



(12) 发明专利申请

(10) 申请公布号 CN 104202359 A

(43) 申请公布日 2014. 12. 10

(21) 申请号 201410393925. 7

(22) 申请日 2014. 08. 12

(71) 申请人 浪潮电子信息产业股份有限公司
地址 250101 山东省济南市高新区舜雅路
1036 号

(72) 发明人 范文洋

(74) 专利代理机构 济南信达专利事务有限公
司 37100

代理人 姜明

(51) Int. Cl.

H04L 29/08(2006. 01)

G06F 3/06(2006. 01)

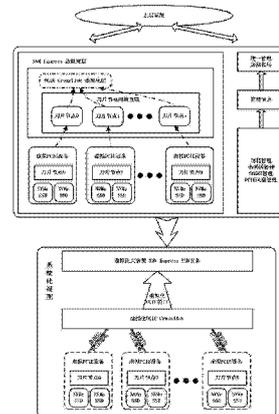
权利要求书1页 说明书3页 附图2页

(54) 发明名称

一种基于刀片服务器的NVMe SSD虚拟化设计方法

(57) 摘要

本发明提供一种基于刀片服务器的NVMeSSD虚拟化设计方法,包括刀片服务器,该刀片服务器内置若干刀片计算节点、高速网络交换节点、PCIEIO 拓展模块、管理模块;基于该刀片服务器的虚拟化的详细映射过程为:设置底层虚拟 PCIE 设备层,该层包括若干刀片节点将本节点的若干个支持NVMe 标准的SSD 闪存作为PCIE 终端设备;设置中间层虚拟化总线桥接层;设置顶层 NVMe 虚拟化层,将底层所有设备整合为单一 NVMe 设备。该一种基于刀片服务器的NVMeSSD 虚拟化设计方法和现有技术相比,有效提高刀片服务器拓展的灵活性,可靠性高,实用性强,适用范围广泛。



1. 一种基于有限状态机的存储虚拟化管理系统,其特征在于包括刀片服务器,该刀片服务器内置若干刀片计算节点、高速网络交换节点、PCIE IO拓展模块、管理模块;其中PCIE IO拓展模块用于拓展NVMe SSD,各个刀片计算节点的PCIE IO拓展模块则由各个刀片节点独立提供;基于该刀片服务器的虚拟化的详细映射过程为:

设置虚拟PCIE设备层,该虚拟PCIE设备层设置在底层,该层包括若干刀片节点,所述刀片节点将本节点的若干个支持NVMe标准的SSD闪存作为PCIE终端设备;

设置虚拟化总线桥接层,该总线桥接层设置在中间层,通过刀片服务器的高速通信网络虚拟PCIE Crosslink设备,即总线桥设备,该层桥接顶层的NVMe虚拟化层以及底层的各个刀片节点的NVMe SSD闪存;

设置NVMe虚拟化层,该NVMe虚拟化层设置在顶层,该层将底层所有设备整合为单一NVMe设备,并为上层系统提供统一的NVMe标准的接口。

2. 根据权利要求1所述的一种基于有限状态机的存储虚拟化管理系统,其特征在于:所述NVMe SSD闪存均为PCIE接口且每个节点的NVMe SSD闪存都虚拟映射到PCIE Crosslink上,该NVMe SSD闪存分布安装在各个刀片节点的PCIE槽位上。

3. 根据权利要求1所述的一种基于有限状态机的存储虚拟化管理系统,其特征在于:所述虚拟化总线桥层的下行端接各个节点的NVMe SSD闪存的PCIE终端设备,上行端为NVMe虚拟化层提供PCIE上行接口。

4. 根据权利要求1所述的一种基于有限状态机的存储虚拟化管理系统,其特征在于:所述NVMe虚拟化层由操作系统直接操作NVMe SSD闪存的软件接口层,这里的软件接口层包括NVMe SSD闪存的虚拟化层和NVMe SSD闪存的管理虚拟化层;该NVMe虚拟化层将各节点分布式分布的NVMe SSD设备映射为单一的NVMe SSD闪存。

一种基于刀片服务器的 NVMe SSD 虚拟化设计方法

技术领域

[0001] 本发明涉及计算机存储虚拟化技术领域,具体的说是一种实用性强、基于刀片服务器的 NVMe SSD 虚拟化设计方法。

背景技术

[0002] 随着大数据时代的来临,对数据的处理能力以及响应速度要求越来越高,刀片服务器广泛运用在云计算与大数据处理领域,刀片服务器可以很好的完成多节点刀片数据处理能力的统一,但是 IO 访问性能由于外围存储速度受限,不能有效的提高。固态存储 SSD 可以极大的提高服务器存储的 IO 处理速度并降低 IO 延迟,NVMe SSD 采用 PCIE 接口,更加靠近处理器 CPU,凭借 PCIE 接口访问的高带宽低延迟优势,更好的发挥了固态存储 SSD 的性能。

[0003] 刀片服务器刀片节点的 PCIE 槽位是独立的,NVMe SSD 应用在刀片服务器上时,如果按照刀片服务器的设计特点,用刀片节点 PCIE 槽位分布式的方式布置 NVMe SSD 的方式,不能体现刀片服务器计算资源集中的优势。

[0004] 而存储虚拟化通过将一(或多个)目标(Target)服务或功能与其它附加的功能集成,统一提供有用的全面功能服务。可以有效的对存储资源进行管理,存储虚拟化技术可以屏蔽异构的存储系统并将其整合为一个统一的存储系统供用户使用。同时,存储虚拟化技术有可以对存储资源进行再分配,将存储资源划分给不同的用户使用。为了方便用户对存储虚拟化系统的管理,目前存储虚拟化系统的实现普遍采用 BS 架构,也即通过浏览器应用程序实现对存储虚拟化系统的管理。

[0005] 因而,如果将存储虚拟化技术应用在计算机服务器存储方面,将有效提高服务器的存储速度,基于此,现提供一种刀片服务器上 NVMe SSD 的虚拟化方法,该方法通过整合刀片节点中分布式安装的 NVMe SSD 资源和管理,通过虚拟化 PCIE Crosslink 桥接各个节点的 NVMe SSD 与虚拟化 NVMe 层,虚拟化 NVMe 层采用 NVMe 接口标准,为系统提供统一接口。

发明内容

[0006] 本发明的技术任务是解决现有技术的不足,提供一种实用性强、基于刀片服务器的 NVMe SSD 虚拟化设计方法。

[0007] 本发明的技术方案是按以下方式实现的,该一种基于刀片服务器的 NVMe SSD 虚拟化设计方法,包括刀片服务器,该刀片服务器内置若干刀片计算节点、高速网络交换节点、PCIE IO 拓展模块、管理模块;其中 PCIE IO 拓展模块用于拓展 NVMe SSD,各个刀片计算节点的 PCIE IO 拓展模块则由各个刀片节点独立提供;基于该刀片服务器的虚拟化的详细映射过程为:

设置虚拟 PCIE 设备层,该虚拟 PCIE 设备层设置在底层,该层包括若干刀片节点,所述刀片节点将本节点的若干个支持 NVMe 标准的 SSD 闪存作为 PCIE 终端设备;

设置虚拟化总线桥接层,该总线桥接层设置在中间层,通过刀片服务器的高速通信网

络虚拟 PCIE Crosslink 设备,即总线桥设备,该层桥接顶层的 NVMe 虚拟化层以及底层的各个刀片节点的 NVMe SSD 闪存;

设置 NVMe 虚拟化层,该 NVMe 虚拟化层设置在顶层,该层将底层所有设备整合为单一 NVMe 设备,并为上层系统提供统一的 NVMe 标准的接口。

[0008] 所述 NVMe SSD 闪存均为 PCIE 接口且每个节点的 NVMe SSD 闪存都虚拟映射到 PCIE Crosslink 上,该 NVMe SSD 闪存分布安装在各个刀片节点的 PCIE 槽位上。

[0009] 所述虚拟化总线桥层的下行端接各个节点的 NVMe SSD 闪存的 PCIE 终端设备,上行端为 NVMe 虚拟化层提供 PCIE 上行接口。

[0010] 所述 NVMe 虚拟化层由操作系统直接操作 NVMe SSD 闪存的软件接口层,这里的软件接口层包括 NVMe SSD 闪存的虚拟化层和 NVMe SSD 闪存的管理虚拟化层;该 NVMe 虚拟化层将各节点分布式分布的 NVMe SSD 设备映射为单一的 NVMe SSD 闪存。

[0011] 本发明与现有技术相比所产生的有益效果是:

本发明的一种基于刀片服务器的 NVMe SSD 虚拟化设计方法为刀片服务器的 SSD 应用提供了高效可靠的应用模型,整体提高刀片服务器的运算以及存储 IO 性能;通过虚拟化 PCIE Crosslink 将各个节点的 NVMe SSD 统一映射到上层系统,方便了 NVMe SSD 在刀片服务器上的拓展以及动态调整;实现软件可编程实时控制,便捷管理热插拔,功耗以及闪存的寿命管理等;数据可以动态灵活的在各个节点 NVMe SSD 盘位上动态迁移,并整合各个节点的 SSD 进行统一使用,并且对热插拔的支持,可以动态调整系统的 SSD 容量,极大的提高了易用灵活性;NVMe 虚拟化层可以灵活配置 RAID 功能,提高数据使用的可靠性以及整体并行 IO 访问的性能;通过虚拟化管理接口,管理闪存的寿命,可以动态更换添加以及删除硬盘,有效的提高设备的可靠性;实用性强,适用范围广泛,易于推广。

附图说明

[0012] 附图 1 是本发明的 NVMe SSD 虚拟化系统拓扑示意图。

[0013] 附图 2 是本发明的虚拟化层结构示意图。

具体实施方式

[0014] 下面结合附图对本发明的一种基于刀片服务器的 NVMe SSD 虚拟化设计方法作以下详细说明。

[0015] 本发明提供一种基于刀片服务器的 NVMe SSD 虚拟化设计方法,该方法整合刀片节点中分布式安装的 NVMe SSD 资源和管理。通过虚拟化 PCIE Crosslink 桥接各个节点的 NVMe SSD 与虚拟化 NVMe 层,虚拟化 NVMe 层采用 NVMe 接口标准,为系统提供统一接口。基于此设计思路,如附图 1 所示,该附图整体展示了刀片服务器中 NVMe SSD 虚拟化的系统构成的拓扑关系。对于上层系统而言,直接面向 NVMe 虚拟化层,该层隔离了系统和底层设备之间的直接联系,将刀片节点,高速网络互联以及 NVMe SSD 设备虚拟化成整体的 NVMe SSD 供系统使用。

[0016] 结合附图 2,本发明的方法中包括刀片服务器,该刀片服务器内置若干刀片计算节点、高速网络交换节点、PCIE IO 拓展模块、管理模块;其中 PCIE IO 拓展模块用于拓展 NVMe SSD,各个刀片计算节点的 PCIE IO 拓展模块则由各个刀片节点独立提供;基于该刀片服务

器的虚拟化的详细映射过程为：

设置虚拟 PCIE 设备层,该虚拟 PCIE 设备层设置在底层,该层包括若干刀片节点,所述刀片节点将本节点的若干个支持 NVMe 标准的 SSD 闪存作为 PCIE 终端设备；

设置虚拟化总线桥接层,该总线桥接层设置在中间层,通过刀片服务器的高速通信网络虚拟 PCIE CrossLink 设备,即总线桥设备,该层桥接顶层的 NVMe 虚拟化层以及底层的各个刀片节点的 NVMe SSD 闪存；

设置 NVMe 虚拟化层,该 NVMe 虚拟化层设置在顶层,该层将底层所有设备整合为单一 NVMe 设备,并为上层系统提供统一的 NVMe 标准的接口。

[0017] 所述 NVMe SSD 闪存均为 PCIE 接口且每个节点的 NVMe SSD 闪存都虚拟映射到 PCIE Crosslink 上,该 NVMe SSD 闪存分布安装在各个刀片节点的 PCIE 槽位上。

[0018] 所述虚拟化总线桥层的下行端接各个节点的 NVMe SSD 的 PCIE 终端设备,上行端为 NVMe 虚拟化层提供 PCIE 上行接口。

[0019] 所述 NVMe 虚拟化层由操作系统直接操作 NVMe SSD 闪存的软件接口层,这里的软件接口层包括 NVMe SSD 闪存的虚拟化层和 NVMe SSD 闪存的管理虚拟化层。通过 NVMe 虚拟化层,各节点分布式分布的 NVMe SSD 设备,映射为单一的 NVMe SSD。通过虚拟化的方案,数据可以动态灵活的在各个节点 NVMe SSD 盘位上动态迁移,并整合各个节点的 SSD 进行统一使用,系统对热插拔的支持,可以动态调整系统的 SSD 容量,极大的提高了易用灵活性。NVMe 虚拟化层还可以灵活配置 RAID 功能,提高数据使用的可靠性以及整体并行 IO 访问的性能。NVMe 的管理接口的虚拟化,实现软件可编程实时控制,便捷管理热插拔,功耗以及闪存的寿命管理等。热插拔的管理也可以配合 NVMe SSD 虚拟化层,有效地提高刀片服务器拓展的灵活性。

[0020] 本发明通过虚拟化方案,有效地统一刀片服务器中各刀片中节点分布式分布的遵循 NVMe 标准设计的 PCIE SSD。刀片服务器各节点通过高速网络互联,具有计算能力集中统一的特点,各个节点中的 NVMe SSD,通过虚拟化技术,统一映射到系统中,被识别成大容量的 NVMe 设备,并可以动态迁移各个刀片节点数据,在热插拔基础上可动态调整整个系统的 SSD 容量,通过这一虚拟化技术既可以提高 NVMe SSD 在刀片服务器中的易用性,极大的提高刀片服务器的性能,也可以拓展刀片服务器中 NVMe SSD 的使用的纵向拓展能力、拓展的灵活性、高效性以及可靠性。同时管理接口的虚拟化,实现软件可编程实时控制,便捷管理热插拔,功耗以及闪存的寿命管理等。

[0021] 以上实施方式仅用于说明本发明,而并非对本发明的限制,有关技术领域的普通技术人员,在不脱离本发明的精神和范围的情况下,还可以做出各种变化和变型,因此所有等同的技术方案也属于本发明的范畴,本发明的专利保护范围应由权利要求限定。

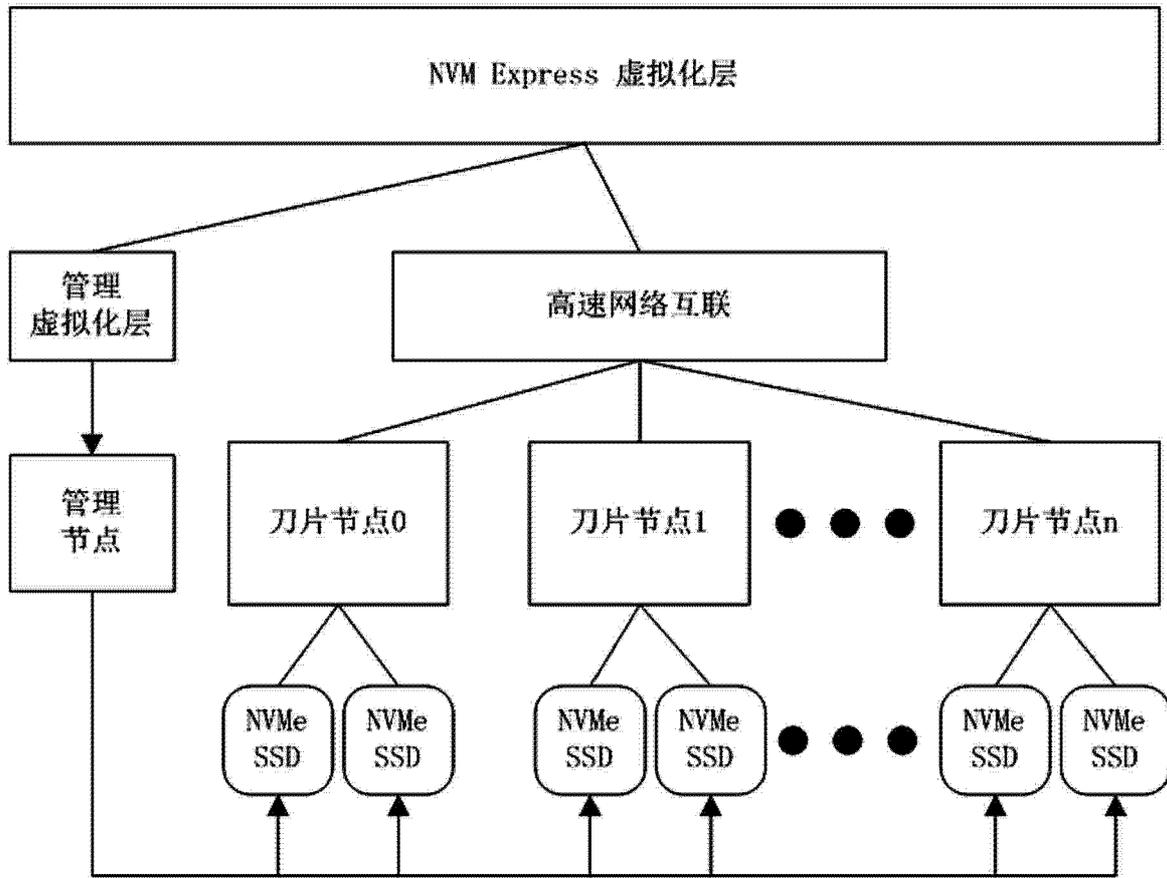


图 1

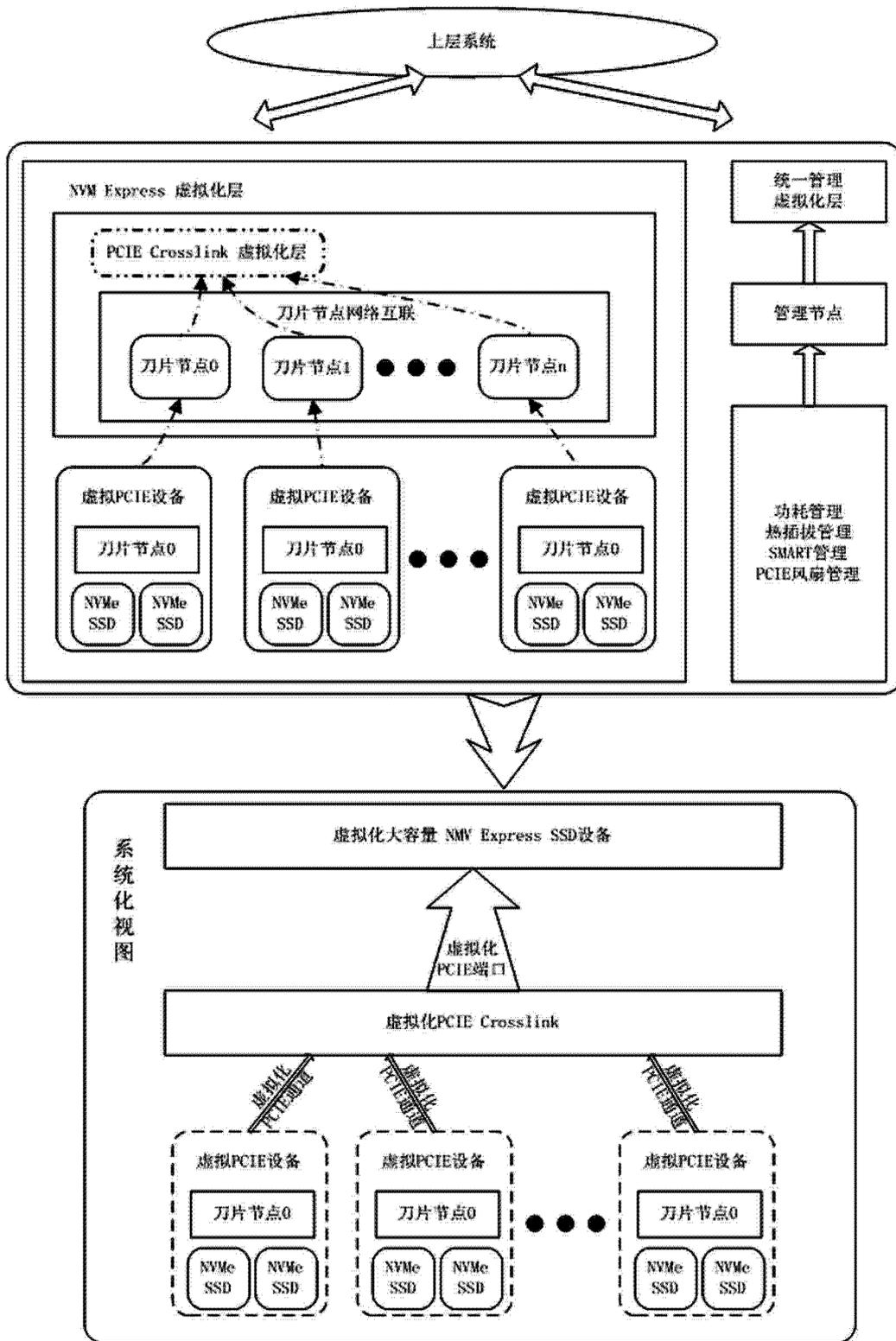


图 2