



(12) 发明专利申请

(10) 申请公布号 CN 115145492 A

(43) 申请公布日 2022. 10. 04

(21) 申请号 202210906226.2

(22) 申请日 2022.07.29

(71) 申请人 山东云海国创云计算装备产业创新中心有限公司

地址 250000 山东省济南市中国(山东)自由贸易试验区济南片区浪潮路1036号浪潮科技园S01楼35层

(72) 发明人 李幸远 王江 孙华锦 李树青

(74) 专利代理机构 北京连和连知识产权代理有限公司 11278

专利代理师 张涛 杨帆

(51) Int. Cl.

G06F 3/06 (2006.01)

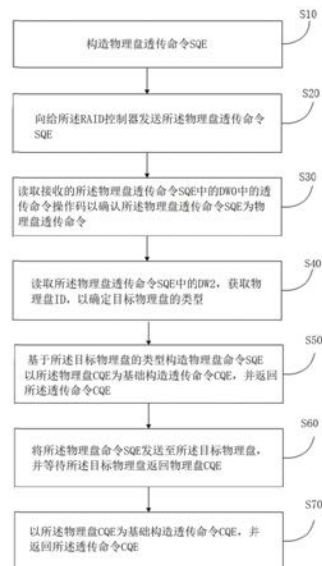
权利要求书2页 说明书11页 附图5页

(54) 发明名称

一种RAID控制器下物理盘管理的方法和系统

(57) 摘要

本发明提供了一种RAID控制器下物理盘管理的方法、系统、存储介质及设备,方法包括:构造物理盘透传命令SQE;向给RAID控制器发送物理盘透传命令SQE;读取接收的物理盘透传命令SQE中的DWO中的透传命令操作码以确认物理盘透传命令SQE为物理盘透传命令;读取物理盘透传命令SQE中的DW2,获取物理盘ID,以确定目标物理盘的类型;基于目标物理盘的类型构造物理盘命令SQE;将物理盘命令SQE发送至目标物理盘,并等待目标物理盘返回物理盘CQE;以及以物理盘CQE为基础构造透传命令CQE,并返回透传命令CQE。根据本发明使主机透过RAID控制器直接向RAID控制器下的物理盘发送命令,从而降低RAID管理接口的复杂读,提升接口的灵活性。



1. 一种RAID控制器下物理盘管理的方法,其特征在于,包括以下步骤:  
构造物理盘透传命令SQE;  
向给所述RAID控制器发送所述物理盘透传命令SQE;  
读取接收的所述物理盘透传命令SQE中的DW0中的透传命令操作码以确认所述物理盘透传命令SQE为物理盘透传命令;  
读取所述物理盘透传命令SQE中的DW2,获取物理盘ID,以确定目标物理盘的类型;  
基于所述目标物理盘的类型构造物理盘命令SQE;  
将所述物理盘命令SQE发送至所述目标物理盘,并等待所述目标物理盘返回物理盘CQE;以及  
以所述物理盘CQE为基础构造透传命令CQE,并返回所述透传命令CQE。
2. 根据权利要求1所述的方法,其特征在于,所述构造物理盘透传命令SQE包括以下步骤:  
将所述透传命令操作码以及主机命令标识符填入所述物理盘透传命令SQE中的DW0;  
将被透传命令中的所述物理盘ID以及被透传命令操作码分别填入所述物理盘透传命令SQE的DW2和DW3中;以及  
将所述被透传命令中的其他字段内容复制到所述物理盘透传命令SQE的对应字段。
3. 根据权利要求2所述的方法,其特征在于,所述构造物理盘命令SQE包括以下步骤:  
将所述物理盘透传命令SQE的DW3中的所述被透传命令操作码填入所述物理盘命令SQE的DW0中;  
重新生成新的物理盘命令标识符并填入所述物理盘SQE的DW0中;以及  
将所述物理盘透传命令SQE中的其他字段内容复制到物理盘命令SQE中对应位置。
4. 根据权利要求3所述的方法,其特征在于,所述构造透传命令CQE包括以下步骤:  
将所述物理盘透传命令SQE的DW0中的所述主机命令标识符填入所述透传命令CQE中的DW3中;  
将所述目标物理盘返回的所述物理盘CQE中的DW0和DW1复制到所述透传命令CQE中的对应位置;以及  
将所述目标物理盘返回的所述物理盘CQE中的DW3中的状态信息位复制到所述透传命令CQE中的对应位置。
5. 根据权利要求1所述的方法,其特征在于,所述构造物理盘透传命令SQE包括以下步骤:  
将所述透传命令操作码以及主机命令标识符填入所述物理盘透传命令SQE中的DW0;  
将被透传命令中的所述物理盘ID以及被透传命令操作码分别填入所述物理盘透传命令SQE的DW2和DW3中;以及  
将所述被透传命令中的FIS帧填入所述物理盘透传命令SQE DW10至DW14或者将所述被透传命令中的CDB帧填入所述物理盘透传命令SQE中的DW10至DW13;  
在所述被透传命令中存在数据传输阶段时,根据NVMe标准构造PRP数据结构并填入所述物理盘透传命令SQE的DW6至DW9字段。
6. 根据权利要求5所述的方法,其特征在于,  
在确定所述目标物理盘的类型为SATA盘时,构造物理盘命令SQE包括从所述物理盘透

传命令SQE的DW10至DW14中取出FIS并填入所述目标物理盘中的SATA控制器,或者

在确定所述目标物理盘的类型为SAS盘时,构造物理盘命令SQE包括从所述物理盘透传命令SQE的DW10至DW13中取出CDB并填入所述目标物理盘中的SAS控制器。

7. 根据权利要求6所述的方法,其特征在于,所述构造物理盘命令SQE还包括以下步骤:

从所述物理盘透传命令SQE的DW6至DW9中取出PRP数据结构,并按照所述SATA控制器要求构造SATA数据缓冲描述符结构或者按照所述SAS控制器要求构造SAS数据缓冲描述符结构。

8. 根据权利要求6所述的方法,其特征在于,所述方法还包括:

在所述目标物理盘接收到所述物理盘命令SQE之后,操作所述SATA控制器或者SAS控制器将所述物理盘命令SQE发送给所述目标物理盘;以及

在所述被透传命令中存在数据传输阶段时,由所述SATA控制器或者SAS控制器完成所述主机的数据缓冲区和所述目标物理盘之间的数据搬移。

9. 根据权利要求6所述的方法,其特征在于,所述构造透传命令CQE包括以下步骤:

将所述SATA控制器或者SAS控制器返回的所述物理盘CQE中的命令执行状态填入所述透传命令CQE中DW3中的状态;

将返回的所述物理盘CQE中的状态数据填入到所述透传命令CQE的DW0中;以及

将返回的所述物理盘CQE中的DW3中的状态信息位复制到所述透传命令CQE中的对应位置。

10. 一种RAID控制器下物理盘管理的系统,其特征在于,包括:

透传命令SQE构造模块,该透传命令SQE构造模块被配置为构造物理盘透传命令SQE;

透传命令SQE发送模块,该透传命令SQE发送模块被配置为将所述物理盘透传命令SQE发送给所述RAID控制器;

命令确认模块,该命令确认模块被配置为读取接收的所述物理盘透传命令SQE中的DW0中的透传命令操作码以确认所述物理盘透传命令SQE为物理盘透传命令;

物理盘类型确认模块,该物理盘类型确认模块被配置为读取所述物理盘透传命令SQE中的DW2,获取物理盘ID,以确定目标物理盘的类型;

物理盘命令SQE构造模块,该物理盘命令SQE被配置为构造模块基于所述目标物理盘的类型构造物理盘命令SQE;

物理盘命令SQE处理模块,该物理盘命令SQE处理模块被配置为将所述物理盘命令SQE发送至所述目标物理盘,并等待所述目标物理盘返回物理盘CQE;以及

透传命令CQE处理模块,该透传命令CQE处理模块被配置为以所述物理盘CQE为基础构造透传命令CQE,并返回所述透传命令CQE。

## 一种RAID控制器下物理盘管理的方法和系统

### 技术领域

[0001] 本发明涉及计算机数据存储技术领域,尤其涉及RAID控制器下物理盘管理技术领域,具体涉及一种RAID控制器下物理盘管理的方法、系统、存储介质及设备。

### 背景技术

[0002] 由于物理盘Disk 1至Disk N直接连接到RAID控制器上,并未直接到主机,因此这些物理盘的发现、管理只能由RAID控制器来操作,主机无法看到和这些物理盘。如果主机要直接要透过RAID控制器向物理盘发送命令,只能在RAID管理命令接口上扩展新的专用命令接口来实现。通常流程是这样的:首先主机向RAID控制器发出命令,要获取物理盘上的特定信息。RAID控制器接收到命令后,根据物理盘类型,构造相应的物理盘命令并发送给物理盘,等物理盘返回数据后,RAID控制器再构造数据返回给主机。这种方法获取物理盘的每种信息都必须由RAID控制器来构造命令,这就导致RAID控制器必须为每种盘的每个命令实现接口。

[0003] 由于RAID控制器所连接的物理盘由多种协议,如NVMe盘、SAS盘和SATA盘,这些盘在命令接口上遵循不同的协议,如果为每种盘的每个命令相对应的在RAID管理接口上都扩展一个命令,会导致RAID管理接口的命令个数无法控制。且由于物理盘厂商的不同,导致其支持的命令集也不同,可能存在厂商自定义命令,这就导致了扩展专用命令的方法存在无法覆盖所有类型物理盘命令的可能性。从而导致RAID控制器内的RAID管理模块需求无法完全确定,从而导致实现困难。

[0004] 因此,针对问题,需要提出一种通用的命令接口,以降低RAID管理接口的复杂读,提升接口的灵活性。

### 发明内容

[0005] 有鉴于此,本发明的目的在于提出一种改进的RAID控制器下物理盘管理的方法、系统、存储介质及设备,以提高测试的覆盖率和灵活性。

[0006] 基于上述目的,一方面,本发明提供了一种RAID控制器下物理盘管理的方法,其中该方法包括以下步骤:

[0007] 构造物理盘透传命令SQE;

[0008] 向给所述RAID控制器发送所述物理盘透传命令SQE;

[0009] 读取接收的所述物理盘透传命令SQE中的DW0中的透传命令操作码以确认所述物理盘透传命令SQE为物理盘透传命令;

[0010] 读取所述物理盘透传命令SQE中的DW2,获取物理盘ID,以确定目标物理盘的类型;

[0011] 基于所述目标物理盘的类型构造物理盘命令SQE;

[0012] 将所述物理盘命令SQE发送至所述目标物理盘,并等待所述目标物理盘返回物理盘CQE;以及

[0013] 以所述物理盘CQE为基础构造透传命令CQE,并返回所述透传命令CQE。

[0014] 在根据本发明的RAID控制器下物理盘管理的方法的一些实施例中，“构造物理盘透传命令SQE”包括以下步骤：

[0015] 将所述透传命令操作码以及主机命令标识符填入所述物理盘透传命令SQE中的DW0；

[0016] 将被透传命令中的所述物理盘ID以及被透传命令操作码分别填入所述物理盘透传命令SQE的DW2和DW3中；以及

[0017] 将所述被透传命令中的其他字段内容复制到所述物理盘透传命令SQE的对应字段。

[0018] 在根据本发明的RAID控制器下物理盘管理的方法的一些实施例中，“构造物理盘命令SQE”包括以下步骤：

[0019] 将所述物理盘透传命令SQE的DW3中的所述被透传命令操作码填入所述物理盘命令SQE的DW0中；

[0020] 重新生成新的物理盘命令标识符并填入所述物理盘SQE的DW0中；以及

[0021] 将所述物理盘透传命令SQE中的其他字段内容复制到物理盘命令SQE中对应位置。

[0022] 在根据本发明的RAID控制器下物理盘管理的方法的一些实施例中，“构造透传命令CQE”包括以下步骤：

[0023] 将所述物理盘透传命令SQE的DW0中的所述主机命令标识符填入所述透传命令CQE中的DW3中；

[0024] 将所述目标物理盘返回的所述物理盘CQE中的DW0和DW1复制到所述透传命令CQE中的对应位置；以及

[0025] 将所述目标物理盘返回的所述物理盘CQE中的DW3中的状态信息位复制到所述透传命令CQE中的对应位置。

[0026] 在根据本发明的RAID控制器下物理盘管理的方法的一些实施例中，“构造物理盘透传命令SQE”包括以下步骤：

[0027] 将所述透传命令操作码以及主机命令标识符填入所述物理盘透传命令SQE中的DW0；

[0028] 将被透传命令中的所述物理盘ID以及被透传命令操作码分别填入所述物理盘透传命令SQE的DW2和DW3中；以及

[0029] 将所述被透传命令中的FIS帧填入所述物理盘透传命令SQE DW10至DW14或者将所述被透传命令中的CDB帧填入所述物理盘透传命令SQE中的DW10至DW13。

[0030] 在根据本发明的RAID控制器下物理盘管理的方法的一些实施例中，在所述被透传命令中存在数据传输阶段时，根据NVMe标准构造PRP数据结构并填入所述物理盘透传命令SQE的DW6至DW9字段。

[0031] 在根据本发明的RAID控制器下物理盘管理的方法的一些实施例中，

[0032] 在确定所述目标物理盘的类型为SATA盘时，构造物理盘命令SQE包括从所述物理盘透传命令SQE的DW10至DW14中取出FIS并填入所述目标物理盘中的SATA控制器，或者

[0033] 在确定所述目标物理盘的类型为SAS盘时，构造物理盘命令SQE包括从所述物理盘透传命令SQE的DW10至DW13中取出CDB并填入所述目标物理盘中的SAS控制器。

[0034] 在根据本发明的RAID控制器下物理盘管理的方法的一些实施例中，“构造物理盘

命令SQE”还包括以下步骤：

[0035] 从所述物理盘透传命令SQE的DW6至DW9中取出PRP数据结构，并按照所述SATA控制器要求构造SATA数据缓冲描述符结构或者按照所述SAS控制器要求构造SAS数据缓冲描述符结构。

[0036] 在根据本发明的RAID控制器下物理盘管理的方法的一些实施例中，所述方法还包括：

[0037] 在所述目标物理盘接收到所述物理盘命令SQE之后，操作所述SATA控制器或者SAS控制器将所述物理盘命令SQE发送给所述目标物理盘。

[0038] 在根据本发明的RAID控制器下物理盘管理的方法的一些实施例中，在所述被透传命令中存在数据传输阶段时，由所述SATA控制器或者SAS控制器完成所述主机的数据缓冲区和所述目标物理盘之间的数据搬移。

[0039] 在根据本发明的RAID控制器下物理盘管理的方法的一些实施例中，“构造透传命令CQE”包括以下步骤：

[0040] 将所述SATA控制器或者SAS控制器返回的所述物理盘CQE中的命令执行状态填入所述透传命令CQE中DW3中的状态；

[0041] 将返回的所述物理盘CQE中的状态数据填入到所述透传命令CQE的DW0中；以及

[0042] 将返回的所述物理盘CQE中的DW3中的状态信息位复制到所述透传命令CQE中的对应位置。

[0043] 本发明的另一方面，还提供了一种RAID控制器下物理盘管理的系统，其中包括：

[0044] 透传命令SQE构造模块，该透传命令SQE构造模块被配置为构造物理盘透传命令SQE；

[0045] 透传命令SQE发送模块，该透传命令SQE发送模块被配置为将所述物理盘透传命令SQE发送给所述RAID控制器；

[0046] 命令确认模块，该命令确认模块被配置为读取接收的所述物理盘透传命令SQE中的DW0中的透传命令操作码以确认所述物理盘透传命令SQE为物理盘透传命令；

[0047] 物理盘类型确认模块，该物理盘类型确认模块被配置为读取所述物理盘透传命令SQE中的DW2，获取物理盘ID，以确定目标物理盘的类型；

[0048] 物理盘命令SQE构造模块，该物理盘命令SQE被配置为构造模块基于所述目标物理盘的类型构造物理盘命令SQE；

[0049] 物理盘命令SQE处理模块，该物理盘命令SQE处理模块被配置为将所述物理盘命令SQE发送至所述目标物理盘，并等待所述目标物理盘返回物理盘CQE；以及

[0050] 透传命令CQE处理模块，该透传命令CQE处理模块被配置为以所述物理盘CQE为基础构造透传命令CQE，并返回所述透传命令CQE。

[0051] 本发明的再一方面，还提供了一种计算机可读存储介质，存储有计算机程序指令，该计算机程序指令被执行时实现上述任一项根据本发明的RAID控制器下物理盘管理的方法。

[0052] 本发明的又一方面，还提供了一种计算机设备，包括存储器和处理器，该存储器中存储有计算机程序，该计算机程序被处理器执行时执行上述任一项根据本发明的RAID控制器下物理盘管理的方法。

[0053] 本发明至少具有以下有益技术效果：本发明提出了一种SOCRAID控制器下物理盘管理模式，通过本发明提供的设计和方法步骤，可以实现将针对不同类型物理盘的任意命令封装在NVMe命令中，通过RAID控制器上针对每种物理盘的转发操作流程，可以完成主机下发的任意物理盘命令处理。命令的处理过程仅和物理盘的类型有关系，和命令无关。因此将免去了RAID控制器实现具体的物理盘命令的操作，从而简化了RAID控制器接口的设计，降低了开发难度。

## 附图说明

[0054] 为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本发明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的实施例。

[0055] 在图中：

[0056] 图1示出了根据本发明的RAID控制器下物理盘管理的方法的实施例的示意性框图；

[0057] 图2示出了根据本发明的RAID控制器下物理盘管理的方法中的物理盘透传命令格式SQE的基本格式；

[0058] 图3示出了在目标物理盘为NVMe盘时，根据本发明的RAID控制器下物理盘管理的方法中的物理盘透传命令格式SQE中命令参数的基本格式；

[0059] 图4示出了在目标物理盘为SAS盘时，根据本发明的RAID控制器下物理盘管理的方法中的物理盘透传命令格式SQE中命令参数的基本格式；

[0060] 图5示出了在目标物理盘为SATA盘时，根据本发明的RAID控制器下物理盘管理的方法中的物理盘透传命令格式SQE中命令参数的基本格式；

[0061] 图6示出了根据本发明的RAID控制器下物理盘管理的方法中的物理盘CQE的基本格式；

[0062] 图7示出了根据本发明的RAID控制器下物理盘管理的方法中从物理盘透传命令SQE到物理盘命令SQE的转换示意图；

[0063] 图8示出了根据本发明的RAID控制器下物理盘管理的方法中FIS在物理盘透传命令SQE中的DW10到DW14中的存放位置；

[0064] 图9示出了根据本发明的RAID控制器下物理盘管理的方法中CDB在物理盘透传命令SQE中的DW10到DW13中的存放位置；

[0065] 图10示出了根据本发明的RAID控制器下物理盘管理的系统的实施例的示意性框图；

[0066] 图11示出了根据本发明的实现RAID控制器下物理盘管理的方法的计算机可读存储介质的实施例的示意图；

[0067] 图12示出了根据本发明的实现RAID控制器下物理盘管理的方法的计算机设备的实施例的硬件结构示意图；

[0068] 图13示出了根据现有技术的RAID控制器与主机和物理盘的连接架构的示意图。

## 具体实施方式

[0069] 为使本发明的目的、技术方案和优点更加清楚明白,以下结合具体实施例,并参照附图,对本发明实施例进一步详细说明。

[0070] 需要说明的是,本发明实施例中所有使用“第一”和“第二”的表述均是为了区分两个相同名称的非相同的实体或者非相同的参量,可见“第一”“第二”仅为了表述的方便,不应理解为对本发明实施例的限定。此外,术语“包括”和“具有”以及他们的任何变形,意图在于覆盖不排他的包含,例如,包含了一系列步骤或单元的过程、方法、系统、产品或设备固有的其他步骤或单元。

[0071] RAID磁盘冗余阵列技术,即将多个磁盘设备组合起来,成为存储阵列组,目的为提升性能、增加冗余以保护数据。

[0072] RAID卡作实现RAID技术的设备通常实现为计算机系统上的PCIe设备,通过PCIe总线连接到主机。RAID卡上的核心部件为RAID控制器。如图13所示的现有技术的RAID控制器,与主机的接口遵循NVMe(一种基于PCIe的存储设备规范,用于高速固态硬盘SSD主机接口。)协议标准,并在NVMe主机接口上扩展了RAID管理命令,用于完成RAID管理相关的操作。

[0073] 本发明的目的是提供一种方法,在一种基于NVMe主机接口的RAID控制器管理接口上,实现一种通用的命令接口,可以使主机透过RAID控制器直接向RAID控制下的物理盘发送命令,从而降低RAID管理接口的复杂读,提升接口的灵活性。

[0074] 对此,本发明的第一方面,提供了一种RAID控制器下物理盘管理的方法100。图1示出了根据本发明的RAID控制器下物理盘管理的方法的实施例的示意性框图。在如图1所示的实施例中,该方法包括:

[0075] 步骤S10:构造物理盘透传命令SQE;

[0076] 步骤S20:向给所述RAID控制器发送所述物理盘透传命令SQE;

[0077] 步骤S30:读取接收的所述物理盘透传命令SQE中的DW0中的透传命令操作码以确认所述物理盘透传命令SQE为物理盘透传命令;

[0078] 步骤S40:读取所述物理盘透传命令SQE中的DW2,获取物理盘ID,以确定目标物理盘的类型;

[0079] 步骤S50:基于所述目标物理盘的类型构造物理盘命令SQE;

[0080] 步骤S60:将所述物理盘命令SQE发送至所述目标物理盘,并等待所述目标物理盘返回物理盘CQE;

[0081] 步骤S70:以所述物理盘CQE为基础构造透传命令CQE,并返回所述透传命令CQE。

[0082] SQE(Submission Queue Entry)为命令提交队列条目,每一个条目对应一个主机对NVMe设备的命令。CQE(Completion Queue Entry)为命令完成队列条目,每一个条目对应一个NVMe控制器上报的命令完成数据结构。NVMe控制器对主机下达的每个命令,完成后都会在CQ中上报CQE。

[0083] 物理盘透传命令用于主机透过RAID控制器直接向RAID控制器下的物理盘发送命令,包括命令格式和命令响应格式。

[0084] 物理盘透传命令格式为NVMe命令,其是一种称为SQE格式的数据结构,SQE长度为64个字节,分为16个双字(DW0~DW15),其中包含了命令的操作码和命令的参数。该数据格式基本式如图2所示。



[0085] 在图2中,

[0086] ①物理盘透传命令操作码属于NVMe标准所规定的厂商自定操作码(Combined Opcode)范围,其值为C0h到FFh中的任意一个;

[0087] ②命令标识符为主机构造命令使分配的一个标识符,用于关联命令和命令的响应;

[0088] ③物理盘标识符类型:由于RAID控制器下可能存在多个和多中物理盘,此处的标识符用于设置物理盘透传命令发送给哪个物理盘;

[0089] ④数据指针:如果在物理盘执行命令的过程中,有数据要从主机发送到物理盘,或者物理盘要返回数据给主机,此处的指针指向主机内存区域的一段缓冲区,用于存放这些数据;

[0090] ⑤命令参数:根据目标物理盘的类型——NVMe,SAS,SATA,此处的命令参数有不同的类型。以下分开介绍:

[0091] 如果目标物理盘为NVMe盘,则命令参数如图3所示,其中①物理盘命NVMe令操作码是要发送给NVMe物理盘的命令操作码;②NVMe命令参数,即①物理盘NVMe命令所对应的命令参数。

[0092] 如果目标物理盘为SAS盘,则命令参数如图4所示。其中CDB(command descriptor block)为一种用于封装SCSI协议命令的数据结构。SCSI(Small Computer System Interface)是一种通常用于硬盘接口的协议规范。

[0093] 如果目标物理盘为SATA盘,则命令参数如图5所示。其中FIS(Frame Information Structure)为一种用于封装SATA硬盘命令的数据结构。

[0094] 物理盘透传命令响应用于RAID控制器向主机返回命令的执行状态,主要用于指示命令是否执行成功。NVMe命令的响应是一种称为CQE的数据结构,长度为16个字节有4个双字节构成(DW0~DW3)。物理盘透传命令的响应格式如图6所示,其中:

[0095] ①命令标识符,来自于物理盘透传命令中的①命令标识符,表示该命令响应使针对哪个命令的响应;

[0096] ②命令状态码表示命令执行的成功与否;

[0097] ③是与命令结果状态相关的参数。

[0098] 在根据本发明的RAID控制器下物理盘管理的方法的一个优选实施例中,构造物理盘透传命令SQE包括以下步骤:

[0099] 将透传命令操作码以及主机命令标识符填入物理盘透传命令SQE中的DW0;

[0100] 将被透传命令中的物理盘ID以及被透传命令操作码分别填入物理盘透传命令SQE的DW2和DW3中;以及

[0101] 将被透传命令中的其他字段内容复制到物理盘透传命令SQE的对应字段。

[0102] 构造物理盘命令SQE包括以下步骤:

[0103] 将物理盘透传命令SQE的DW3中的被透传命令操作码填入物理盘命令SQE的DW0中;

[0104] 重新生成新的物理盘命令标识符并填入物理盘SQE的DW0中;以及

[0105] 将物理盘透传命令SQE中的其他字段内容复制到物理盘命令SQE中对应位置。

[0106] 构造透传命令CQE包括以下步骤:

[0107] 将物理盘透传命令SQE的DW0中的主机命令标识符填入透传命令CQE中的DW3中;

[0108] 将目标物理盘返回的物理盘CQE中的DW0和DW1复制到透传命令CQE中的对应位置；  
以及

[0109] 将目标物理盘返回的物理盘CQE中的DW3中的状态信息位复制到透传命令CQE中的对应位置。

[0110] 上述优选实施例对应于物理盘为NVMe盘时主机通过盘透传命令直接向物理盘发送命令的流程。具体地，该流程如下：

[0111] 1、主机构造物理盘透传命令SQE：

[0112] 将物理盘透传命令操作码、主机生成的主机命令标识符填入物理盘透传命令SQE的DW0，将物理盘ID、被透传命令操作码分别填入物理盘透传命令SQE的DW2和DW3中，将被透传命令的其他字段，DW1、DW4~DW15内容复制到物理盘透传命令SQE对应的字段；

[0113] 2、主机将构造好的物理盘透传命令SQE发送给RAID控制器；

[0114] 3、RAID控制器接收到物理盘透传命令SQE，读取命令DW0中的操作码，判断出这是一个物理盘透传命令；

[0115] 4、RAID控制器读取物理盘透传命令中的DW2，获取物理盘ID，并根据此ID判断出目标物理盘是一个NVMe盘；

[0116] 5、RAID控制器构造发送给物理盘的物理盘命令SQE：

[0117] 从物理盘透传命令SQE的DW3中取出被透传命令的操作码，填入物理盘命令SQE的DW0中，RAID控制器重新生成新的物理盘命令标识符，填入物理盘命令SQE的DW0中，从物理盘透传命令SQE中取出DW1、DW4~DW15，复制到物理命令SQE中对应位置，图7示出从物理盘透传命令SQE到物理盘命令SQE的转换；

[0118] 6、将物理盘命令SQE发送给物理盘，并等待物理盘返回物理盘CQE；

[0119] 7、RAID控制器构造透传命令CQE，并返回给主机：

[0120] 从物理盘透传命令SQE DW0中取出透传命令标识符，填入CQE中DW3中，将返回的物理盘CQE中的DW0、DW1复制到透传命令CQE中的对应位置，将返回的物理盘CQE中的DW3中的状态信息位复制到透传命令CQE中的对应位置。

[0121] 另外，在根据本发明的RAID控制器下物理盘管理的方法的另一优选实施例中，

[0122] 构造物理盘透传命令SQE包括以下步骤：

[0123] 将透传命令操作码以及主机命令标识符填入物理盘透传命令SQE中的DW0；

[0124] 将被透传命令中的物理盘ID以及被透传命令操作码分别填入物理盘透传命令SQE的DW2和DW3中；以及

[0125] 将被透传命令中的FIS帧填入物理盘透传命令SQE DW10至DW14或者将被透传命令中的CDB帧填入物理盘透传命令SQE中的DW10至DW13。

[0126] 在被透传命令中存在数据传输阶段时，根据NVMe标准构造PRP数据结构并填入物理盘透传命令SQE的DW6至DW9字段。

[0127] 在确定目标物理盘的类型为SATA盘时，构造物理盘命令SQE包括从物理盘透传命令SQE的DW10至DW14中取出FIS并填入目标物理盘中的SATA控制器，或者在确定目标物理盘的类型为SAS盘时，构造物理盘命令SQE包括从物理盘透传命令SQE的DW10至DW13中取出CDB并填入目标物理盘中的SAS控制器。

[0128] 构造物理盘命令SQE还包括以下步骤：

[0129] 从物理盘透传命令SQE的DW6至DW9中取出PRP数据结构,并按照SATA控制器要求构造SATA数据缓冲描述符结构或者按照SAS控制器要求构造SAS数据缓冲描述符结构。

[0130] 方法还包括在目标物理盘接收到物理盘命令SQE之后,操作SATA控制器或者SAS控制器将物理盘命令SQE发送给目标物理盘。

[0131] 在被透传命令中存在数据传输阶段时,由SATA控制器或者SAS控制器完成主机的数据缓冲区和目标物理盘之间的数据搬移。

[0132] 构造透传命令CQE包括以下步骤:

[0133] 将SATA控制器或者SAS控制器返回的物理盘CQE中的命令执行状态填入透传命令CQE中DW3中的状态;

[0134] 将返回的物理盘CQE中的状态数据填入到透传命令CQE的DW0中;以及

[0135] 将返回的物理盘CQE中的DW3中的状态信息位复制到透传命令CQE中的对应位置。

[0136] 上述优选实施例对应于物理盘为SATA盘或SAS盘时主机通过盘透传命令直接向物理盘发送命令的流程。

[0137] 具体地,物理盘为SATA盘时主机通过物理盘透传命令直接向物理盘发送命令的流程如下:

[0138] 1、主机构造物理盘透传命令SQE:

[0139] 将物理盘透传命令操作码、主机生成的命令标识符填入物理盘透传命令SQE的DW0,将物理盘ID、被透传命令操作码分别填入物理盘透传命令SQE的DW2和DW3中,将SATA命令FIS帧,填入SQE DW10~DW14。FIS长度位20个字节,FIS在CQE的DW10到DW14中的存放位置如图8所示,如果命令存在数据传输阶段,则根据NVMe标准构造PRP数据结构,填入物理盘透传命令SQE的DW6~DW9字段;

[0140] 2、主机将构造好的物理盘透传命令SQE发送给RAID控制器;

[0141] 3、RAID控制器接收到物理盘透传命令SQE,读取物理盘透传命令的DW0中的操作码,判断出这是一个物理盘透传命令;

[0142] 4、RAID控制器读取物理盘透传命令中的DW2,获取物理盘ID。并根据此ID判断出目标物理盘是一个SATA盘;

[0143] 5、RAID控制器构造物理盘命令SQE发送给SATA物理盘:

[0144] 从物理盘透传命令SQE的DW10~DW14中取出FIS,填入SATA控制器,如果存在数据传输阶段,则从物理盘透传命令SQE的DW6~DW9中取出PRP数据结构,并按照SATA控制器要求构造SATA数据缓冲描述符结构;

[0145] 6、操作SATA控制器,将命令发送给SATA盘。如果存在数据传输阶段,由SATA控制器完成主机数据缓冲区和盘之间的数据搬移;

[0146] 7、RAID控制器构造透传命令CQE,并返回给主机:

[0147] 将SATA控制器返回的命令执行状态填入透传命令CQE中DW3中的状态,如果SATA控制器返回有详细的状态数据,将数据填入到透传命令CQE的DW0中,将返回的物理盘CQE中的DW3中的状态信息位复制到透传命令CQE中的对应位置。

[0148] 具体地,物理盘为SAS盘时主机通过物理盘透传命令直接向物理盘发送命令的流程如下:

[0149] 1、主机构造物理盘透传命令SQE:

[0150] 将物理盘透传命令操作码、主机生成的命令标识符填入物理盘透传命令SQE的DW0,将物理盘ID、被透传命令操作码分别填入物理盘透传命令SQE的DW2和DW3中,将SCSI命令CDB帧,填入SQE DW10~DW13。CDB长度位最多16个字节,CDB在CQE的DW10到DW13中的存放位置如图9所示,如果命令存在数据传输阶段,则根据NVMe标准构造PRP数据结构,填入物理盘透传命令SQE的DW6~DW9字段;

[0151] 2、主机将构造好的物理盘透传命令SQE发送给RAID控制器;

[0152] 3、RAID控制器接收到物理盘透传命令SQE,读取物理盘透传命令的DW0中的操作码,判断出这是一个物理盘透传命令;

[0153] 4、RAID控制器读取物理盘透传命令中的DW2,获取物理盘ID,并根据此ID判断出目标物理盘是一个SAS盘;

[0154] 5、RAID控制器构造物理盘命令SQE发送给SAS物理盘:

[0155] 从透传命令SQE的DW10~DW13中取出CDB,填入SAS控制器, 如果存在数据传输阶段,则从透传命令SQE的DW6~DW9中取出PRP数据结构,并按照SAS控制器要求构造SAS数据缓冲描述符结构;

[0156] 6、操作SAS控制器,将物理盘命令SQE发送给SAS盘。如果存在数据传输阶段,由SAS控制器完成主机数据缓冲区和盘之间的数据搬移;

[0157] 7、RAID控制器构造透传命令CQE,并返回给主机:

[0158] 将SAS控制器返回的命令执行状态填入透传命令CQE中DW3中的状态,如果SAS控制器返回有详细的状态数据,将数据填入到透传命令CQE的DW0中,将返回的物理盘CQE中的DW3中的状态信息位复制到透传命令CQE中的对应位置。

[0159] 本发明的第二方面,还提供了一种RAID控制器下物理盘管理的系统200。图10示出了根据本发明的RAID控制器下物理盘管理的系统200的实施例的示意性框图。如图10所示,该系统包括:

[0160] 透传命令SQE构造模块210,其被配置为构造物理盘透传命令SQE;

[0161] 透传命令SQE发送模块220,其被配置为将物理盘透传命令SQE发送给RAID控制器;

[0162] 命令确认模块230,其被配置为读取接收的物理盘透传命令SQE中的DW0中的透传命令操作码以确认物理盘透传命令SQE为物理盘透传命令;

[0163] 物理盘类型确认模块240,其被配置为读取物理盘透传命令SQE中的DW2,获取物理盘ID,以确定目标物理盘的类型;

[0164] 物理盘命令SQE构造模块250,其被配置为构造模块基于目标物理盘的类型构造物理盘命令SQE;

[0165] 物理盘命令SQE处理模块260,其被配置为将物理盘命令SQE发送至目标物理盘,并等待目标物理盘返回物理盘CQE;以及

[0166] 透传命令CQE处理模块270,其被配置为以物理盘CQE为基础构造透传命令CQE,并返回透传命令CQE。

[0167] 本发明实施例的第三个方面,还提供了一种计算机可读存储介质,图11示出了根据本发明实施例提供的RAID控制器下物理盘管理的方法的计算机可读存储介质的示意图。如图11所示,计算机可读存储介质300存储有计算机程序指令310,该计算机程序指令310可以被处理器执行。该计算机程序指令310被执行时实现上述任意一项实施例的方法。

[0168] 应当理解,在相互不冲突的情况下,以上针对根据本发明的RAID控制器下物理盘管理的方法阐述的所有实施方式、特征和优势同样地适用于根据本发明的RAID控制器下物理盘管理的系统和存储介质。

[0169] 本发明实施例的第四个方面,还提供了一种计算机设备400,包括存储器420和处理器410,该存储器中存储有计算机程序,该计算机程序被该处理器执行时实现上述任意一项实施例的方法。

[0170] 如图12所示,为本发明提供的执行RAID控制器下物理盘管理的方法的计算机设备的一个实施例的硬件结构示意图。以如图12所示的计算机设备400为例,在该计算机设备中包括一个处理器410以及一个存储器420,并还可以包括:输入装置430和输出装置440。处理器410、存储器420、输入装置430和输出装置440可以通过总线或者其他方式连接,图6中以通过总线连接为例。输入装置430可接收输入的数字或字符信息,以及产生与RAID控制器下物理盘管理的有关的信号输入。输出装置440可包括显示屏等显示设备。

[0171] 存储器420作为一种非易失性计算机可读存储介质,可用于存储非易失性软件程序、非易失性计算机可执行程序以及模块,如本申请实施例中的资源监控方法对应的程序指令/模块。存储器420可以包括存储程序区和存储数据区,其中,存储程序区可存储操作系统、至少一个功能所需要的应用程序;存储数据区可存储资源监控方法的使用所创建的数据等。此外,存储器420可以包括高速随机存取存储器,还可以包括非易失性存储器,例如至少一个磁盘存储器件、闪存器件、或其他非易失性固态存储器件。在一些实施例中,存储器420可选包括相对于处理器410远程设置的存储器,这些远程存储器可以通过网络连接至本地模块。上述网络的实例包括但不限于互联网、企业内部网、局域网、移动通信网及其组合。

[0172] 处理器410通过运行存储在存储器420中的非易失性软件程序、指令以及模块,从而执行服务器的各种功能应用以及数据处理,即实现上述方法实施例的资源监控方法。

[0173] 本领域技术人员还将明白的是,结合这里的公开所描述的各种示例性逻辑块、模块、电路和算法步骤可以被实现为电子硬件、计算机软件或两者的组合。为了清楚地说明硬件和软件的这种可互换性,已经就各种示意性组件、方块、模块、电路和步骤的功能对其进行了一般性的描述。这种功能是被实现为软件还是被实现为硬件取决于具体应用以及施加给整个系统的设计约束。本领域技术人员可以针对每种具体应用以各种方式来实现的功能,但是这种实现决定不应被解释为导致脱离本发明实施例公开的范围。

[0174] 最后需要说明的是,本文的计算机可读存储介质(例如,存储器)可以是易失性存储器或非易失性存储器,或者可以包括易失性存储器和非易失性存储器两者。作为例子而非限制性的,非易失性存储器可以包括只读存储器(ROM)、可编程ROM(PROM)、电可编程ROM(EPROM)、电可擦写可编程ROM(EEPROM)或快闪存储器。易失性存储器可以包括随机存取存储器(RAM),该RAM可以充当外部高速缓存存储器。作为例子而非限制性的,RAM可以以多种形式获得,比如同步RAM(DRAM)、动态RAM(DRAM)、同步DRAM(SDRAM)、双数据速率SDRAM(DDR SDRAM)、增强SDRAM(ESDRAM)、同步链路DRAM(SLDRAM)、以及直接Rambus RAM(DRRAM)。所公开的方面的存储设备意在包括但不限于这些和其它合适类型的存储器。

[0175] 结合这里的公开所描述的各种示例性逻辑块、模块和电路可以利用被设计成用于执行这里功能的下列部件来实现或执行:通用处理器、数字信号处理器(DSP)、专用集成电路(ASIC)、现场可编程门阵列(FPGA)或其它可编程逻辑器件、分立门或晶体管逻辑、分立的

硬件组件或者这些部件的任何组合。通用处理器可以是微处理器,但是可替换地,处理器可以是任何传统处理器、控制器、微控制器或状态机。处理器也可以被实现为计算设备的组合,例如,DSP和微处理器的组合、多个微处理器、一个或多个微处理器结合DSP和/或任何其它这种配置。

[0176] 以上是本发明公开的示例性实施例,但是应当注意,在不背离权利要求限定的本发明实施例公开的范围的前提下,可以进行多种改变和修改。根据这里描述的公开实施例的方法权利要求的功能、步骤和/或动作不需以任何特定顺序执行。此外,尽管本发明实施例公开的元素可以以个体形式描述或要求,但除非明确限制为单数,也可以理解为多个。

[0177] 应当理解的是,在本文中使用的,除非上下文清楚地支持例外情况,单数形式“一个”旨在也包括复数形式。还应当理解的是,在本文中使用的“和/或”是指包括一个或者一个以上相关联地列出的项目的任意和所有可能组合。上述本发明实施例公开实施例序号仅仅为了描述,不代表实施例的优劣。

[0178] 所属领域的普通技术人员应当理解:以上任何实施例的讨论仅为示例性的,并非旨在暗示本发明实施例公开的范围(包括权利要求)被限于这些例子;在本发明实施例的思路下,以上实施例或者不同实施例中的技术特征之间也可以进行组合,并存在如上的本发明实施例的不同方面的许多其它变化,为了简明它们没有在细节中提供。因此,凡在本发明实施例的精神和原则之内,所做的任何省略、修改、等同替换、改进等,均应包含在本发明实施例的保护范围之内。

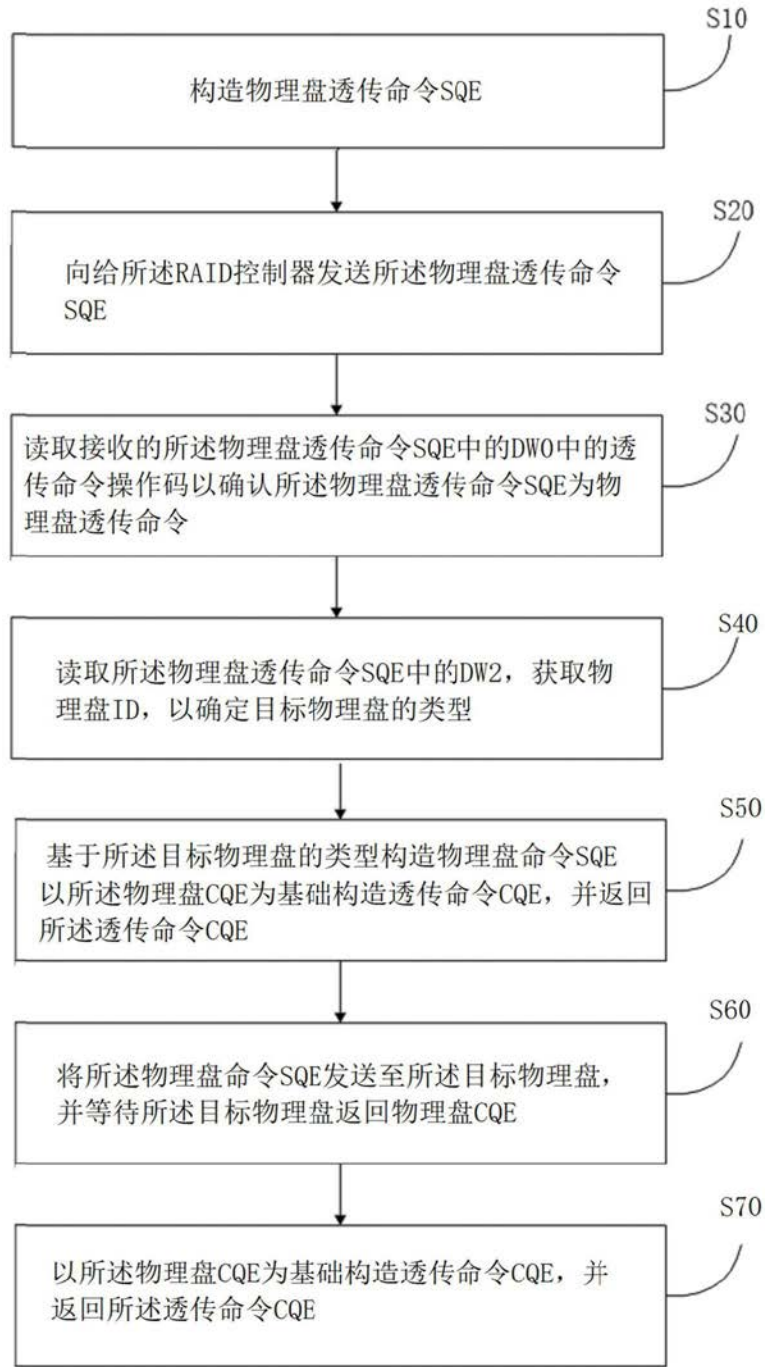


图1



图2



图3

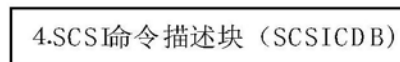


图4

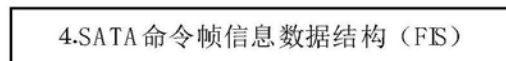


图5

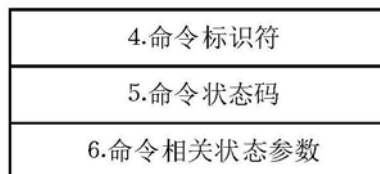


图6





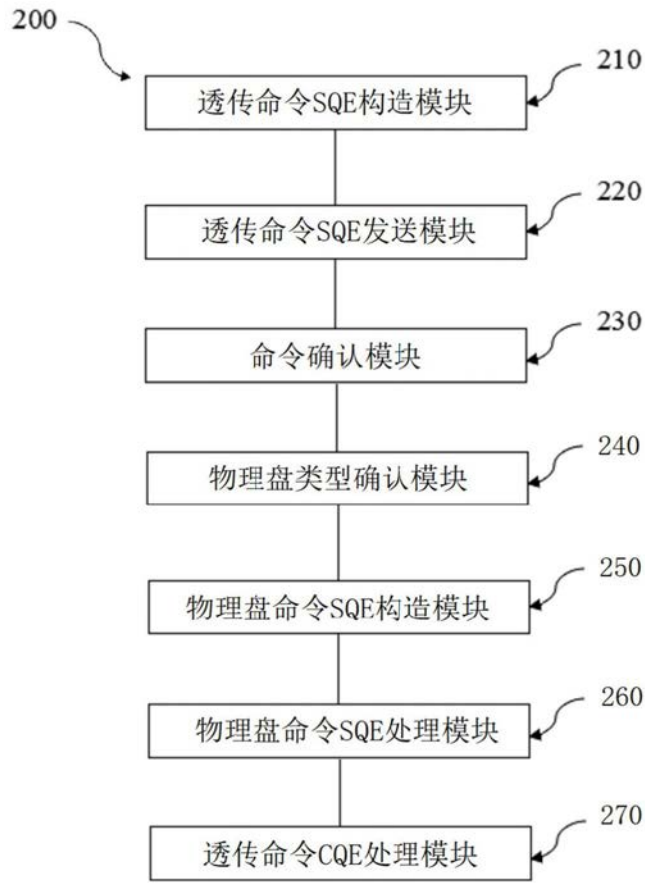


图10

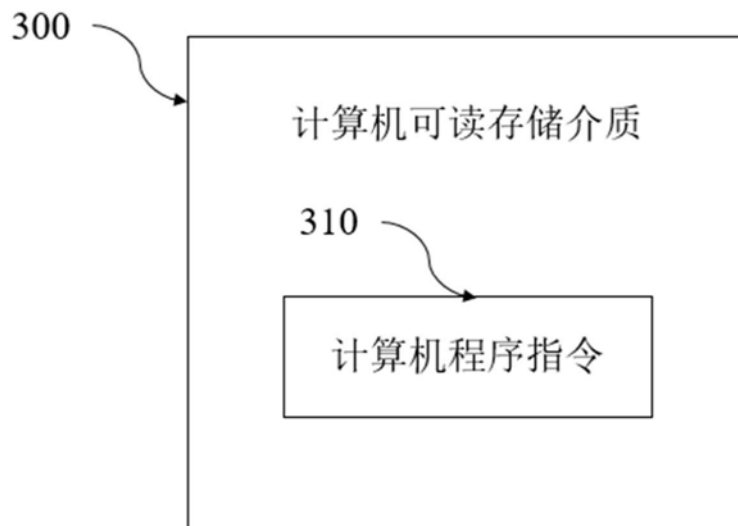


图11

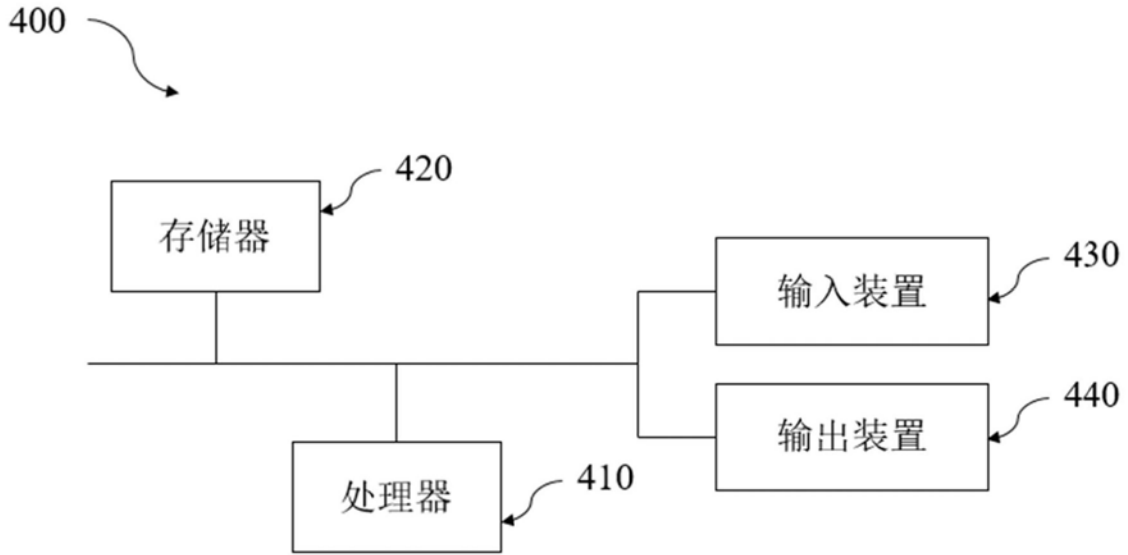


图12

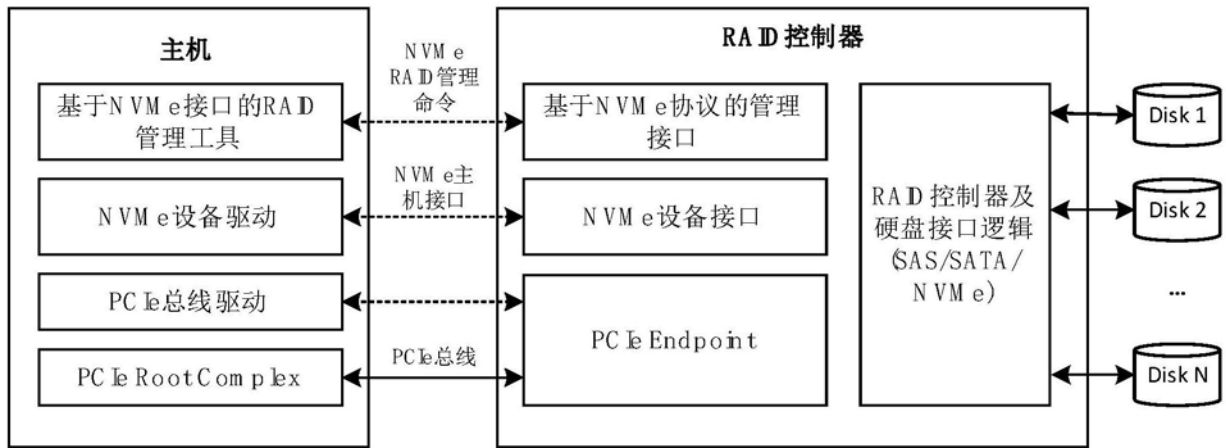


图13