



(12) 发明专利

(10) 授权公告号 CN 102591668 B

(45) 授权公告日 2015.04.08

(21) 申请号 201110001297.X

US 2010199037 A1, 2010.08.05,

(22) 申请日 2011.01.05

US 2005144292 A1, 2005.06.30,

(73) 专利权人 阿里巴巴集团控股有限公司

审查员 田冰

地址 英属开曼群岛大开曼岛资本大厦一座
四层 847 号邮箱

(72) 发明人 陈伟才 陈波 康华

(74) 专利代理机构 北京同达信恒知识产权代理
有限公司 11291

代理人 郭润湘

(51) Int. Cl.

G06F 9/445(2006.01)

(56) 对比文件

US 5664195 A, 1997.09.02,

权利要求书2页 说明书9页 附图4页

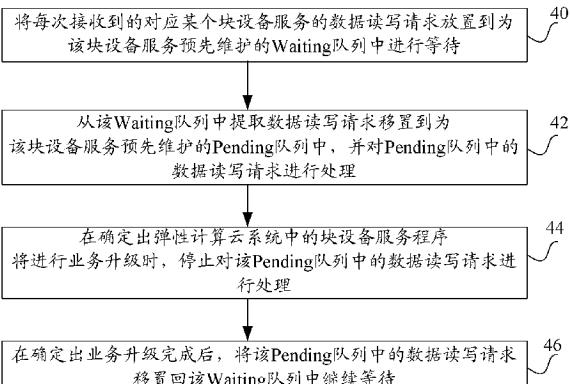
US 5664195 A, 1997.09.02,

(54) 发明名称

对弹性计算云系统升级的装置、方法及系统

(57) 摘要

本申请公开了一种弹性计算云系统中的服务器，包括块设备驱动装置和至少一个块设备服务装置，块设备驱动装置用于在接收到对应一个块设备服务装置的数据读写请求时，将接收到的数据读写请求放置到为该块设备服务装置预先维护的等待队列中，并从等待队列中提取数据读写请求移置到为该块设备服务装置预先维护的未决队列中，并将未决队列中的数据读写请求转发给该块设备服务装置处理；并在确定出该块设备服务装置将升级时，停止将未决队列中的数据读写请求转发给该块设备服务，以及在确定出业务升级完成后，将未决队列中的数据读写请求移置回等待队列中；块设备服务装置根据块设备驱动装置发来的数据读写请求，进行对应请求数据的读写操作。



1. 一种基于块设备存储服务实现数据存储的弹性计算云系统中的服务器，其特征在于，包括块设备驱动装置和至少一个块设备服务装置，其中：

块设备驱动装置，用于在接收到对应一个块设备服务装置的数据读写请求时，将接收到的数据读写请求放置到为该块设备服务装置预先维护的等待队列中，并从所述等待队列中提取数据读写请求并移置到为该块设备服务装置预先维护的未决队列中，并将所述未决队列中的数据读写请求转发给该块设备服务装置进行处理，并在接收到该块设备服务装置对接收到的数据读写请求的处理结果后，将转发的数据读写请求从所述未决队列中删除；并在确定出该块设备服务装置将进行业务升级时，停止将所述未决队列中的数据读写请求转发给该块设备服务装置进行处理，以及在确定出将进行业务升级或者业务升级完成后，将所述未决队列中的数据读写请求移置回所述等待队列中；

块设备服务装置，用于根据块设备驱动装置转发来的数据读写请求，针对块设备存储资源池进行对应请求数据的读写操作；

其中，所述块设备驱动装置具体通过检测自身与该块设备服务装置之间的通信连接，并在检测到自身与该块设备服务装置之间的通信连接断开后，确定该块设备服务装置将进行业务升级，以及在检测到自身与该块设备服务装置之间的通信连接恢复连接后，确定该块设备服务装置业务升级完成。

2. 如权利要求 1 所述的服务器，其特征在于，还包括客户端装置，用于控制块设备服务装置启动，并在启动后建立所述块设备驱动装置与启动的该块设备服务装置之间的通信连接；以及控制块设备驱动装置停止，并在停止后拆除所述块设备驱动装置与该块设备服务装置之间的通信连接。

3. 一种对基于块设备存储服务实现数据存储的弹性计算云系统进行升级的方法，其特征在于，包括：

在每次接收到对应一个块设备服务的数据读写请求时，将接收到的数据读写请求放置到为该块设备服务预先维护的等待队列中；

从所述等待队列中提取数据读写请求并移置到为该块设备服务预先维护的未决队列中，并将所述未决队列中的数据读写请求转发给该块设备服务进行处理，并在接收到该块设备服务对接收到的数据读写请求的处理结果后，将转发的数据读写请求从所述未决队列中删除；以及

在确定出该块设备服务将进行业务升级时，停止将所述未决队列中的数据读写请求转发给该块设备服务进行处理；以及

在确定出将进行业务升级或者业务升级完成后，将所述未决队列中的数据读写请求移置回所述等待队列中；

其中，确定出该块设备服务将进行业务升级、及业务升级完成，具体包括：

检测与该块设备服务之间的通信连接；

在检测到与该块设备服务之间的通信连接断开后，确定该块设备服务将进行业务升级；以及

在检测到与该块设备服务之间的通信连接恢复连接后，确定该块设备服务业务升级完成。

4. 如权利要求 3 所述的方法，其特征在于，从所述等待队列中提取数据读写请求并移

置到预先维护的未决队列中,具体包括:

从所述等待队列的队列头中提取数据读写请求;并在所述等待队列中将提取出的数据读写请求删除;
将提取出的数据读写请求移置到预先维护的未决队列的队列头中。

5. 一种对基于块设备存储服务实现数据存储的弹性计算云系统进行升级的装置,其特征在于,包括:

等待队列维护单元,用于在每次接收到对应一个块设备服务的数据读写请求时,将接收到的数据读写请求放置到为该块设备服务预先维护的等待队列中;

未决队列维护单元,用于从所述等待队列维护单元维护的等待队列中提取数据读写请求并移置到为该块设备服务预先维护的未决队列中;

读写请求处理单元,用于将所述未决队列维护单元维护的未决队列中的数据读写请求转发给该块设备服务进行处理,并在接收到该块设备服务对接收到的数据读写请求的处理结果后,将转发的数据读写请求从所述未决队列中删除;并在确定出该块设备服务将进行业务升级时,停止将所述未决队列中的数据读写请求转发给该块设备服务进行处理,以及在确定出将进行业务升级或者业务升级完成后,将所述未决队列中的数据读写请求移置回所述等待队列中;

其中,所述读写请求处理单元具体通过检测所述装置与该块设备服务之间的通信连接,并在检测到所述通信连接断开后,确定该块设备服务将进行业务升级,以及在检测到所述通信连接恢复连接后,确定该块设备服务业务升级完成。

6. 如权利要求5所述的装置,其特征在于,所述未决队列维护单元具体从所述等待队列的队列头中提取数据读写请求,并在所述等待队列中将提取出的数据读写请求删除,以及将提取出的数据读写请求移置到预先维护的未决队列的队列头中。

7. 一种基于块设备存储服务实现数据存储的弹性计算云系统,其特征在于,包括:

存储资源池,用于基于块设备存储服务方式存储弹性计算云系统中的数据;和

至少一个物理服务器,包括:块设备驱动装置和至少一个块设备服务装置;其中,所述块设备驱动装置用于在每次接收到对应一个块设备服务装置的数据读写请求时,将接收到的数据读写请求放置到为该块设备服务装置预先维护的等待队列中,并从所述等待队列中提取数据读写请求并移置到为该块设备服务装置预先维护的未决队列中,并根据所述未决队列中的数据读写请求,针对所述存储资源池进行对应请求数据的读写操作;并在确定出该块设备服务装置将进行业务升级时,停止对所述未决队列中的数据读写请求进行处理,以及在将进行业务升级或者业务升级完成后,将所述未决队列中的数据读写请求移置回所述等待队列中;

块设备服务装置,用于根据所述块设备驱动装置转发来的数据读写请求,针对所述存储资源池进行对应请求数据的读写操作;

其中,所述块设备驱动装置具体通过检测自身与该块设备服务装置之间的通信连接,并在检测到自身与该块设备服务装置之间的通信连接断开后,确定该块设备服务装置将进行业务升级,以及在检测到自身与该块设备服务装置之间的通信连接恢复连接后,确定该块设备服务装置业务升级完成。

对弹性计算云系统升级的装置、方法及系统

技术领域

[0001] 本申请涉及弹性计算云技术领域和块设备的存储服务技术领域，尤其是涉及一种对基于块设备存储服务实现数据存储的弹性计算云系统进行升级的装置、方法及系统。

背景技术

[0002] 弹性计算云 (Elastic Computing Cloud) 是一种云计算服务，即基于虚拟化技术，将本地的物理服务器虚拟化成多个虚拟服务器来使用，以此来提高资源的使用率，压缩成本。其中基于弹性计算云技术虚拟出来的虚拟服务器（即虚拟机）的存储服务都是基于块设备的远程存储服务，业界称之为弹性块存储服务 (Elastic Block Storage)。

[0003] 块设备的远程存储服务，是基于块设备驱动 (Block Device Driver) 提供的一种块级别的存储服务。本地的物理服务器和 / 或虚拟服务器可以将服务器的数据基于块的方式存储到远程的存储资源池中，从而节约本地的存储资源。简而言之，块设备的远程存储服务提供了一个类似物理裸硬盘的服务来存储本地的物理服务器和 / 或虚拟服务器的数据，且对数据的格式、文件系统的格式没有任何限制。

[0004] 随着弹性计算云技术的快速发展，弹性块存储服务对远程存储服务的高可用性提出了更高的要求。通常情况下，基于弹性计算云技术虚拟出来的虚拟服务器都数以万计，并将虚拟出来的这数以万计的虚拟服务器提供给第三方使用（例如政府部门、传统大中小型企业、互联网中小站长、个人用户等）。而众多的虚拟服务器组成的集群都是依托块设备的远程存储服务而运转起来的，显然，远程存储服务承载并流转着整个弹性计算云系统的数据，远程存储服务的可用性关系到整个弹性计算云系统的可用性。

[0005] 现有技术中，对基于块设备的远程存储服务实现数据存储的弹性计算云系统进行升级时，需要首先通知服务使用用户因升级过程将导致服务暂停，然后将整个弹性计算云系统中的各个虚拟机服务器的服务暂停，这样才能对存储服务进行服务升级（例如 BUG 的修复以及提供新的功能等），在重新部署后的服务正常工作后，再将整个弹性计算云系统中的虚拟机服务器恢复正常运转，从而完成一个升级过程。此外在升级过程中，如果用户正在使用弹性计算云系统中的虚拟服务器提供的 Web 服务访问 Web 网页时，就会导致用户不能继续访问 Web 网页，而如果用户正在使用弹性计算云系统中的虚拟服务器进行分布式计算时，也会导致用户不得不终止当前正在进行的分布式计算。

[0006] 由此可见，现有技术中对基于块设备的远程存储服务实现数据存储的弹性计算云系统进行升级时，其升级过程十分繁琐，升级过程中需要耗费较长时间，一般需要几个小时的时间才能完成升级操作，升级效率比较低，并且还会导致用户在升级过程中不得不中断当前正在进行处理的业务，因此不能保证在升级过程中为用户提供高可用的服务，降低了用户的使用体验。

发明内容

[0007] 本申请实施例提供一种对基于块设备存储服务实现数据存储的弹性计算云系统

进行升级的装置、方法及系统,以通过在线升级的方式提高升级效率,且避免升级对用户正在使用的业务中断所导致的用户使用体验差的问题。

[0008] 为解决上述问题,本申请实施例提供了一种基于块设备存储服务实现数据存储的弹性计算云系统中的服务器,包括块设备驱动装置和至少一个块设备服务装置,其中:块设备驱动装置,用于在接收到对应一个块设备服务装置的数据读写请求时,将接收到的数据读写请求放置到为该块设备服务装置预先维护的等待队列中,并从所述等待队列中提取数据读写请求并移置到为该块设备服务装置预先维护的未决队列中,并将所述未决队列中的数据读写请求转发给该块设备服务装置进行处理,并在接收到该块设备服务装置对接收到的数据读写请求的处理结果后,将转发的数据读写请求从所述未决队列中删除;并在确定出该块设备服务装置将进行业务升级时,停止将所述未决队列中的数据读写请求转发给该块设备服务进行处理,以及在确定出将进行业务升级或者业务升级完成后,将所述未决队列中的数据读写请求移置回所述等待队列中;块设备服务装置,用于根据块设备驱动装置转发来的数据读写请求,针对块设备存储资源池进行对应请求数据的读写操作。

[0009] 为解决上述问题,本申请实施例还提供了一种对基于块设备存储服务实现数据存储的弹性计算云系统进行升级的方法,包括在每次接收到对应一个块设备服务的数据读写请求时,将接收到的数据读写请求放置到为该块设备服务预先维护的等待队列中;从所述等待队列中提取数据读写请求并移置到为该块设备服务预先维护的未决队列中,并将所述未决队列中的数据读写请求转发给该块设备服务进行处理,并在接收到该块设备服务对接收到的数据读写请求的处理结果后,将转发的数据读写请求从所述未决队列中删除;以及在确定出该块设备服务将进行业务升级时,停止将所述未决队列中的数据读写请求转发给该块设备服务进行处理;以及在确定出将进行业务升级或者业务升级完成后,将所述未决队列中的数据读写请求移置回所述等待队列中。

[0010] 为解决上述问题,本申请实施例还提供了一种对基于块设备存储服务实现数据存储的弹性计算云系统进行升级的装置,包括等待队列维护单元,用于在每次接收到对应一个块设备服务的数据读写请求时,将接收到的数据读写请求放置到为该块设备服务预先维护的等待队列中;未决队列维护单元,用于从所述等待队列维护单元维护的等待队列中提取数据读写请求并移置到为该块设备服务预先维护的未决队列中;读写请求处理单元,用于将所述未决队列维护单元维护的未决队列中的数据读写请求转发给该块设备服务进行处理,并在接收到该块设备服务对接收到的数据读写请求的处理结果后,将转发的数据读写请求从所述未决队列中删除;并在确定出该块设备服务将进行业务升级时,停止将所述未决队列中的数据读写请求转发给该块设备服务进行处理,以及在确定出将进行业务升级或者业务升级完成后,将所述未决队列中的数据读写请求移置回所述等待队列中。

[0011] 为解决上述问题,本申请实施例提供了一种基于块设备存储服务实现数据存储的弹性计算云系统,包括存储资源池,用于基于块设备存储服务方式存储弹性计算云系统中的数据;和至少一个物理服务器,用于在每次接收到对应一个块设备服务的数据读写请求时,将接收到的数据读写请求放置到为该块设备服务预先维护的等待队列中,并从所述等待队列中提取数据读写请求并移置到为该块设备服务预先维护的未决队列中,并根据所述未决队列中的数据读写请求,针对所述存储资源池进行对应请求数据的读写操作;并在确定出该块设备服务将进行业务升级时,停止对所述未决队列中的数据读写请求进行处理,

以及在将进行业务升级或者业务升级完成后,将所述未决队列中的数据读写请求移置回所述等待队列中。

[0012] 本申请实施例提出的技术方案,在基于块设备存储服务实现数据存储的弹性计算云系统中的物理服务器中,通过分别为每个虚拟出来的块设备服务维护两个队列,一个是数据读写请求等待(Waiting)队列,用来维护针对一个发向块设备服务并等待其处理的数据读写请求,另一个是悬而未决(Pending)队列,用来维护物理服务器针对块设备服务当前正在处理的数据读写请求,在该块设备服务需要进行升级时,停止对为该块设备服务维护的Pending队列中的数据读写请求进行处理,并在该块设备服务升级完成并恢复运作后,将Pending队列中正在处理的数据读写请求再次放入到Waiting队列中重新接受调度处理,从而实现了在不中断存储服务的情况下完成对弹性计算云系统中存储服务的业务升级处理,这提高了服务升级的效率,且避免了因升级造成用户正在使用业务的中断,在升级过程中能够为用户提供高可用的服务,进而提高了用户的使用体验。

附图说明

[0013] 图1为本申请实施例提出的技术方案所应用的基于块设备存储服务实现数据存储的弹性计算云系统的拓扑结构图;

[0014] 图2为本申请实施例提出的改进后的物理服务器的组成结构示意图;

[0015] 图3为本申请实施例中改进后的物理服务器向本地或远程存储资源池读写数据的内部处理流程示意图;

[0016] 图4为本申请实施例改进后的物理服务器中的块设备驱动装置的具体工作原理流程图;

[0017] 图5为本申请实施例改进后的物理服务器中的块设备驱动装置在块设备服务装置进行业务升级前后对维护的Waiting队列、Pending队列进行调度的处理示意图;

[0018] 图6为本申请实施例中改进的物理服务器中新增的块设备驱动装置的具体组成结构示意图;

[0019] 图7为本申请实施例中在进行服务器业务升级时,控制物理服务器中的块设备驱动装置的操作示意图。

具体实施方式

[0020] 本申请针对现有技术中,在对基于块设备存储服务实现数据存储的弹性计算云系统进行升级时,由于没有热部署实现在线升级方式,而是需要对弹性计算云系统中的各个虚拟服务器停机后进行业务升级,升级完成后再启动各个虚拟服务器,这无论对于存储服务提供方来说,还是对于存储服务使用方来说,其维护成本都会大幅增加,并大大降低了服务的可用性。本申请针对此问题,提出在对基于块设备存储服务实现数据存储的弹性计算云系统中实现块设备存储服务的热部署,以实现在线升级存储服务,即在不中断存储服务的同时进行存储服务的升级处理,以此来提升服务提供方存储服务的可用性,并增强存储服务的透明性,使得存储服务的使用方不用关注后端服务的升级过程。

[0021] 所谓块设备存储服务的热部署即块设备服务的在线升级(Online Upgrade)方式,由于块设备存储服务是基于块设备驱动技术的,因此弹性计算云系统中的物理服务器中包

含了内核态的块设备驱动服务程序，内核态的块设备驱动服务程序需要将数据读写请求转发给物理服务器包含的用户态的块设备服务程序，由用户态的块设备服务程序将读写请求转发给远程的网络存储资源池或本地物理磁盘存储资源池等等来进而处理。其中，内核态的块设备驱动服务程序相对于用户态的块设备服务程序而言，其更为稳定；而用户态的块设备服务程序需要频繁的版本升级来提供更优质的存储服务。按照本申请实施例提供的技术方案，其在弹性计算云系统中的物理服务器中提供了一个可以支持在线升级的块设备驱动服务程序，以提高存储服务的可用性，并改善用户的使用体验。

[0022] 如图 1 所示，为本申请实施例提出的技术方案所应用的基于块设备存储服务实现数据存储的弹性计算云系统的拓扑结构图，其中，弹性计算云系统中包括若干基于弹性计算云技术将本地的物理服务器虚拟出来的虚拟服务器（图中所示为虚拟服务器 1、虚拟服务器 2、虚拟服务器 3……虚拟服务器 n），各个虚拟服务器会将本地提供的基于块设备存储服务技术的数据存储服务使用远程的存储资源池或本地的存储资源池来实现，即各个虚拟服务器会将本地需要保存的数据存储到远程的存储资源池或本地设置的存储资源池中，并在后续有读取数据的请求时，从远程的存储资源池或本地设置的存储资源池中读取相关数据。这样各个虚拟服务器就可以共享远程或本地设置的存储资源池，而无需占用本地物理服务器本身的存储资源，从而较好地提高了弹性计算云系统的资源利用率。

[0023] 为实施本申请提出的技术方案，需要对上述介绍的弹性计算云系统中的物理服务器的功能进行改进，以实现对存储服务进行在线升级的目的。

[0024] 如图 2 所示，为本申请实施例提出的改进后的物理服务器的组成结构示意图，具体包括内核态的块设备驱动装置 20、用户态的客户端装置 21 和至少一个块设备服务装置 22、以及本地或远程存储资源池 23，其中块设备驱动装置 20 可以控制多个块设备服务装置 22 的工作，各组成部分通过下述工作原理来实现存储服务的在线升级：

[0025] 内核态的块设备驱动装置 20，其接收来自物理服务器底层的读写调度层的数据读写请求，并将数据读写请求通过例如 TCP 协议等转发给用户态的块设备服务装置 22。为了实现支持热部署的在线升级功能，块设备驱动装置 20 分别为物理服务器虚拟出来的每个块设备服务装置 22 维护了两个队列，其中一个为等待 (Waiting) 队列，块设备驱动装置 20 每次接收到物理服务器底层的读写调度层发来的对应某个块设备服务装置 22 的数据读写请求时，就将该接收到的读写请求放入到为该块设备服务装置 22 预先维护的 Waiting 队列中等待后续处理；另外一个为悬而未决 (Pending) 队列，块设备驱动装置 20 每次处理该块设备服务装置 22 的数据读写请求时，都会先从为该块设备服务装置 22 预先维护的 Waiting 队列中提取数据读写请求并移放到为该块设备服务装置 22 预先维护的 Pending 队列中，并将该 Pending 队列中的请求转发给该块设备服务装置 22 进行处理，可见 Pending 队列中的读写请求都是物理服务器当前正在处理的读写请求，当块设备服务装置 22 正确处理完块设备驱动装置 20 转发过来的未决队列的读写请求后，会向块设备驱动装置 20 发送正确处理的结果消息，块设备驱动装置 20 接收到正确处理的结果消息后，会将转发的这些处理好的读写请求从未决队列中删除掉，接着处理悬而未决队列中的其他读写请求，如此循环。后续当块设备驱动装置 20 确定出该块设备服务装置 22 将进行业务升级时，停止将针对该块设备服务装置 22 而预先维护的 Pending 队列中的数据读写请求转发给块设备服务装置 22 进行处理，以及在后续进而确定出将进行业务升级或业务升级完成后，块设备驱动装置 20

将为该块设备服务装置 22 预先维护的 Pending 队列中的请求重新加入到为该块设备服务装置 22 预先维护的 Waiting 队列中等待后续重新被调用并处理。

[0026] 较优地,块设备驱动装置 20 可以通过检测自身与用户态的块设备服务装置 22 之间的 TCP 通信连接是否正常来判断块设备服务装置 22 是否将进行升级操作或是已经完成升级,当在检测到自身与块设备服务装置 22 之间的 TCP 通信连接断开后,可以确定块设备服务装置 22 将进行业务升级,在检测到自身与块设备服务装置 22 之间的 TCP 通信连接恢复连接后,可以确定块设备服务装置 22 业务升级已完毕。

[0027] 用户态的块设备服务装置 22,其通过例如 TCP 协议等接收来自内核态的块设备驱动装置 20 转发过来的数据读写请求,即接收块设备驱动装置 20 处理的对应该块设备服务装置 22 预先维护的 Pending 队列中的数据读写请求,在接收到读写请求后,对该请求进行解析操作,解析包括验证该读写请求是否合法、具体是读请求还是写请求,读写请求的起始位置及长度信息等,并根据解析结果向本地或远程的存储资源池进行数据的读操作或者写操作。如果是读请求,则从本地或远程的存储资源池 23 中将请求的数据读取出来,并通过例如 TCP 协议等将读取到的数据及处理结果返回给内核态的块设备驱动装置 20 ;如果是写请求,则将请求写入的数据存储到本地或远程的存储资源池 23 中,并通过例如 TCP 协议等将写请求的处理结果返回给内核态的块设备驱动装置 20 。

[0028] 用户态的客户端装置 21,其可以通过内核态的块设备驱动装置 20 控制各个块设备服务装置 22 启动工作或停止工作。在控制块设备服务装置 22 启动工作时,用户通过客户端装置 21 发起创建块设备服务的请求后,客户端装置 21 会先与块设备服务装置 22 建立 TCP 连接,然后再把创建好的 TCP Socket 句柄通过 ioctl 指令传递给块设备驱动装置 20,此后客户端装置 21 就可以关闭该 TCP Socket 套接字,这样客户端装置 21 就实现了在内核态的块设备驱动装置 20 与用户态的块设备服务装置 22 之间建立起了 TCP 连接,后续所有发向块设备驱动装置 20 的数据读写请求都可以通过 TCP 连接转发给块设备服务装置 22。在控制块设备服务装置 22 停止工作时,用户通过客户端装置 21 发起停止某个块设备服务的请求后,客户端装置 21 会通过 ioctl 指令向块设备驱动装置 20 发起停止请求,块设备驱动装置 20 接收到该请求后,会先向块设备服务装置 22 发送停止请求,块设备服务装置 22 接收到该请求后关闭 TCP Socket 套接字后安全退出;块设备驱动装置 20 收到块设备服务装置 22 的停止反应后,停止处理 Pending 队列中的读写请求,这样块设备服务装置 22 就停止工作了。当块设备服务装置 22 处于升级的过程中,客户端装置 21 可每隔一秒尝试与块设备服务装置 22 建立 TCP 链接。如果新的 TCP 链接建立成功后,即表明块设备服务装置 22 升级完成,然后客户端装置 21 将新创建好的 TCP Socket 句柄通过 ioctl 指令传递给块设备驱动装置 20。这样块设备服务装置 22 升级完成后,块设备驱动装置 20 与块设备服务装置 22 建立了新的 TCP 链接进行数据读写请求的通信处理。

[0029] 当然,该客户端装置 21 为本申请物理服务器中的可选组成部分,除了使用客户端装置 21 来实现用户对用户态的块设备服务装置 22 的工作状态进行控制外,还可以基于其他例如定时功能、特定事件触发功能等触发机制自动实现对用户态的块设备服务装置 22 的工作状态进行控制。

[0030] 本地或远程存储资源池 23,一般情况下存储资源池 23 是类似于分布式文件系统的存储服务器,该资源池 23 既可以是本地部署也可以是远程部署,大多数情形下都是远程

部署的存储资源池。存储资源池 23 中存储有弹性计算云系统中的各个虚拟服务器中的服务数据。具体地,在接收到物理服务器中的某个块设备服务装置 22 的数据读请求时,就从自身中读取相关请求的数据,并将读取到的数据反馈给该块设备服务装置 22 ;在接收到物理服务器中的某个块设备服务装置 22 的数据写请求时,就将请求写入的数据存储到自身中。

[0031] 如图 3 所示,为本申请实施例中改进后的物理服务器向本地或远程存储资源池读写数据的内部处理流程示意图,其中读取数据的具体流程为:

[0032] 当本地的物理服务器中的文件系统层将对应某个块设备服务装置的数据读请求通过通用块层、W/R 调度层发送到块设备驱动装置后,块设备驱动装置会将该请求放入为该块设备服务装置预先维护的 Waiting 队列中等待后续处理,并在后续从该 Waiting 队列中提取出该数据读请求放入到为该块设备服务装置预先维护的 Pending 队列中并对该读请求进行处理,即将该读请求转发给该块设备服务装置,该块设备服务装置会根据读请求从存储池中读取相关请求的数据,并将读取结果通过 TCP 连接返回给块设备驱动装置,块设备驱动装置进而将读取结果通过 W/R 调度层、通用块层转发给文件系统层,文件系统层再将读取结果返回给系统调用。

[0033] 写入数据的具体流程为:

[0034] 本地的物理服务器将数据写请求和请求写入的数据通过系统调用中的文件系统层、通用块层、W/R 调度层存储到块设备驱动装置,块设备驱动装置会将该写请求放入为该块设备服务装置预先维护的 Waiting 队列中等待后续处理,并在后续从该 Waiting 队列中提取出该数据写请求放入到为该块设备服务装置预先维护的 Pending 队列中并对该写请求进行处理,即将该数据写请求和待写入的数据转发给该块设备服务装置,该块设备服务装置通过解析数据写请求实现将待写入的数据保存到存储资源池中。

[0035] 如图 4 所示,为本申请实施例改进后的物理服务器中的块设备驱动装置的具体工作原理流程图,其中,该块设备驱动装置会分别为每个块设备服务装置预先维护两个队列,一个队列为 Waiting 队列,用于放入每次发向对应块设备服务装置并等待其处理的数据读写请求,使其进入等待处理排队状态,另一个队列为 Pending 队列,用于放入针对对应块设备服务装置当前正在处理的数据读写请求,基于维护的该两个队列,块设备驱动装置的具体工作原理如下:

[0036] 步骤 40、块设备驱动装置在每次接收到物理服务器底层发来的针对某个块设备服务装置的数据读写请求时,将接收到的数据读写请求放置到为该块设备服务装置预先维护的 Waiting 队列中进行等待;

[0037] 步骤 42,块设备驱动装置在处理对应该块设备服务装置的数据读写请求时,先从该 Waiting 队列中提取数据读写请求并移置到为该块设备服务装置预先维护的 Pending 队列中,并将该 Pending 队列中的数据读写请求转发给用户态的该块设备服务装置进行处理,以及在接收到该块设备服务装置对接收到的数据读写请求正确处理的结果后,将转发的数据读写请求从 Pending 队列中删除;

[0038] 步骤 44,块设备驱动装置在确定出该块设备服务装置将进行业务升级时,停止将该 Pending 队列中的所有数据读写请求转发给块设备服务装置进行处理;

[0039] 步骤 46,块设备驱动装置在确定出将进行业务升级或业务升级完成后,将该

Pending 队列中的数据读写请求再移置回该 Waiting 队列中继续等待后续被再次调用并被处理。具体地，块设备驱动装置可以通过检测自身与该块设备服务装置之间的通信连接状态，来判断该块设备服务装置将进行业务升级还是已经升级结束。更具体地在检测到自身与该块设备服务装置之间的通信连接断开后，可以确定该块设备服务装置将进行业务升级，以及在检测到自身与该块设备服务装置之间的通信连接恢复连接后，可以确定该块设备服务装置已经完成业务升级。

[0040] 如图 5 所示，为本申请实施例改进后的物理服务器中的块设备驱动装置在块块设备服务装置进行业务升级前后对维护的 Waiting 队列、Pending 队列进行调度的处理示意图。内核态的块设备驱动装置为了支持块设备服务装置的热部署（即在线升级），需要对接收到的数据读写请求的状态进行维护。在块设备驱动装置中，分别针对每个块设备服务装置采用两个数据读写队列来维护读写请求的状态，一个是等待队列（Waiting Queue），用来为每个块设备服务装置维护物理服务器中的 W/R 调度层下发下来的 W/R 请求，这些请求在 Waiting 队列中排队等待块设备驱动装置进行后续处理；另一个是悬而未决队列（Pending Queue），用来维护块设备驱动装置针对每个块设备服务装置当前正在处理的 W/R 请求，Pending 队列中的这些 W/R 请求没有处理完成，可能处于块设备驱动装置正等待对应块设备服务装置的处理应答状态。其中，在某个块设备服务装置未进行业务升级期间，块设备驱动装置在处理 W/R 请求时，会先从为该块设备服务装置预先维护的 Waiting 队列的头部取出一个 W/R 请求，并将该取出的 W/R 请求从该 Waiting 队列中删除掉，然后将该 W/R 请求加入到为该块设备服务装置预先维护的 Pending 队列的头部，并将该 Pending 队列中的各个 W/R 请求转发给块设备服务装置进行处理，并在接收到该块设备服务装置对接收到的数据读写请求正确处理的结果后，将转发的数据读写请求从 Pending 队列中删除。当该块设备服务装置需要业务升级时，块设备驱动装置将停止将对应 Pending 队列中的各个 W/R 请求转发给该块设备服务装置进行处理，即不再向该块设备服务装置发送新的 W/R 请求。当该块设备服务装置将进行升级或者升级完毕后，块设备驱动装置将对应的 Pending 队列中的所有 W/R 请求（如图 5 中所示的请求 1、2、3、4、5、6、...）重新加入到对应的 Waiting 队列中，以在后续重新接收块设备驱动装置的再次调度处理。

[0041] 由此可见，通过在弹性计算云系统中的物理服务器中增加块设备驱动装置，并通过该块设备驱动装置对为每个块设备服务装置预先维护的两个队列进行调度，可以实现在不中断为数据存储使用者提供存储服务的同时，完成后台的存储服务业务升级处理，从而提高了弹性计算云系统中数据存储的可用性。此外，上述方案对于使用者而言也是透明的，使数据存储使用者几乎感知不到存储服务业务升级对于自己使用数据存储业务所造成的影响，从而提高了用户的使用体验。

[0042] 相应地，如图 6 所示，为本申请实施例中改进的物理服务器中新增的块设备驱动装置的具体组成结构示意图，具体包括等待队列维护单元 60，用于分别为每个块设备服务装置预先维护一个 Waiting 队列，用于放入每次接收到的针对相应块设备服务装置的数据读写请求，使其进入等待处理排队状态；未决队列维护单元 61，用于分别为每个块设备服务装置预先维护另一个 Pending 队列，用于放入针对相应块设备服务装置当前正在处理的数据读写请求；具体地，从 Waiting 队列维护单元 60 为某个块设备服务装置维护的 Waiting 队列中提取数据读写请求并移置到为该块设备服务装置预先维护的 Pending 队列

中；具体地，未决队列维护单元 61 可以从对应的 Waiting 队列的队列头中提取数据读写请求，并在该 Waiting 队列中将提取出的数据读写请求删除掉，进而将提取出的数据读写请求移到对应预先维护的 Pending 队列的队列头中。读写请求处理单元 62，用于将 Pending 队列维护单元 61 维护的对应 Pending 队列中的数据读写请求转发给块设备服务装置进行处理，并在接收到该块设备服务装置对接收到的数据读写请求正确处理的结果后，将转发的数据读写请求从 Pending 队列中删除；并在确定出弹性计算云系统中的块设备服务装置将进行业务升级时，停止将 Pending 队列中的数据读写请求转发给块设备服务装置进行处理，以及在确定出将进行业务升级或者业务升级完成后，将对应的 Pending 队列中的数据读写请求移置回对应的 Waiting 队列中继续等待后续再次被读写请求处理单元 62 的调度处理。具体地读写请求处理单元 62 具体可以通过检测块设备驱动装置与块设备服务装置之间的通信连接状态，在检测到该通信连接断开后，可以确定块设备服务装置将进行业务升级，以及在检测到该通信连接恢复连接后，可以确定块设备服务装置业务升级完成。

[0043] 如图 7 所示，为本申请实施例中在进行服务器业务升级时，控制物理服务器中的块设备驱动装置的操作示意图，其操作流程具体如下：

[0044] 步骤 70，在存储服务升级之前，升级维护人员通过用户态的客户端装置向内核态的块设备驱动装置发出 ioctl 指令，指示块设备驱动装置维持 (Hold) 所有的 W/R 请求。

[0045] 步骤 72，接下来，升级维护人员就按照正常的升级流程对存储服务进行业务升级（业务升级包括停止服务，升级服务程序，发布新的服务等）。这样就完成了一次在线对存储服务进行业务升级的热部署，而在热部署的过程中，块设备驱动装置会将维持 (Hold) 所有的 W/R 请求。

[0046] 步骤 74，当部署好后，块设备驱动装置再将所有的 W/R 请求放开，数据存储服务将被继续处理。

[0047] 本领域的技术人员应明白，本申请的实施例可提供为方法、装置（设备）、或计算机程序产品。因此，本申请可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且，本申请可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质（包括但不限于磁盘存储器、CD-ROM、光学存储器等）上实施的计算机程序产品的形式。

[0048] 本申请是参照根据本申请实施例的方法、装置（设备）和计算机程序产品的流程图和 / 或方框图来描述的。应理解可由计算机程序指令实现流程图和 / 或方框图中的每一流程和 / 或方框、以及流程图和 / 或方框图中的流程和 / 或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器，使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能的装置。

[0049] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中，使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品，该指令装置实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能。

[0050] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上，使得在计

算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理，从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和 / 或方框图一个方框或多个方框中指定的功能的步骤。

[0051] 尽管已描述了本申请的优选实施例，但本领域内的技术人员一旦得知了基本创造性概念，则可对这些实施例作出另外的变更和修改。所以，所附权利要求意欲解释为包括优选实施例以及落入本申请范围的所有变更和修改。

[0052] 显然，本领域的技术人员可以对本申请进行各种改动和变型而不脱离本申请的精神和范围。这样，倘若本申请的这些修改和变型属于本申请权利要求及其等同技术的范围之内，则本申请也意图包含这些改动和变型在内。

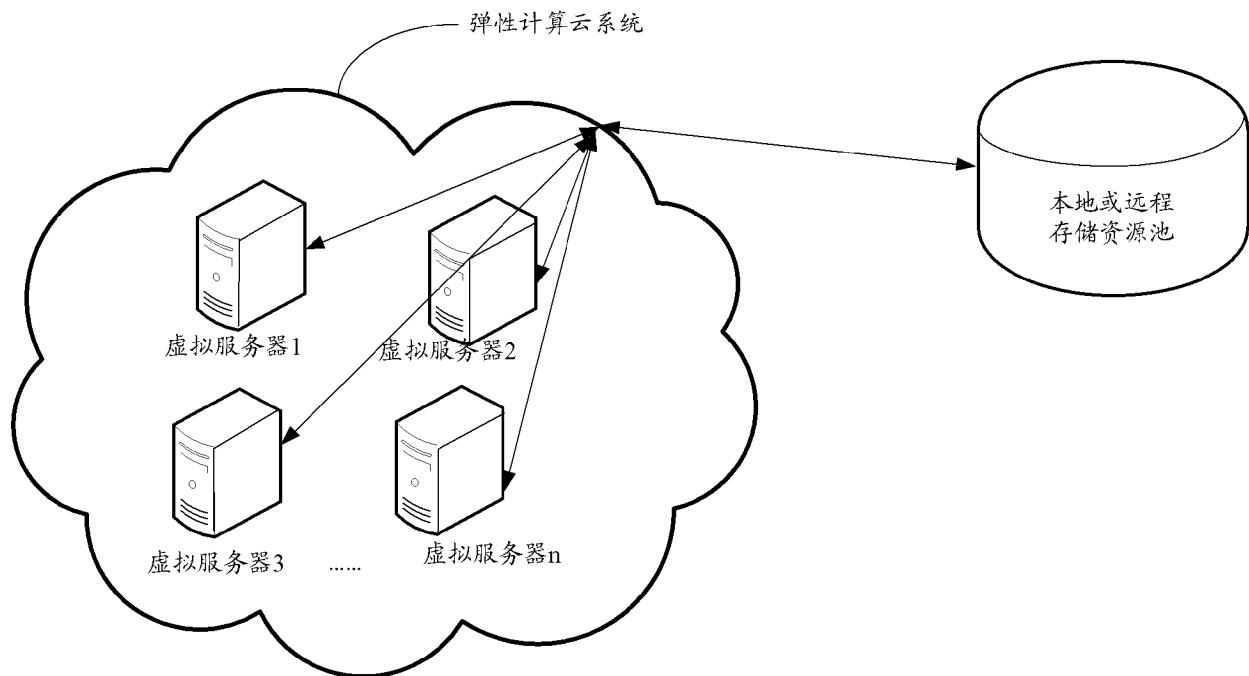


图 1

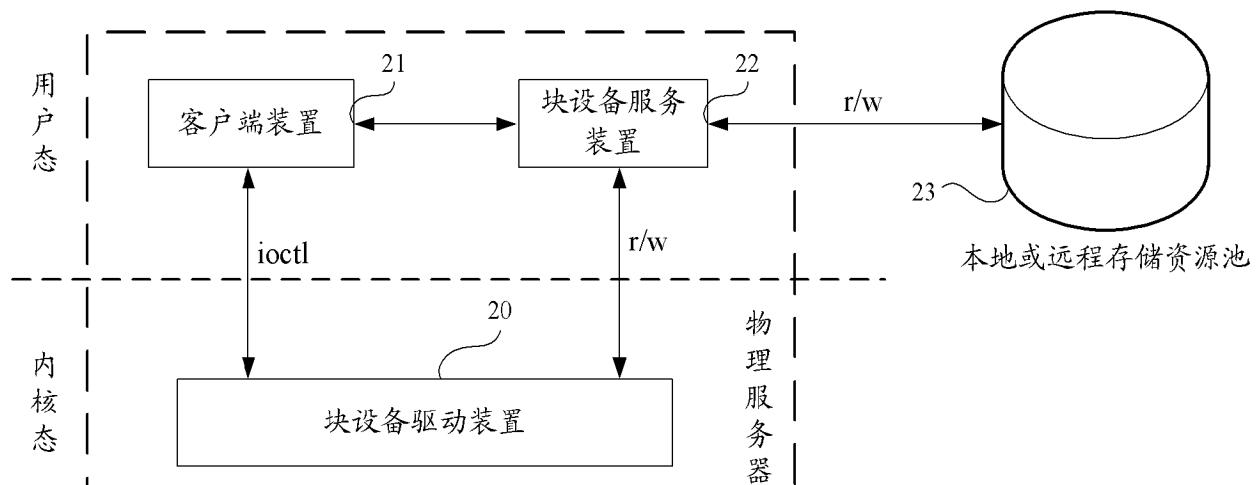


图 2

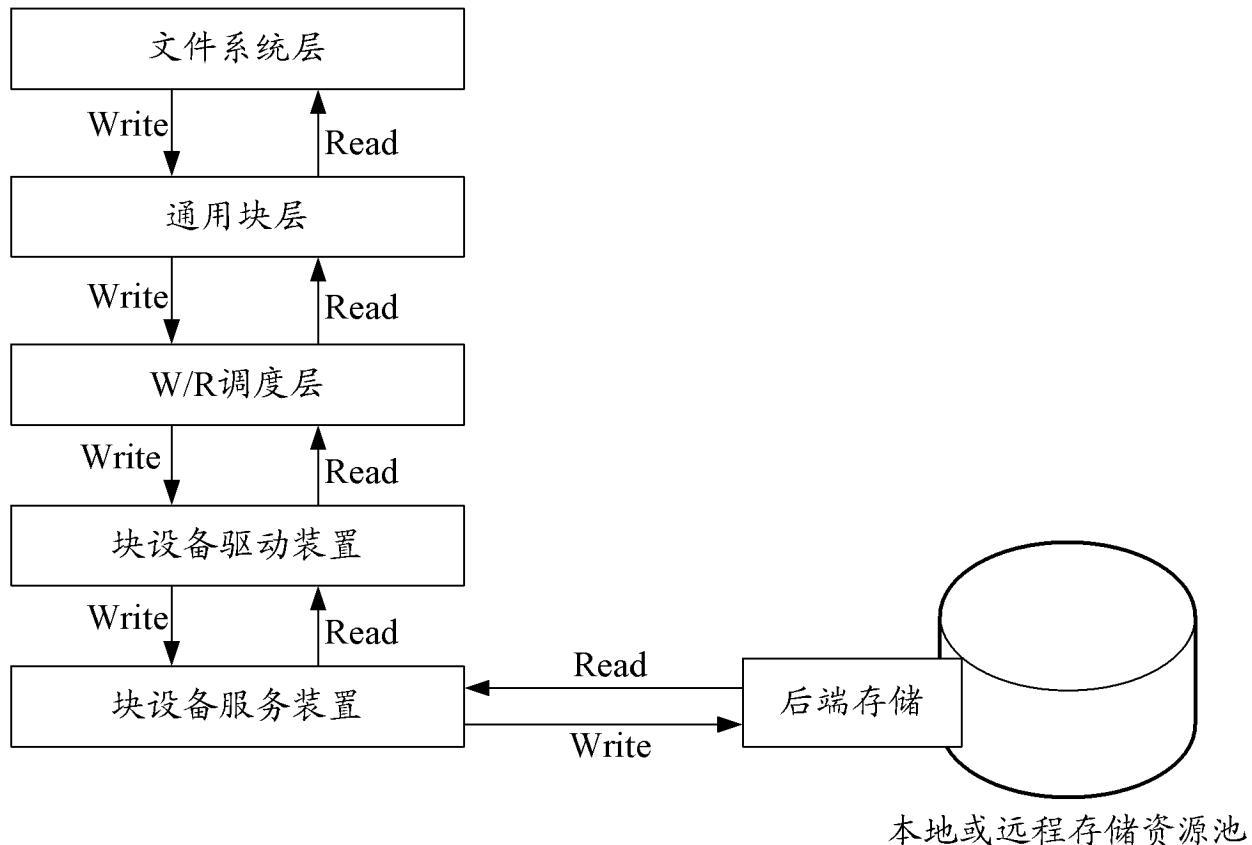


图3

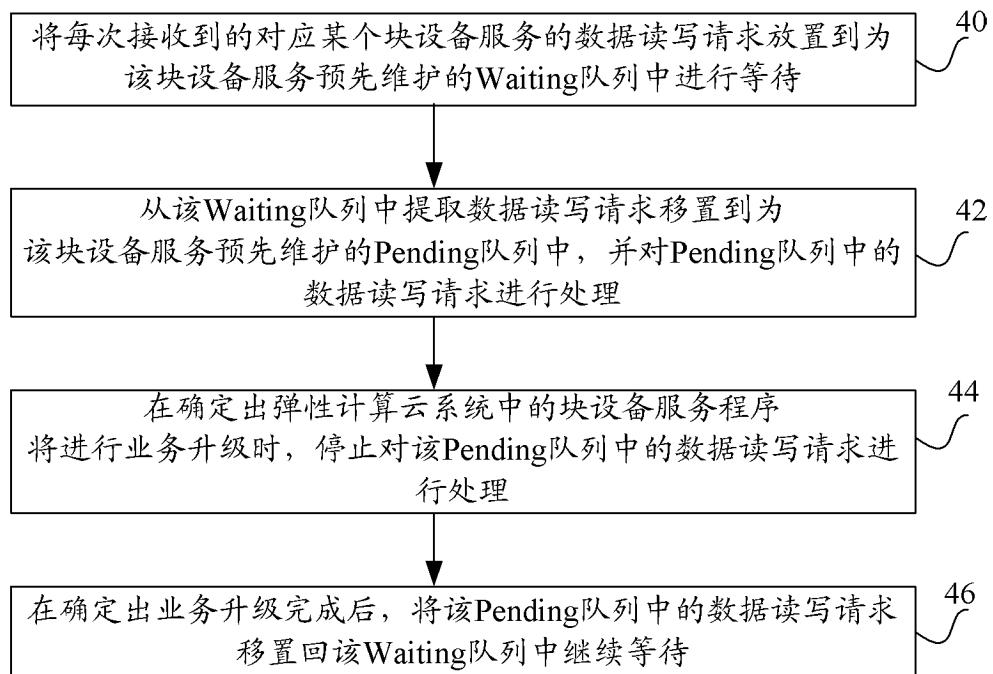


图4

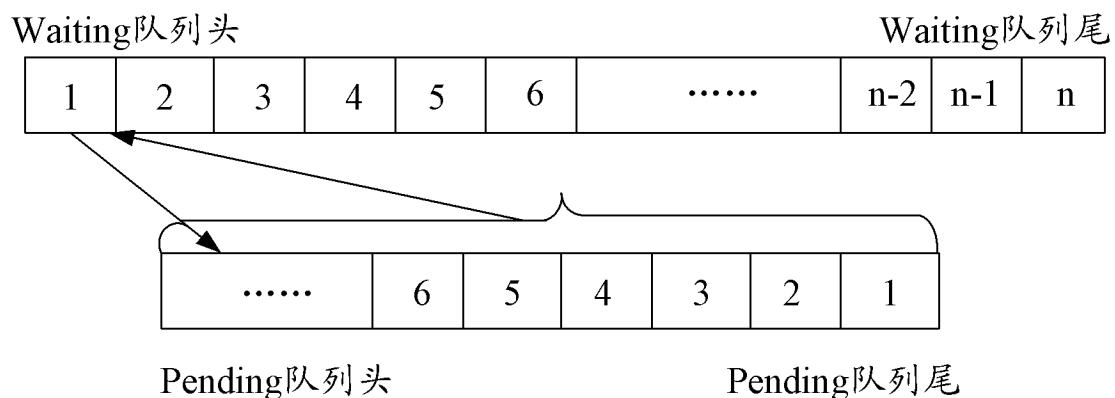


图 5

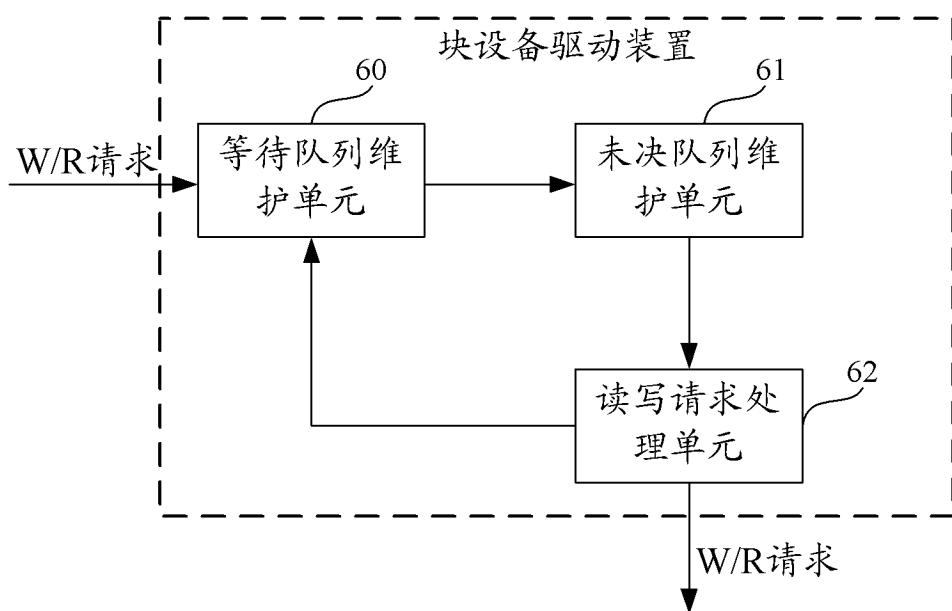


图 6

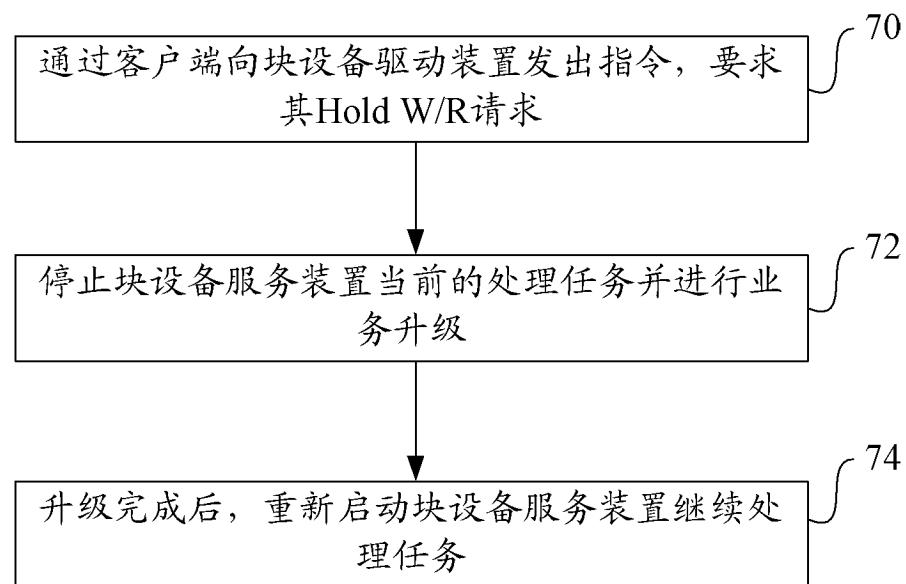


图 7