境可以是客户机或服务器集群的示例中,本领域的技术人员然而将明白,在任何其他类型的环境(诸如但不限于虚拟网络)中可以采用本文所公开的系统和方法。

[0020] 数据可以在多个请求器之间共享。如本文所使用的,请求器可以包括任何应用、工作负载、线程、或请求对目标数据的访问的其他过程或实体。虽然本文所描述的示例可以关于充当请求器的“应用”或“客户机”或“过程”或“工作负载”来描述,但是本公开不如此被限制。当请求器访问目标数据时,目标数据可以被锁定,由此禁止其他请求器访问它直到锁定被释放。锁定目标数据可以用来防范冲突—也就是,在进行访问的请求器已经执行它的操

作之前防范由另一请求器对目标数据的修改。通常,当锁定被释放时,存储系统不再保留目标数据与请求器之间的任何关联。

[0021] 在一些实例中,当影响请求器与存储系统的通信的故障发生时,请求器可能无法在对目标数据的锁定被释放之前重新连接到存储系统。作为结果,请求器变得与目标数据解除关联。例如,如果存储系统使用由请求器在访问请求中提出的或参考的工作负载实例标识符来维持锁定,则当锁定被释放时,工作负载实例标识符通常可以从目标数据被删除和/或解除关联。进一步地,因为锁定被释放,所以在建立新连接时,请求器不能假定目标数据自从它上次被请求器访问以来未被修改。因此,不是恢复执行以及与存储系统的通信,而是请求器必须建立新通信会话并且假定目标数据处于未知状态。

[0022] 进一步地,当请求器与存储系统通信时,请求器可能期望不会接收到通信延迟或故障的指示,并且可能替代地假定它与存储系统(和/或托管请求器的客户机上的输入/输出(I/O)管理器)的契约保证访问请求将被重试直到它们成功。作为结果,通信故障将使得请求器继续等待来自存储系统的响应。这可能影响请求器的行为并且引起执行延迟。例如,请求器可能“被冻结”并且无法执行任何处理直到故障被解决。本文所公开的系统和方法提供了识别请求器正在尝试恢复与存储系统的通信的情形的机制、以及向请求器通知通信延迟或故障的机制,由此减轻对执行的影响并且使得请求器能够在故障之后恢复执行和通信。

[0023] 图1图示了可以用于实施一些示例的系统100。系统100包括客户机102和存储系统108。在所图示的示例中,客户机102包括请求器,诸如由物理或虚拟计算设备托管的应用104和输入/输出(I/O)管理器106。I/O管理器106中继并且管理应用104与存储系统108之间的通信。存储系统108存储由应用104访问的信息。虽然在图1中仅客户机102上的应用104被示出为与存储系统108通信,但是在其他示例中,可以存在从存储系统108访问信息的多于一个应用和多于一个客户机。在一些示例中,存储系统108可以包括在本地连接到客户机102的存储装置、通过网络可访问的远离客户机102的存储装置、或在集群化文件服务器之间共享的存储装置。

[0024] 根据一个示例,存储系统108用于向应用104提供一致的可用性。这通过利用客户机102和存储系统108上的组件来存储信息而被完成,如果存在通信故障、文件系统卸除、重启、或任何其他类型的故障,则该信息可以用于重建客户机102与存储系统108之间的通信。

[0025] 如下面更详细描述,所存储的信息和通信故障通知允许客户机102具有一致的文件访问,并且使得客户机102上运行的应用能够处置可能发生的通信故障。在示例中,客户机102可以使用服务器消息块(SMB)2.0/3.0协议与存储系统108通信。

[0026] 为了说明一个示例,应用104可以向I/O管理器106指示:与存储系统108的通信会话期间的通信故障应当被报告。在一些示例中,该指示可以指明用以等待请求成功的指定时间段,在该指定时间段之后,请求应当被报告为已经失败。在缺少这种指示的示例中,应用104可能依赖于I/O管理器106来向存储系统108重试访问请求。这可能导致应用104无限期地等待,而没有与延迟的性质或长度有关的信息。

[0027] 应用104可以向存储系统108发送对目标数据的访问请求。对目标数据的访问请求可以包括要对目标数据执行的多个文件操作。这些操作例如可以是打开以读取/写入数据、列举属性、租赁请求以允许数据在本地的缓存、或其他文件访问操作。在一些示例中,访问请求可以包括或参考请求标识符,诸如工作负载实例标识符。

[0028] 存储系统108接收对目标数据的访问请求。在一些示例中,存储系统108可以确定目标数据是否被锁定并且因此与(例如,用于不同请求器的)不同请求标识符相关联,不同请求器诸如客户机102或不同客户机上运行的另一应用。如果目标数据被锁定,则存储系统108取决于所采用的协议可以拒绝文件访问请求,可以打破锁定,或者可以采取其他动作。

[0029] 在确定访问请求应当被准予时,存储系统108向客户机102发送准予对目标数据的访问的响应。请求标识符可以由存储系统108存储在持久存储装置中。当请求标识符由存储系统108存储时,它在本文可以被称为所存储的标识符。在一些示例中,目标数据可以在准予访问请求之前或之后被锁定。在一些示例中,锁定可以被配置为在请求中的所有操作完成时被释放(由此允许其他请求器具有对目标数据的访问)。进一步地,锁定可以在指明的时间段之后到期、或者在指明的时间段与请求中的所有操作被完成的时间中的较短者时到期。在更其他的示例中,锁定可以提供对目标数据的客户机独有访问,直到客户机释放锁定或直到打破锁定的事件发生。例如,依照SMB协议,来自存储系统108的独有OpLock可以被提供给客户机102。客户机102可以假定它是具有对目标数据的访问的仅有过程,并且客户机102可以在向存储系统108送交对目标数据的改变之前缓存所有改变。如果另一客户机/过程尝试打开相同的目标数据,则存储系统108向客户机102发送消息(称为打破或撤销),该消息使之前给予客户机102的独有OpLock无效。客户机102然后清除对目标数据的所有改变并且将它们送交存储系统108。

[0030] 在某个时刻,客户机102可能经历与存储系统108的通信故障。该故障在一些示例中可能起因于断开、文件系统的卸除、或重启。作为通信故障的结果,I/O管理器106向应用104提供故障的通知。在示例中,这种通知仅在应用104“已选择参加”并且向I/O管理器106提供了如下指示时发生:应用104在与存储系统108的通信故障的情况下将被通知。在示例中,故障的通知可以包括描述故障的性质和/或预期长度的附加故障信息。应用104可以使用附加故障信息来确定采取什么动作,包括继续等待、放弃特定请求、重试请求、重启应用104,等等。除了其他选择之外,在接收到通知时,应用104还可以暂停执行。在其他示例中,应用104还可以(或替代地)暂停I/O操作或相关处理。在示例中,故障还可能使得存储系统108释放应用104对目标数据的锁定,或者锁定可能在故障期间到期。

[0031] 最后,故障可能被解决,并且客户机102与存储系统108之间的连接可以被重建。I/O管理器106确定故障是否已经被解决,并且如果是,则向应用104提供解决通知。在示例中,应用104还可以例如通过重试访问请求来确定故障是否已经被解决。在接收到解决通知之后,应用104可以恢复执行。在仅I/O操作被暂停的一些示例中,应用104还可以恢复I/O操作或相关处理。应用104发送恢复请求来恢复针对目标数据的与存储系统108的通信。在示例中,恢复请求可以由I/O管理器106或通过I/O管理器106来发送。恢复请求可以包括访问请求,该访问请求包括(或参考)在故障之前使用的请求标识符,并且可以包括向存储系统指示访问请求是恢复请求的标记。存储系统108接收恢复请求。存储系统108确定目标数据自从故障以来是否已经改变。目标数据是否已经改变的确定也可以部分地在接收到恢复请求之前被做出。

[0032] 如果存储系统108确定目标数据尚未改变,则存储系统108准予恢复请求。在示例中,存储系统108通过将恢复请求中的请求标识符与所存储的标识符进行比较来确定目标数据是否已经改变,该所存储的标识符来自客户机对目标数据的最为近期的访问请求。如

果另一过程或应用在故障的时间与恢复请求的时间之间改变了目标数据,则存储系统108将已经从客户机102的最为近期的访问请求删除了所存储的标识符。如此,如果所存储的标识符与包括在恢复请求中或以其他方式与恢复请求相关联的标识符相匹配,则存储系统准予恢复请求。准予恢复请求向应用有效地保证目标数据自从所存储的标识符针对其而初始被存储的访问请求以来尚未改变。因此,客户机102在继续它的之前操作时确信数据完整性。例如,如果客户机102已经缓存了在故障之前未被完成的对目标数据的写入,则这些写入现在可以被完成。如果存储系统108确定目标数据已经改变(例如,恢复请求中的请求标识符不与持久存储装置中所存储的标识符相匹配),则存储系统108拒绝恢复请求。如本文所使用的,“匹配”不需要包括精确复制。更确切地,标识符的匹配仅要求比较标识符的实体足够满意于两个标识符指代相同请求和/或请求器的给定水平的特异性(specificity)。

[0033] 在示例中,客户机102可能正在利用存储系统108上的一个或多个虚拟驱动器托管一个或多个虚拟机。当存储系统108出故障时,可靠性层(诸如SMB连续可用性)可能放弃尝试复原。这使得应用104接收到I/O故障,这些I/O故障导致有效地“冻结”具有在存储系统108上打开的目标数据的虚拟机。通常,虚拟机被要求经历重启过程来处置故障。然而,本公开允许虚拟机被暂停并且一旦故障被解决则被恢复。以这种方式,虚拟机内的操作系统重启可以被避免。然而,为了完成这一点,数据完整性必须被维持。

[0034] 当虚拟机被暂停时,存储系统108上针对该虚拟机的所有处置可能被关闭,但是虚拟机内的操作系统状态可以被保存。例如,虚拟硬驱动器上所安装的新技术文件系统(NTFS)做出如下假设:在虚拟机的恢复时,虚拟硬驱动器的所有内容都是相同的。如果虚拟硬驱动器的内容在虚拟机被暂停时改变,则虚拟硬驱动器可能被损坏。本公开使得有可能高效地确定在虚拟机被暂停之后恢复它是否为相同的。

[0035] 前述描述仅是图1中示出的示例可以如何操作的一个示例。如下面更详细描述,示例可以涉及不同的步骤或操作。这些可以使用任何适当的软件或硬件组件或模块来实施。

[0036] 图2图示了可以用于实施一些示例的系统200。系统200包括节点202和208。节点202包括由物理或虚拟计算设备托管的应用204和I/O管理器206。应用204可以是虚拟化应用。节点208包括其他实用程序(utility)210,并且还可以包括一个或多个应用和I/O管理器(未示出)。节点202和208可以是物理和/或虚拟计算设备的客户机集群的一部分。节点202和208使用例如SMB协议通过网络212与存储系统214通信。存储系统214存储由节点202和208上的一个或多个应用访问的信息。在示例中,节点202和208可以被考虑为如本文所描述的请求器。节点202和208建立与存储系统214的连接以从可缩放文件服务器216访问信息。虽然在图2中仅节点202和208被示出为与存储系统214通信,但是在其他示例中,可以存在从存储系统214访问信息的多于(或少于)两个客户机。

[0037] 如图2中示出的,存储系统214包括节点214A和214B,它们提供用于可缩放文件服务器216的高可用性和冗余这两者。在示例中,存储系统214提供由客户机202和208访问的可缩放文件服务器216。可缩放文件服务器216可以包括多个集群化服务器,它们可以协作以提供来自分布式文件系统218的文件信息。分布式文件系统218包括被存储在物理存储装置226上的文件信息。物理存储装置226的文件系统222被安装在节点214B上。在一些示例中,文件系统222的格式可以是新技术文件系统(NTFS)或弹性文件系统(ReFS)。节点214B充

当分布式文件系统218的协调器,并且中继从分布式文件系统218到文件系统222的文件操作,例如,读取、写入、以及元数据操作。在一些示例中,节点214A可以直接对物理存储装置226执行文件操作,但是节点214B可以具有独有写入访问,而使得写入请求由节点214A转发给节点214B,而不是直接被发送给物理存储装置226。另外,请求标识符过滤器220将访问请求与然后被存储在持久存储装置224中的标识符相关联。例如,持久存储装置224可以包括由附接到NTFS文件系统的恢复键过滤器来利用的恢复键过滤器数据库。虽然图2中示出了两个节点,但是在其他示例中,存储系统214可以包括多于两个节点或少于两个节点。

[0038] 根据示例,节点214A和214B用于提供可缩放文件服务器216的一致可用性。这通过利用节点202、208、214A和214B上的组件来存储信息而被完成,如果存在故障(诸如,网络212的故障、分布式文件系统218或文件系统222的卸除、或节点214A和214B中的一个节点的故障或重启),则该信息可以用于重建节点202和208上的应用与节点214A和214B上的可缩放文件服务器216之间的通信。进一步地,应用204可以向IO管理器206指示:向应用204通知影响与存储系统214的通信的故障。如下面更详细描述,所存储的信息和通信故障通知允许节点202和208具有一致的文件访问,并且使得在节点202和208上运行的应用能够处置可能发生的通信故障。

[0039] 示例中的节点(例如214A和214B)每个为客户机提供对文件数据的访问,并且被配置为向客户机提供文件数据的一致可用性。为了说明一个示例,应用204可以向IO管理器206指示:与存储系统214的通信会话期间的故障应当被报告给应用204。在示例中,应用可能必须“选择参加”以接收故障的指示。在一些示例中,该指示可以指明用以等待请求成功的时间段,在该时间段之后,请求应当被报告为已经失败。应用204可以向节点214A发送对目标数据的访问请求。访问请求可以与请求标识符(诸如工作负载实例标识符)相关联或包括它。例如,当新工作负载被开始时,元组可以被创建,其包括:<工作负载放置Id、工作负载Id、工作负载实例Id>。对目标数据的访问请求可以包括要对目标数据执行的多个文件操作。这些操作例如可以是打开以读取/写入数据、列举属性、租赁请求以允许数据在本地的缓存、或其他文件访问操作。在示例中,所请求的目标数据可以包括虚拟硬驱动器图像,该虚拟硬驱动器图像包括来宾(guest)虚拟机。

[0040] 节点214A接收对目标数据的访问请求。在一些示例中,可缩放文件服务器216可以确定目标数据是否被锁定(例如,正在由不同的应用或过程访问)。如果目标数据被锁定,则可缩放文件服务器216可以拒绝文件访问请求,或者采取由所采用的特定协议规定的其他动作。

[0041] 在确定请求应当被准予时,节点214A向节点202发送准予对目标数据的访问的响应。请求标识符过滤器220把来自节点202的访问请求所包括的或与其相关联的请求标识符存储在持久存储装置224中。在一些示例中,目标数据然后可以被锁定。对目标数据的锁定(例如,由于故障)最后可能在来自应用204的所有操作已经完成之前被释放,但是所存储的标识符可以继续由持久存储装置224保留。进一步地,锁定可能在指明的时间段之后到期。

[0042] 在某个时刻,节点202可能经历与存储系统214的通信故障。故障可能起因于例如断开、文件系统的卸除、或重启。作为通信故障的结果,IO管理器206向应用204提供故障的通知。在接收到通知时,应用204可以暂停一个或多个来宾虚拟机的执行。除了暂停来宾虚拟机之外或者代替暂停来宾虚拟机,应用204也可以暂停IO操作或相关处理。

[0043] 最后,节点202与存储系统214之间的连接可能被复原。当IO管理器206确定故障已经被解决时,它向应用204提供解决通知。在示例中,应用204可以通过其他方法来确定故障已经被解决。在接收到解决通知之后,应用204可以恢复执行,包括恢复一个或多个来宾虚拟机的执行。在示例中,应用204还可以恢复IO操作或相关处理。应用204发送恢复请求以恢复与存储系统214的通信并且寻求对目标数据的访问。来自应用204的恢复请求包括恢复请求标识符或与其相关联,恢复请求标识符可以是在故障之前在第一访问请求中使用的相同标识符。另外,恢复请求可以包括向存储系统214指示该请求是恢复请求的标记。在这一示例中,节点214B接收恢复请求,并且节点214B确定目标数据自从故障以来是否已经改变。

[0044] 如果节点214B确定目标数据尚未改变,则节点214B准予恢复请求并且向应用204提供对目标数据的访问。在一些示例中并且如下面进一步解释的,目标数据自从应用204的之前访问请求以来尚未改变的确定可以通过确定恢复请求所包括的(或与其相关联的)标识符继续被存储在持久存储装置224中(并且尚未从其删除)来做出。当恢复请求被准予时,应用204和节点214B可以完成在故障之前被中断并且未完成的任何操作。当恢复请求被准予时,目标数据可以由节点214B代表应用204再次锁定。最后,对目标数据的锁定可以再次(例如,由于另外的故障、由应用204进行的操作的完成,等等)被释放,但是所存储的标识符可以继续由持久存储装置224保留。进一步地,锁定可能在指明的时间段之后到期。

[0045] 在一些示例中,节点214B可以确定目标数据在应用204的最为近期的访问请求与恢复请求之间已经改变。例如,在某个时刻,节点208上的其他实用程序210可能在节点202上的应用204重建它与存储系统214的连接之前变更目标数据。这可能例如发生在如下的情况下:应用204作为它的最为近期的访问请求的结果而具有对目标数据的锁定;然而,该锁定在故障发生之后、但是在应用204对目标数据的所有操作完成之前到期或被释放。锁定的终止可能允许其他实用程序210在恢复请求从应用204被接收之前变更目标数据。如果目标数据已经改变,则节点214B可以从持久存储装置224移除请求标识符。在示例中,在其他实用程序210变更目标数据时(甚至在恢复请求发生之前),来自应用204对目标数据的最为近期的访问请求的所存储的标识符可以从持久存储装置被删除。例如,存储系统214可以持续地或定期地监测对目标数据(存储系统214针对其而当前将所存储的标识符存储在持久存储装置224中)的改变,并且在改变被做出时删除这种所存储的标识符。在该情况下,目标数据是否已经改变的确定要求简单地检查恢复请求所包括的(或与其相关联的)标识符是否匹配于持久存储装置224中用于目标数据的所存储的标识符。在示例中,不构成对目标数据的变更或改变的操作可以包括读取、缩小(deflating)、压缩、去重、或快照操作,并且代表另一应用或其他实用程序210的这种操作将不会导致应用204的针对目标数据的所存储的标识符从持久存储装置224被删除。如果目标数据已经改变,则存储系统214可以拒绝恢复请求,并且应用可以采取进一步的动作。例如,如果虚拟机在故障期间已经被暂停并且故障之后的恢复请求被存储系统214拒绝,则虚拟机内的操作系统可以再次被暂停,在此时,重启虚拟机的确定可以被做出。

[0046] 如可以明白的,系统200的上述描述不意图限制本文所描述的示例。图2及其描述仅意图说明一些示例的实施方式。在其他示例中,不同类型的信息可以被存储在系统200中的不同组件上。因此,示例不限于图2中示出和描述的事物。

[0047] 图3和图4图示了根据示例的操作流程300和400。操作流程300和400可以在任何适

合的计算环境中被执行。例如,这些操作流程可以由诸如图1和图2中所图示的环境来执行。因此,操作流程300和400的描述可以参考图1和图2的组件中的至少一个组件。然而,将理解,图1和图2的实施方式是用于操作流程300和400的非限制性环境。

[0048] 再者,虽然操作流程300和400按特定顺序依次地被图示和描述,但是在其他示例中,这些操作可以按不同顺序、多次地、和/或并行地被执行。进一步地,一个或多个操作在一些示例中可以被省略或组合。另外,应当理解,序数(诸如“第一”)不意图暗示顺序或次序,除非另有指明,并且被用来在类似的元件之间进行区分。例如,“第一访问请求”不需要是初始访问请求,而是应当被解读为不同于“第二访问请求”或“恢复请求”。

[0049] 操作流程300图示了针对与存储系统进行通信的应用的操作。在示例中,图3中图示的流程300可以由客户机上正运行的应用来执行,例如,节点202上的应用204(图2)。

[0050] 流程300在操作302处开始,在操作302处,存储系统故障应当被报告给应用的指示被传达。该指示可以向IO管理器做出,诸如IO管理器206(图2)。在一些示例中,该指示可以指明用以等待访问请求成功的指定时间段,在该指定时间段之后,访问请求应当被报告为已经失败。

[0051] 流程从操作302转到操作304,在操作304处,对目标数据的第一访问请求被发送。第一访问请求与第一标识符相关联。在一些示例中,第一标识符可以被包括作为第一访问请求的一部分。对目标数据的访问请求可以包括要对目标数据执行的多个文件操作。这些操作可以是例如打开以读取/写入数据、列举属性、租赁请求以允许数据在本地的缓存、或其他文件访问操作。

[0052] 在操作306处,准予对所请求的目标数据的访问的响应被接收。移至操作308,第一故障的通知被接收。该故障可能起因于例如通信故障、文件系统卸除、或重启。流程300然后移至操作310,在操作310处,在接收到故障通知之后,暂停应用的确定被做出。在其他示例中,操作310还可以包括暂停IO操作和相关处理。例如,应用204在接收到故障通知之后可以暂停来宾虚拟机(图2)。在一些示例中,故障通知可以包括故障是否预期在特定时间段内被解决的指示。这种信息准许应用决定是暂停IO操作、重启、还是采取其他动作。

[0053] 在操作312处,第一故障已经被解决的指示被接收。移至操作314,应用被恢复。在其他示例中,操作314还可以包括恢复IO操作和相关处理。例如,应用204在接收到第一故障已经被解决的指示之后可以恢复来宾虚拟机(图2)。在示例中,操作314可以继续进行而无需首先接收到故障已经被解决的指示。

[0054] 流程300然后移至操作316,在操作316处,对目标数据的恢复请求被发送。恢复请求与第一标识符相关联并且包括该请求是恢复请求的指示(诸如标记)。在一些示例中,第一标识符可以被包括作为恢复请求的一部分。在操作316之后,流程300转到操作318,在操作318处,确定恢复请求是否被准予。如果恢复请求被准予,则准予对所请求的目标数据的访问的响应被接收320。接收到准予恢复请求的响应的应用可以将请求的准予考虑为是如下的确认:目标数据在应用对目标数据的最为近期的访问请求与恢复请求之间未改变。应用然后可以继续对目标数据的任何未完成的操作。如果恢复请求被拒绝,则进一步的动作被采取322。例如,接收到对恢复请求的拒绝的应用可以选择重新开始。在示例中,应用还可以通过用户接口向用户提供故障的指示并且提供用户采取进一步动作的选项。

[0055] 操作流程400图示了针对与应用进行通信的存储系统的步骤。在示例中,流程400

可以由存储系统来执行,诸如一个或多个文件服务器,例如,节点214A和/或节点214B(图2)。

[0056] 流程400在操作402处开始,在操作402处,对目标数据的第一访问请求被接收。第一访问请求与请求标识符相关联。在一些示例中,请求标识符可以被包括作为第一访问请求的一部分。对目标数据的访问请求可以包括要对目标数据执行的多个文件操作。这些操作可以是例如打开以读取/写入数据、列举属性、租赁请求以允许数据在本地的缓存、或其他文件访问操作。在操作402之后,流程400转到决定404,在决定404处,关于目标数据是否被锁定而做出决定。

[0057] 如果在决定404处确定目标数据被锁定,则流程转到操作406,在操作406处,第一访问请求被拒绝。在其他示例中,操作406可以包括打破锁定,在这种情况下,流程将会进行到操作408。

[0058] 如果在决定404处确定目标数据不被锁定,则流程继续到操作408,在操作408处,锁定针对请求器而被设置在目标数据上。锁定可以在指明的时间段之后到期,或者可以根据特定协议而被终止。移至操作410,访问所请求的目标数据被准予。流程然后继续到操作412,在操作412处,请求标识符被存储在持久存储装置中。

[0059] 流程转到操作414,在操作414中,故障发生。该故障可能起因于例如连接中断、与存储系统相关联的文件系统的卸载、或存储系统的重启,等等。在示例中,流程400包括操作416,在操作416中,在操作408处创建的锁定被释放并且请求标识符被保留在持久存储装置中。例如,请求器在故障之前可能尚未完成对目标数据的所有操作;然而,锁定由于时间段的到期、由于故障它本身、或由于其他因素而被释放。流程然后继续到决定418,在决定418处,关于目标数据自从它在操作410处上次被访问以来是否被修改而做出决定。

[0060] 如果在决定418处确定目标数据被修改,则流程转到操作420,在操作420处,所存储的请求标识符从持久存储装置被移除。例如,如果目标数据在锁定在操作416中被释放之后由另一过程改变,则将目标数据与请求器相关联的所存储的标识符可以从持久存储装置被删除。

[0061] 流程400然后(从决定操作418或操作420)继续到操作422,在操作422处,恢复请求被接收。恢复请求与恢复请求标识符相关联,其可以与在操作402中接收的请求标识符相同。在一些示例中,恢复请求标识符可以被包括作为恢复请求的一部分。进一步地,在示例中,决定418可以发生在接收到恢复请求之后。在操作422之后,流程400转到决定424,在决定424处,请求标识符是否匹配于持久存储装置中所存储的标识符的确定被做出。

[0062] 如果在决定424处确定请求标识符不匹配,则流程转到操作426,在操作426处,恢复请求被拒绝。

[0063] 如果在决定424处确定请求标识符匹配,则在一些示例中,流程继续到操作428,在操作428处,锁定针对请求器而被设置在目标数据上。锁定可以在指明的时间段之后到期,或者可以根据特定协议而被终止。移至操作430,访问所请求的目标数据被准予。在其他示例中,流程400包括操作432,在操作432中,在操作428处创建的锁定最后被释放并且请求标识符被保留在持久存储装置中。在操作430(或在其他示例中是操作432)之后,操作流程结束。

[0064] 图5-图7和相关联的描述提供了本发明的示例可以被实践在其中的各种操作环境

的讨论。然而,关于图5-图7被说明和讨论的设备和系统用于示例和说明的目的,并且不是对可以用于实践本文所述的本发明的示例的大量计算设备配置的限制。

[0065] 图5是图示了本公开的示例可以利用其而被实践的计算设备502(例如,节点202、208、214A和214B)的物理组件的框图。下面描述的计算设备组件可以适合于上面描述的计算设备。在基本配置中,计算设备502可以包括至少一个处理单元504和系统存储器506。取决于计算设备的配置和类型,系统存储器506可以包括但不限于易失性存储装置(例如,随机访问存储器)、非易失性存储装置(例如,只读存储器)、闪存存储器、或这些存储器的任何组合。系统存储器506可以包括操作系统507以及一个或多个程序模块508,一个或多个程序模块508适合于运行软件应用520,诸如应用204、I/O管理器206、以及其他实用程序210。操作系统507例如可以适合于控制计算设备502的操作。再者,本发明的示例可以连同图形库、其他操作系统、或任何其他应用程序而被实践,并且不限于任何特定应用或系统。这种基本配置在图5中由虚线522内的那些组件图示。计算设备502可以具有另外的特征或功能。例如,计算设备502还可以包括另外的(可移除和/或不可移除的)数据存储设备,诸如,例如磁盘、光盘或磁带。这种另外的存储装置在图5中由可移除存储设备509和不可移除存储设备510图示。

[0066] 如上面所陈述的,多个程序模块和数据文件可以被存储在系统存储器506中。当在处理单元504上执行时,程序模块508(例如,应用204、I/O管理器206、以及其他实用程序210)可以执行处理,包括但不限于图3和图4中图示的操作流程300和400的阶段中的一个或多个阶段。可以根据本发明的示例来使用的其他程序模块可以包括电子邮件和联系人应用、文字处理应用、电子制表应用、数据库应用、幻灯片应用、绘图或计算机辅助应用程序,等等。

[0067] 再者,本发明的示例可以被实践在包括分立电子元件的电路中、包含逻辑门的封装或集成电子芯片中、利用微处理器的电路中、或包含电子元件或微处理器的单个芯片上。例如,本发明的示例可以经由片上系统(SOC)来实践,其中图5中所图示的组件中的每个或许多组件可以被集成到单个集成电路上。这种SOC设备可以包括一个或多个处理单元、图形单元、通信单元、系统虚拟化单元、以及各种应用功能,它们中的全部被集成(或“烧录”)到芯片基板上作为单个集成电路。当经由SOC操作时,本文所描述的功能可以经由与计算设备502的其他组件一起被集成在单个集成电路(芯片)上的专用逻辑来操作。本公开的示例还可以使用能够执行逻辑操作(诸如,例如,“与”、“或”、以及“非”)的其他技术而被实践,包括但不限于机械、光学、流体、以及量子技术。另外,本发明的示例可以被实践在通用计算机内或者任何其他电路或系统中。

[0068] 计算设备502还可以具有一个或多个输入设备512,诸如键盘、鼠标、笔、声音输入设备、触摸输入设备等。(多个)输出设备514(诸如显示器、扬声器、打印机等)也可以被包括。前述设备是示例并且其他设备可以被使用。计算设备504可以包括允许与其他计算设备518通信的一个或多个通信连接516。适合的通信连接516的示例包括,但不限于,RF发射器、接收器、和/或收发器电路系统;通用串行总线(USB)、并行和/或串行端口。

[0069] 如本文所使用的术语“计算机可读介质”可以包括计算机存储介质。计算机存储介质可以包括易失性和非易失性的、可移除和不可移除的介质,其按照用于存储信息(诸如计算机可读指令、数据结构、或程序模块)的任何方法或技术而被实施。系统存储器506、可移除存储设备509、以及不可移除存储设备510全部是计算机存储介质示例(即,存储器存储装

置)。计算机存储介质可以包括RAM、ROM、电可擦除只读存储器(EEPROM)、闪存存储器或其他存储器技术、CD-ROM、数字多功能盘(DVD)或其他光存储装置、磁带盒、磁带、磁盘存储装置或其他磁存储设备、或者可以用于存储信息并且可以由计算设备502访问的任何其他制品。任何这种计算机存储介质可以是计算设备502的一部分。计算机存储介质不包括载波或其他传播的或调制的数据信号。

[0070] 通信介质可以由计算机可读指令、数据结构、程序模块、或调制数据信号(诸如载波或其他传输机制)中的其他数据来具体化,并且包括任何信息递送介质。术语“调制数据信号”可以描述如下的信号:其使一个或多个特性以对信号中的信息进行编码的这种方式而被设置或改变。通过示例而不是限制的方式,通信介质可以包括有线介质(诸如有线网络或直接有线连接)和无线介质(诸如声学、射频(RF)、红外、以及其他无线介质)。

[0071] 图6A和图6B图示了本发明的示例可以利用其而被实践的移动计算设备600,例如,移动电话、智能电话、平板个人计算机、膝上型计算机,等等。例如,移动计算设备600可以用于实施客户机节点202和208。参考图6A,图示了用于实施示例的移动计算设备600的一个示例。在基本配置中,移动计算设备600是具有输入元件和输出元件这两者的手持式计算机。移动计算设备600通常包括显示器605以及一个或多个输入按钮610,它们允许用户将信息输入到移动计算设备600中。移动计算设备600的显示器605也可以用作输入设备(例如,触摸屏显示器)。如果被包括,则可选的侧部输入元件615允许进一步的用户输入。侧部输入元件615可以是旋转开关、按钮、或任何其他类型的手动输入元件。在替换性示例中,移动计算设备600可以包含更多或更少的输入元件。例如,显示器605在一些示例中可以不是触摸屏。在又另一替换性示例中,移动计算设备600是便携式电话系统,诸如蜂窝电话。移动计算设备600还可以包括可选的小键盘635。可选的小键盘635可以是物理小键盘或在触摸屏显示器上生成的“软”小键盘。在各种示例中,输出元件包括用于示出图形用户界面(GUI)的显示器605、视觉指示器620(例如,发光二极管)、和/或音频换能器625(例如,扬声器)。在一些示例中,移动计算设备600包含用于向用户提供触觉反馈的振动换能器。在又另一示例中,移动计算设备600包含用于向外部设备发送信号或从其接收信号的输入和/或输出端口,诸如音频输入(例如,麦克风插口)、音频输出(例如,耳机插口)、以及视频输出(例如,HDMI端口)。

[0072] 图6B是图示了移动计算设备的一个示例的架构的框图。也就是说,移动计算设备600可以包含系统(即,架构)602来实施一些示例。在一个示例中,系统602被实施为能够运行一个或多个应用(例如,浏览器、电子邮件、安排日程表、联系人管理器、消息收发客户机、游戏、以及媒体客户机/播放器)的“智能电话”。在一些示例中,系统602被集成为计算设备,诸如集成个人数字助理(PDA)和无线电话。

[0073] 一个或多个应用程序666可以被加载到存储器662中并且在操作系统664上或与其相关联地运行。应用程序的示例包括电话拨号器程序、电子邮件程序、个人信息管理(PIM)程序、文字处理程序、电子制表程序、互联网浏览器程序、消息收发程序,等等。系统602还包括存储器662内的非易失性存储区域668。非易失性存储区域668可以用于存储在系统602被断电的情况下不应当丢失的持久信息。应用程序666可以使用信息并将信息存储在非易失性存储区域668中,诸如电子邮件或由电子邮件应用使用的其他消息,等等。同步应用(未示出)也驻留在系统602上并且被编程为与驻留在主机上的对应同步应用进行交互,以保持非

易失性存储区域668中所存储的信息与主机处所存储的对应信息同步。如应当明白的,其他应用可以被加载到存储器662中并且在移动计算设备600上运行,包括本文所描述的应用204、I/O管理器206、以及其他实用程序210。

[0074] 系统602具有电源670,其可以被实施为一个或多个电池。电源670可以进一步包括外部电力源,诸如AC适配器或者对电池进行补充或再充电的被供电的对接支座(docking cradle)。

[0075] 系统602可以包括外围设备端口678,其执行促进系统602与一个或多个外围设备之间的连接的功能。去往和来自外围设备端口672的传输在操作系统664的控制下进行。换句话说,由外围设备端口678接收的通信可以经由操作系统664传播到应用程序666,并且反之亦然。

[0076] 系统602还可以包括无线电672,其执行发射和接收射频通信的功能。无线电672经由通信载波或服务供应商来促进系统602与“外部世界”之间的无线连接。去往和来自无线电672的传输在操作系统664的控制下进行。换句话说,由无线电672接收的通信可以经由操作系统664传播到应用程序666,并且反之亦然。

[0077] 视觉指示器620可以用来提供视觉通知,和/或音频接口674可以用于经由音频换能器625产生可听通知。在所图示的示例中,视觉指示器620是发光二极管(LED)并且音频换能器625是扬声器。这些设备可以直接耦合到电源670,从而在被激活时,即使处理器660和其他组件可能为了节省电池功率而关闭,它们也保持接通至由通知机制规定的持续期。LED可以被编程为无限期地保持接通,直到用户采取动作来指示设备的通电状况。音频接口674用于向用户提供可听信号并且从用户接收可听信号。例如,除了被耦合到音频换能器625之外,音频接口674还可以耦合到麦克风以接收可听输入,诸如以促进电话交谈。根据本发明的示例,如下面将描述的,麦克风还可以充当促进对通知的控制的音频传感器。系统602可以进一步包括视频接口676,其使能机载相机630的操作以记录静止图像、视频流,等等。

[0078] 实施系统602的移动计算设备600可以具有另外的特征或功能。例如,移动计算设备600还可以包括另外的(可移除和/或不可移除的)数据存储设备,诸如,磁盘、光盘或磁带。这种另外的存储装置在图6B中由非易失性存储区域668图示。

[0079] 由移动计算设备600生成或捕获并且经由系统602存储的数据/信息可以如上面所描述的在本地存储在移动计算设备600上,或者数据可以存储在可以由设备经由无线电672或经由移动计算设备600与分离的计算设备之间的有线连接来访问的任何数目的存储介质上,该分离的计算设备与移动计算设备600相关联,例如,分布式计算网络(诸如互联网)中的服务器计算机。如应当明白的,这种数据/信息可以经由移动计算设备600,经由无线电672或经由分布式计算网络来访问。类似地,这种数据/信息可以根据公知的数据/信息传送和存储手段(包括电子邮件和协同数据/信息共享系统)而在计算设备之间容易地被传送以用于存储和使用。

[0080] 图7图示了用于向一个或多个客户机设备提供应用的系统的架构的一个示例,该应用如上面所描述的可靠地访问存储系统上的目标数据并且处置通信故障。与应用204、I/O管理器206、其他实用程序210、以及存储系统214相关联地被访问、与之交互、或被编辑的目标数据可以被存储在不同的通信信道或其他存储类型中。例如,各种文档可以使用目录服务722、web门户724、邮箱服务726、即时消息收发存储库728、或社交联网站点730而被存储。

应用204、IO管理器206、其他实用程序210、以及存储系统214可以使用如本文所描述的用于使能数据利用的这些类型的系统等中的任何一种。服务器720可以提供存储系统214,用于由操作在通用计算设备502和(多个)移动设备600上的客户机节点202和208通过网络715使用。通过示例的方式,网络715可以包括互联网或任何其他类型的局域或广域网,并且客户机节点202和208可以被实施为具体化在个人计算机、平板计算设备中的计算设备502,和/或由移动计算设备600(例如,智能电话)来实施。客户机计算设备502或600的这些实施例中的任何实施例可以从存储库716获得内容。

[0081] 除了其他示例之外,本公开还提出了系统和方法,它们用于由应用指示会话期间的故障应当被报告给应用;向存储系统发送对目标数据的第一访问请求,第一访问请求与第一标识符相关联;接收准予对目标数据的访问的第一访问响应;接收第一故障的通知;基于第一故障的通知,暂停应用;恢复应用;向存储系统发送对目标数据的恢复请求,恢复请求与第一标识符相关联;以及接收准予对目标数据的访问的恢复响应。在示例中,故障包括下列至少之一:从存储系统断开、与存储系统相关联的文件系统的卸除、或存储系统的重启。在进一步的示例中,第一标识符被包括在第一访问请求中。在进一步的示例中,指示包括向输入/输出(I/O)管理器提供以下指示:由于会话期间的故障而被延迟的I/O请求如果在指定时间段内未完成则应当被报告给应用。另外,接收第一故障的通知可以包括接收第二访问请求尚未被完成的通知。进一步地,暂停应用可以包括暂停I/O操作和相关处理。另外,系统和方法可以包括完成在故障之前未完成的被缓存的写入请求。

[0082] 本文进一步的描述公开了系统和方法,它们用于接收对目标数据的第一访问请求,目标数据由存储系统所存储,并且第一访问请求与请求标识符相关联;将请求标识符存储在持久存储装置中作为所存储的第一标识符;准予对目标数据的第一访问请求;在故障之后,接收对目标数据的恢复请求,恢复请求与请求标识符相关联;通过将恢复请求中的请求标识符与持久存储装置中的一个或多个标识符进行比较,来确定目标数据自从故障以来是否已经改变;当目标数据自从故障以来尚未改变时,准予对目标数据的恢复请求;以及当目标数据自从故障以来已经改变时,拒绝恢复请求。在示例中,存储系统是服务器集群上的共享文件系统,并且确定目标数据自从故障以来是否已经改变包括在恢复请求被接收之前从持久存储装置删除所存储的第一标识符。进一步地,在示例中,故障包括以下至少之一:连接中断、与存储系统相关联的文件系统的卸除、或存储系统的重启。在另外的示例中,系统和方法包括锁定目标数据;以及在准予第一访问请求之后,释放对目标数据的锁定而不从持久存储装置删除第一标识符。另外,在示例中,对目标数据的锁定持续第一时间段,并且第一时间段可以为零。

[0083] 本公开还包括一种系统,该系统包括被配置为执行计算机可执行指令的至少一个处理器;存储计算机可执行指令的至少一个计算机可读存储介质,这些计算机可执行指令在由至少一个处理器执行时提供一种存储系统,该存储系统被配置为:接收对目标数据的第一访问请求,目标数据由存储系统所存储,并且第一访问请求包括请求标识符;将请求标识符存储在持久存储装置中作为所存储的第一标识符;准予对目标数据的第一访问请求;在故障之后,接收对目标数据的恢复请求,恢复请求包括请求标识符并且包括恢复请求标记;通过将恢复请求中的请求标识符与持久存储装置中的一个或多个标识符进行比较,来确定目标数据自从故障以来是否已经改变;当目标数据自从故障以来尚未改变时,准予对

目标数据的恢复请求;以及当目标数据自从故障以来已经改变时,拒绝恢复请求。在示例中,确定目标数据自从故障以来是否已经改变包括在恢复请求被接收之前从持久存储装置删除所存储的第一标识符。另外,故障可以包括以下至少之一:连接中断、与存储系统相关联的文件系统的卸除、或存储系统的重启。进一步地,存储系统可以是服务器集群上的共享文件系统。另外,对目标数据的锁定可以持续第一时间段,并且第一时间段可以为零。

[0084] 贯穿这一说明书已经对“一个示例”或者“一种示例”做出参考,意味着特定所描述的特征、结构或特性被包括在至少一个示例中。因此,这种短语的使用可以指代多于仅一个示例。再者,所描述的特征、结构或特性在一个或多个示例中可以按任何适合的方式被组合。

[0085] 然而,相关领域的技术人员可以认识到,示例可以没有具体细节中的一个或多个、或利用其他方法、资源、材料等而被实践。在其他实例中,公知的结构、资源或操作没有详细被示出或描述,仅为了遵守不使示例的方面模糊不清。

[0086] 尽管图示和描述了示例和应用,但是将理解,示例不限于上面描述的精确配置和资源。在本文所公开的方法和系统的布置、操作和细节中可以做出对本领域的技术人员明显的各种修改、改变和变化,而不偏离所要求保护的示例的范围。

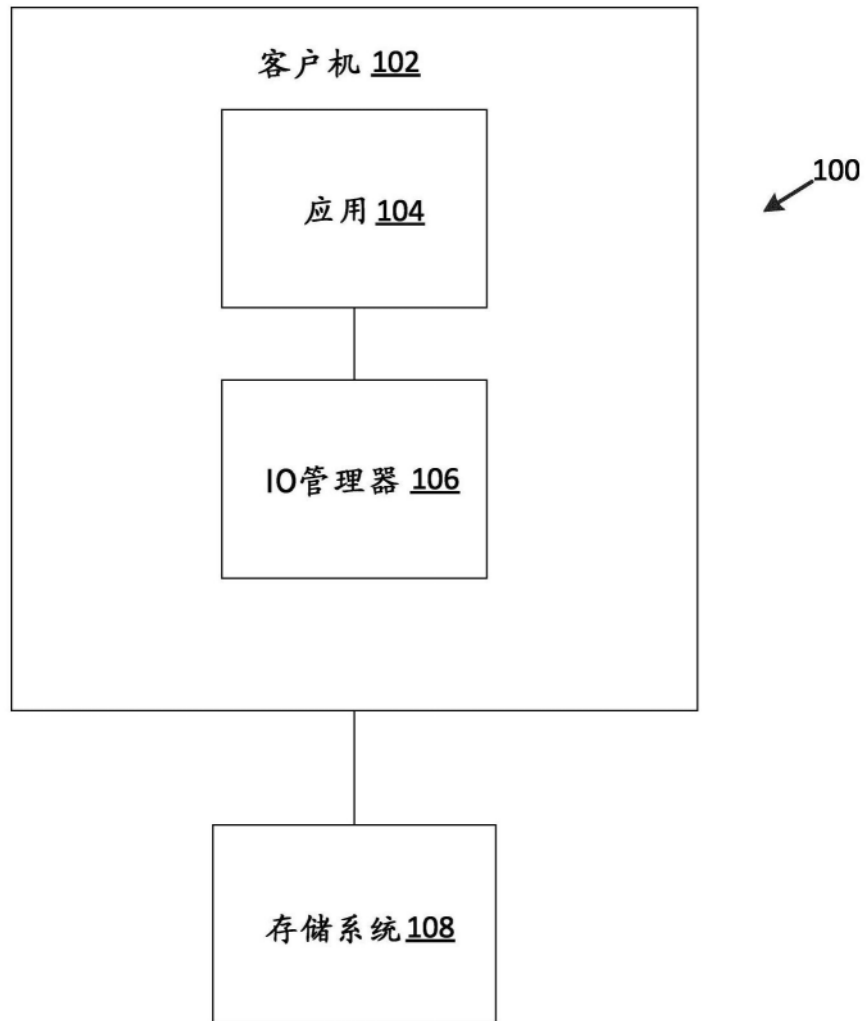


图1

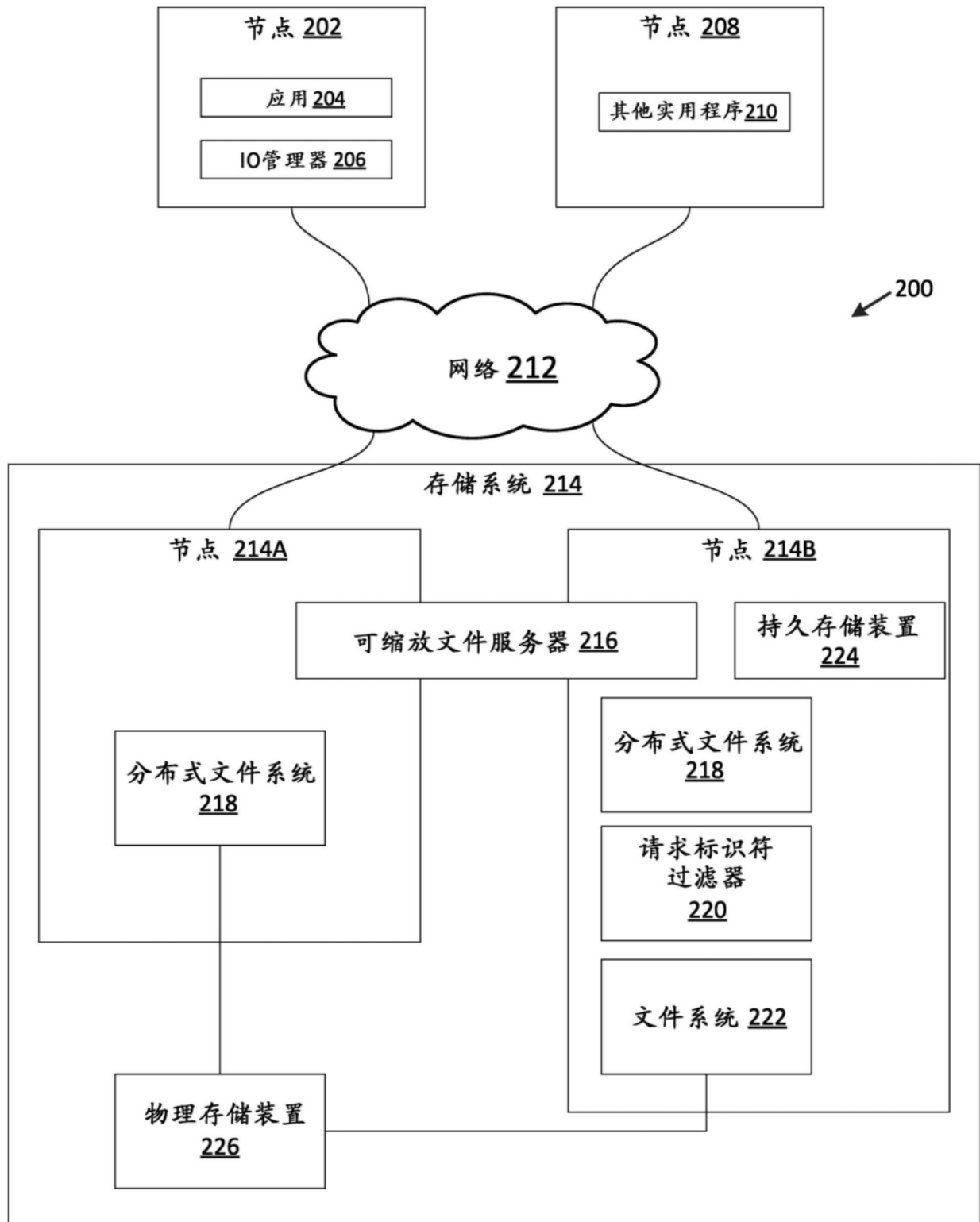


图2

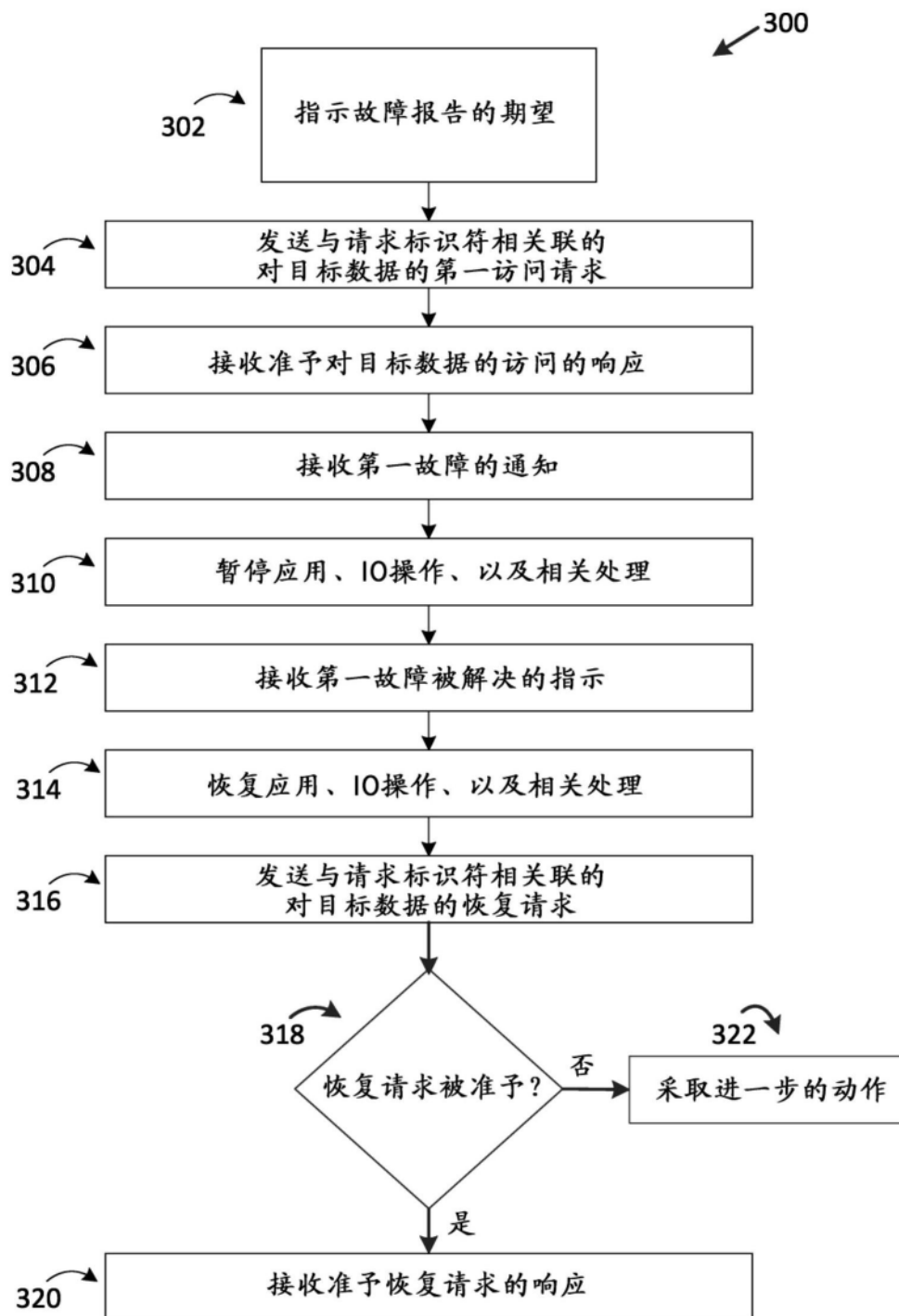


图3

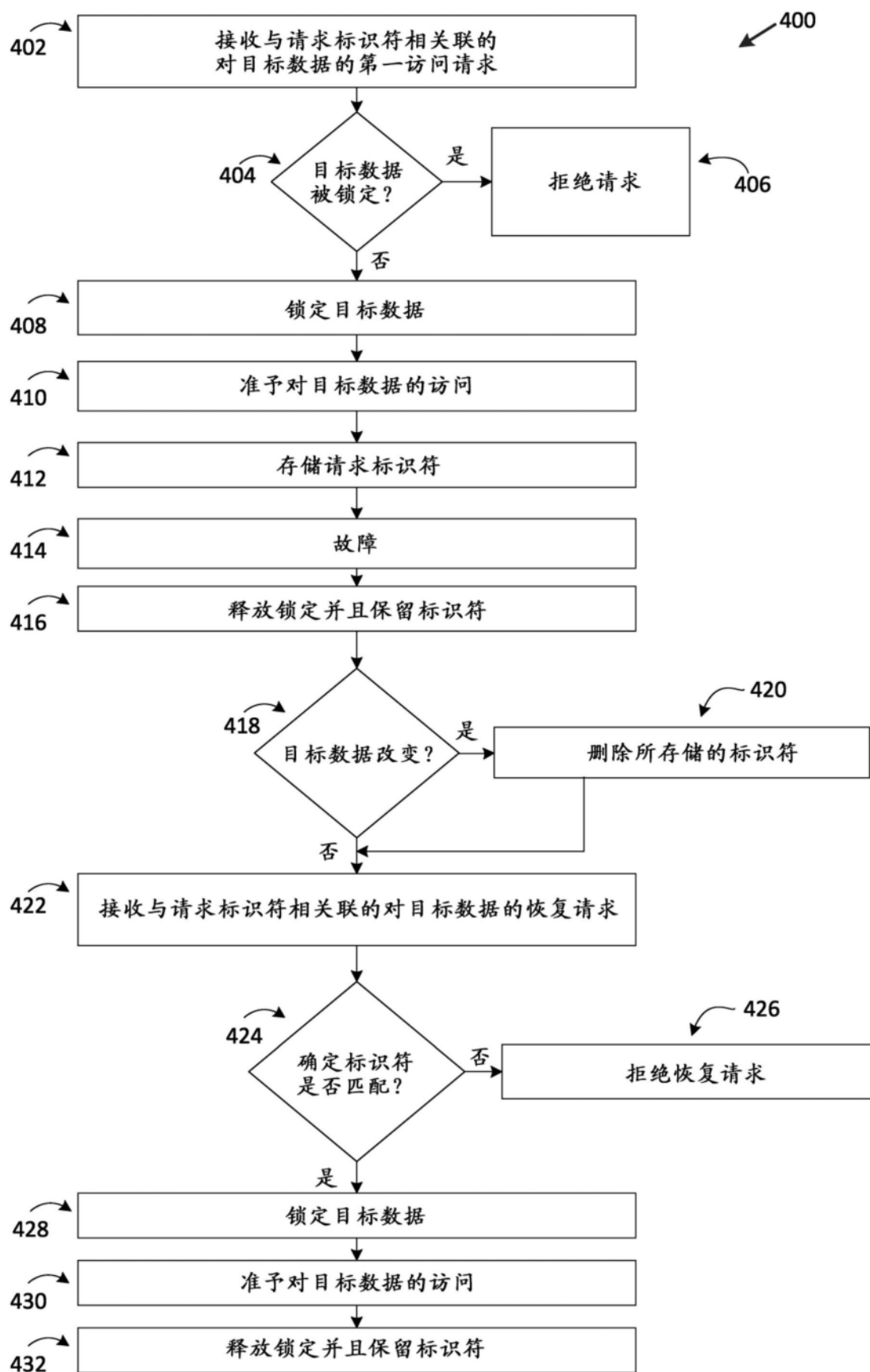


图4

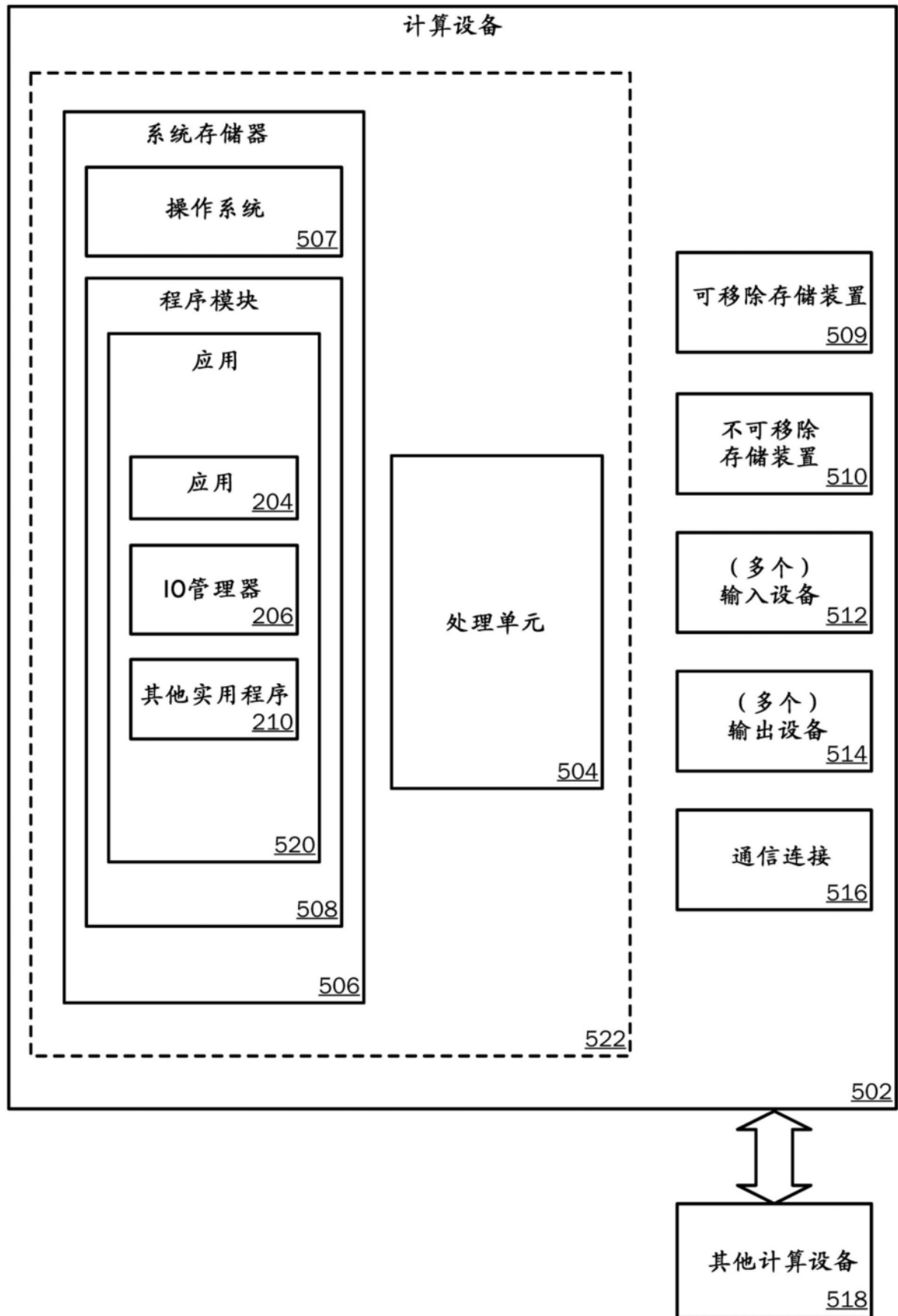


图5

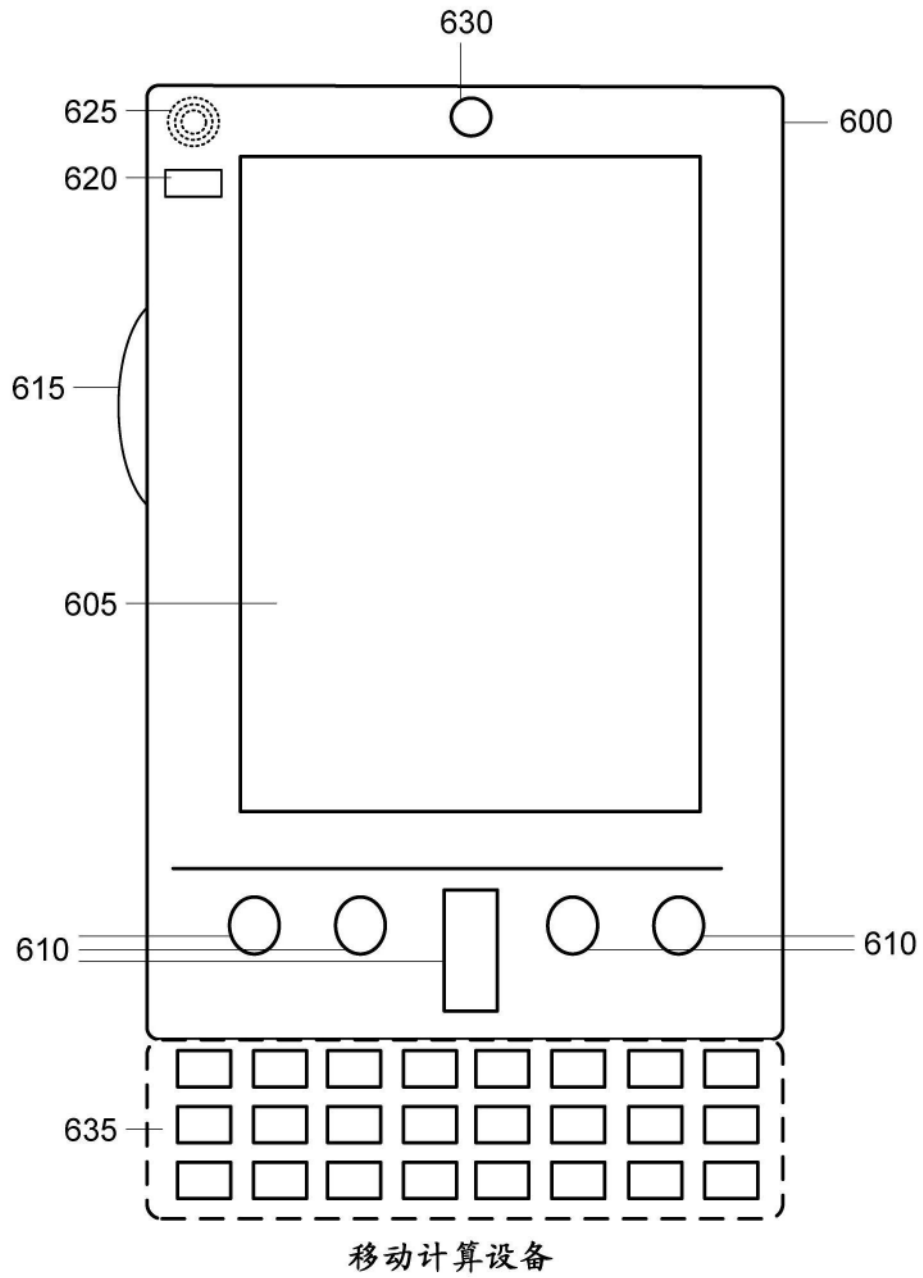


图6A

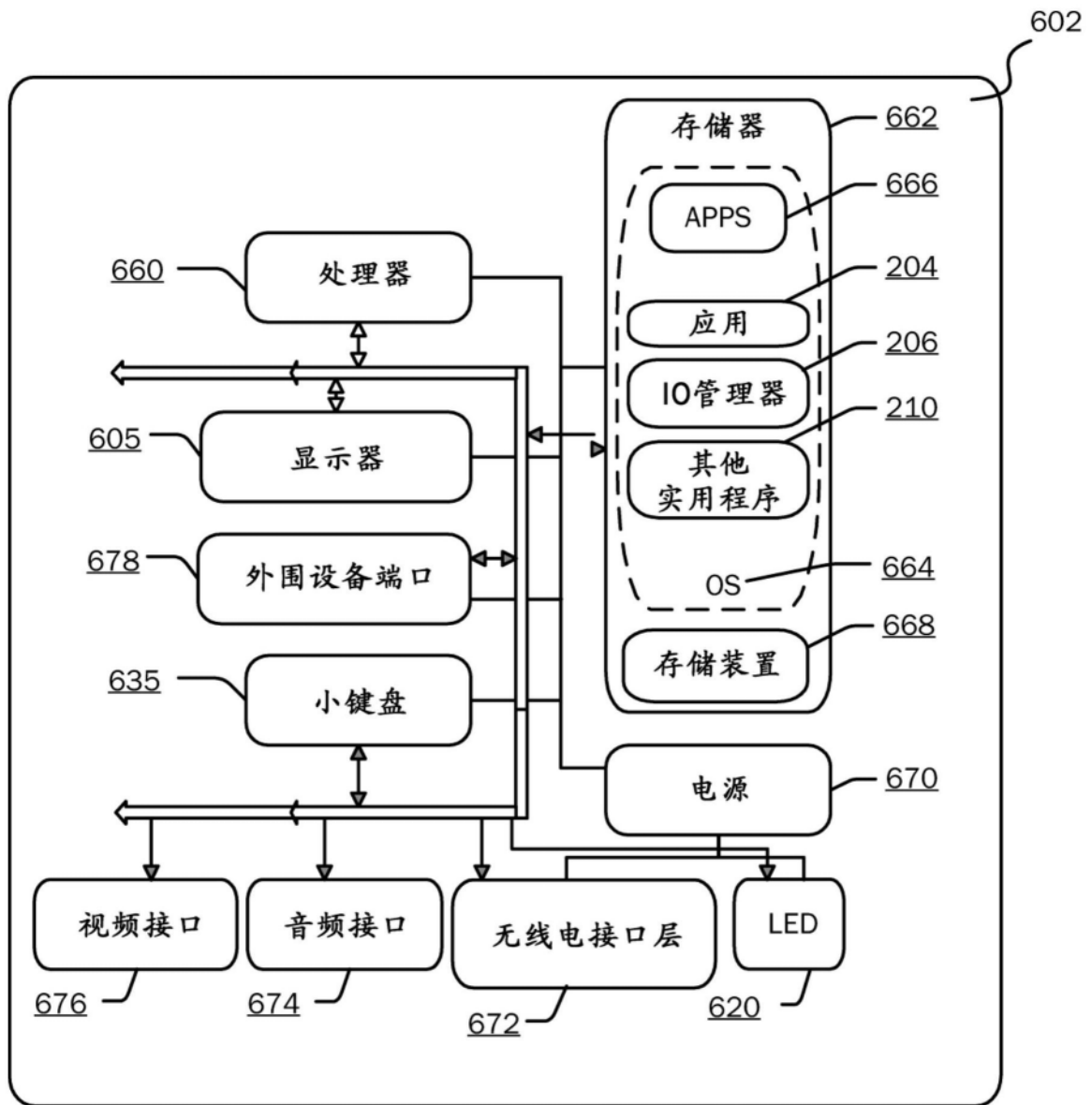


图6B

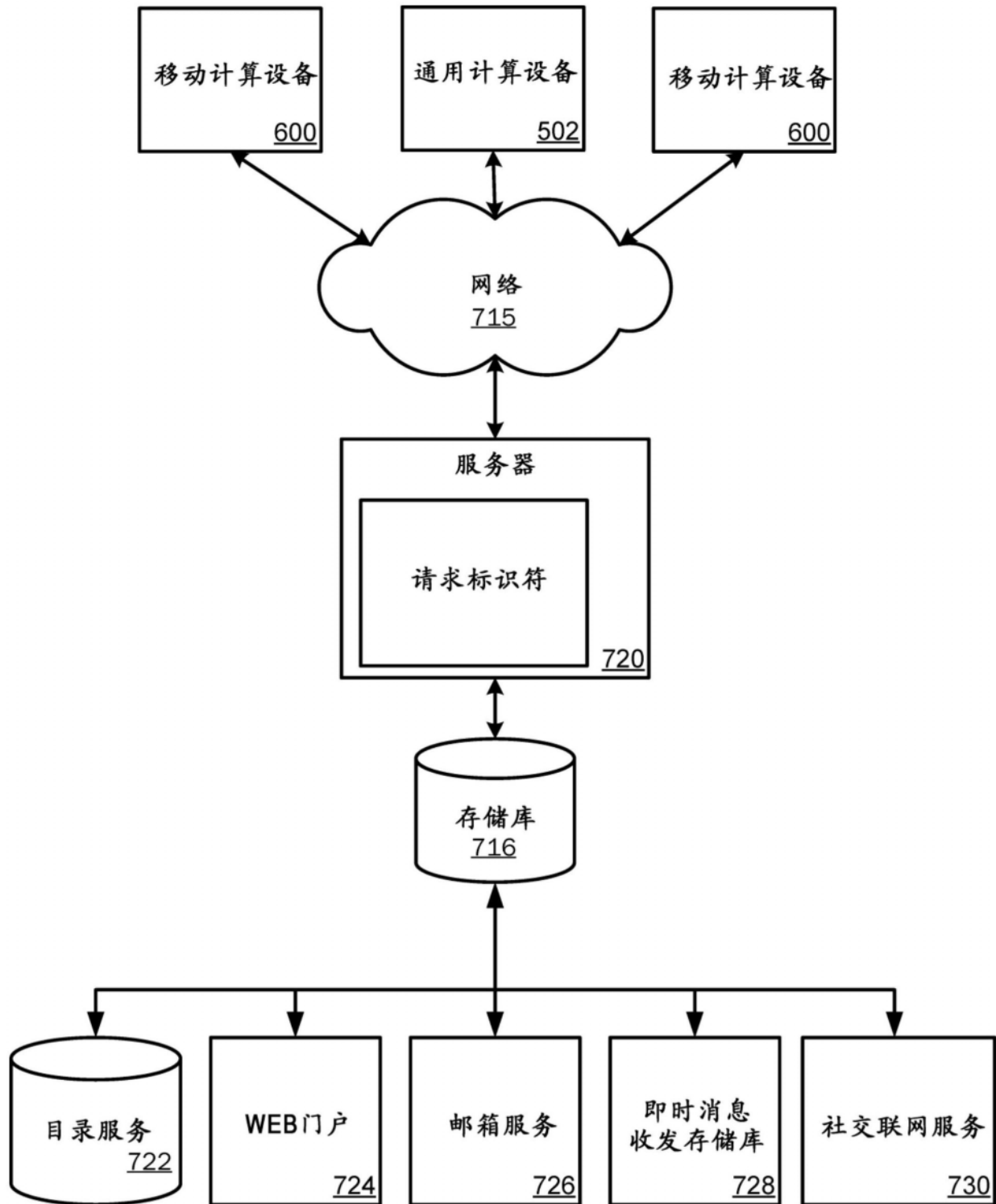


图7