



(12)发明专利申请

(10)申请公布号 CN 106161169 A

(43)申请公布日 2016. 11. 23

(21)申请号 201610872839.3

(22)申请日 2016.09.30

(71)申请人 郑州云海信息技术有限公司

地址 450018 河南省郑州市郑东新区心怡
路278号16层1601室

(72)发明人 周亮 薛广营

(74)专利代理机构 北京集佳知识产权代理有限
公司 11227

代理人 罗满

(51)Int.Cl.

H04L 12/40(2006.01)

H04Q 11/00(2006.01)

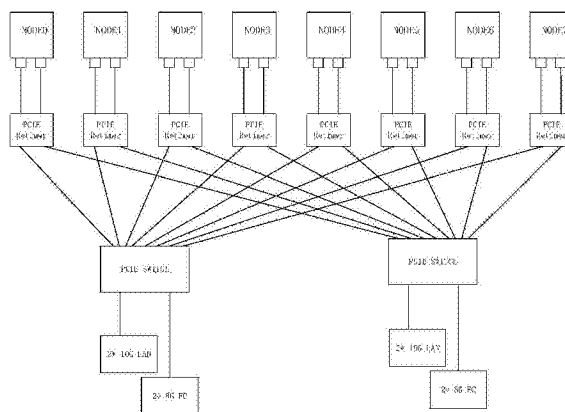
权利要求书1页 说明书4页 附图2页

(54)发明名称

一种多主机网络交换系统

(57)摘要

本发明公开了一种多主机网络交换系统,包括若干个计算节点、若干个PCIE SWITCH模块以及PCIE卡组;每个计算节点分别通过PCIE总线与各个PCIE SWITCH模块的上行端口相连;每个PCIE SWITCH模块的下行端口均通过PCIE总线连接一组PCIE卡组;PCIE卡组内的PCIE卡分别与对应的网络设备连接;PCIE SWITCH模块用于将通过对应的PCIE卡组连接的网络设备虚拟化共享至各个计算节点。本发明中将PCIE卡组独立于计算节点之外,提高了PCIE卡的配置灵活性,同时成本低且便于管理。



1. 一种多主机网络交换系统,其特征在于,包括若干个计算节点、若干个PCIE SWITCH模块以及PCIE卡组;

每个所述计算节点分别通过PCIE总线与各个PCIE SWITCH模块的上行端口相连;

每个所述PCIE SWITCH模块的下行端口均通过PCIE总线连接一组所述PCIE卡组;所述PCIE卡组内的PCIE卡分别与对应的网络设备连接;

所述PCIE SWITCH模块用于将通过对应的PCIE卡组连接的网络设备虚拟化共享至各个所述计算节点。

2. 根据权利要求1所述的系统,其特征在于,包括两个所述PCIE SWITCH模块。

3. 根据权利要求1所述的系统,其特征在于,还包括:

分别与各个所述计算节点一一对应的若干个PCIE retimer模块,每个所述PCIE retimer模块一端通过PCIE总线与对应的计算节点相连,另一端分别通过PCIE总线连接各个所述PCIE SWITCH模块的上行端口;

所述PCIE retimer模块用于对所述计算节点发送的PCIE信号进行增强处理。

4. 根据权利要求1-3任一项所述的系统,其特征在于,所述PCIE卡组包括若干个以太网卡以及若干个光纤网卡;

相应的,与所述以太网卡对应的网络设备为以太网交换机;与所述光纤网卡对应的网络设备为光纤交换机。

5. 根据权利要求4所述的系统,其特征在于,所述PCIE总线具体为PCIE3.0x8总线。

6. 根据权利要求3所述的系统,其特征在于,所述计算节点、所述PCIE retimer模块以及所述PCIE SWITCH模块设置于服务器的主板上,所述PCIE卡组设置于子板上。

一种多主机网络交换系统

技术领域

[0001] 本发明涉及服务器网络连接技术领域,特别是涉及一种多主机网络交换系统。

背景技术

[0002] 服务器需要通过网卡连接相应的交换机,进而连入网络内。由于PCIE总线传输距离有限,传统服务器的以太网卡或光纤网卡都集成在计算节点内部,即然后通过外部以太网或光纤交换机与网络连接。如图1所示,图1为目前的多主机网络交换系统的结构示意图。其中,NODE0-NODE7为计算节点,LAN为以太网卡,FC为光纤网卡。

[0003] 然而,在该种结构中,将以太网卡和光纤网卡集成于计算节点内的话,计算节点内需要设置相应的卡槽,导致占用了计算节点内很大的位置且这种结构不便于依据用户的不同需求配置PCIE卡(例如以太网卡或光纤网卡)的数量及类型,灵活性差。并且每个计算节点内均需要集成一组PCIE卡,PCIE卡的数量多,成本高;同时,由于每个计算节点内的以太网卡以及光纤网卡均要与对应的交换机连接,当计算节点过多或要构建大型集群时,会导致外部线缆多且杂乱,管理不便。其中,这里的计算节点代表一个物理主机。

[0004] 因此,如何提供一种配置灵活性高、成本低且便于管理的多主机网络交换系统是本领域技术人员目前需要解决的问题。

发明内容

[0005] 本发明的目的是提供一种多主机网络交换系统,将PCIE卡组单独设置,PCIE卡的配置灵活性高,成本低且便于管理。

[0006] 为解决上述技术问题,本发明提供了一种多主机网络交换系统,包括若干个计算节点、若干个PCIE SWITCH模块以及PCIE卡组;

[0007] 每个所述计算节点分别通过PCIE总线与各个PCIE SWITCH模块的上行端口相连;

[0008] 每个所述PCIE SWITCH模块的下行端口均通过PCIE总线连接一组所述PCIE卡组;所述PCIE卡组内的PCIE卡分别与对应的网络设备连接;

[0009] 所述PCIE SWITCH模块用于将通过对应的PCIE卡组连接的网络设备虚拟化共享至各个所述计算节点。

[0010] 优选地,包括两个所述PCIE SWITCH模块。

[0011] 优选地,还包括:

[0012] 分别与各个所述计算节点一一对应的若干个PCIE retimer模块,每个所述PCIE retimer模块一端通过PCIE总线与对应的计算节点相连,另一端分别通过PCIE总线连接各个所述PCIE SWITCH模块的上行端口;

[0013] 所述PCIE retimer模块用于对所述计算节点发送的PCIE信号进行增强处理。

[0014] 优选地,所述PCIE卡组包括若干个以太网卡以及若干个光纤网卡;

[0015] 相应的,与所述以太网卡对应的网络设备为以太网交换机;与所述光纤网卡对应的网络设备为光纤交换机。

[0016] 优选地,所述PCIE总线具体为PCIE3.0x8总线。

[0017] 优选地,所述计算节点、所述PCIE retimer模块以及所述PCIE SWITCH模块设置于服务器的主板上,所述PCIE卡组设置于子板上。

[0018] 本发明提供了一种多主机网络交换系统,将PCIE卡组设置于计算节点的外部,各个计算节点分别通过PCIE SWITCH模块连接PCIE卡组;PCIE SWITCH模块通过PCIE卡组连接相应的网络设备,并将连接的网络设备虚拟化共享至各个计算节点。即在本发明中,将PCIE卡组单独设置,独立于计算节点之外,设置PCIE卡槽时不再受到计算节点的体积等因素的限制,大大提高了系统配置的灵活性。并且,本发明通过PCIE SWITCH模块实现了将一组PCIE卡组共享至各个计算节点,即每个PCIE SWITCH模块均能实现将与其连接的几个网络设备共享至各个计算节点的操作,在此过程中,仅需各个计算节点与该PCIE SWITCH模块相连即可,故减少了需要设置的PCIE卡的数量,降低了成本,同时大大减少所接线缆的数量,便于进行管理。

附图说明

[0019] 为了更清楚地说明本发明实施例中的技术方案,下面将对现有技术和实施例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附图。

[0020] 图1为目前的多主机网络交换系统的结构示意图;

[0021] 图2为本发明提供的一种多主机网络交换系统的结构示意图;

[0022] 图3为本发明提供的一种多主机网络交换系统的具体实施例的结构示意图。

具体实施方式

[0023] 本发明的核心是提供一种多主机网络交换系统,将PCIE卡组单独设置,PCIE卡的配置灵活性高,成本低且便于管理。

[0024] 为使本发明实施例的目的、技术方案和优点更加清楚,下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0025] 本发明提供了一种多主机网络交换系统,该系统包括若干个计算节点、若干个PCIE SWITCH模块以及PCIE卡组;

[0026] 每个计算节点分别通过PCIE总线与各个PCIE SWITCH模块的上行端口相连;

[0027] 每个PCIE SWITCH模块的下行端口均通过PCIE总线连接一组PCIE卡组;PCIE卡组内的PCIE卡分别与对应的网络设备连接;

[0028] PCIE SWITCH模块用于将通过对应的PCIE卡组连接的网络设备虚拟化共享至各个计算节点。

[0029] 其中,PCIE SWITCH模块支持MR-IOV(Multi Root I/O Virtualization,多根I/O虚拟化技术),PCIE卡也支持MR-IOV,使得PCIE SWITCH模块能够对网络设备进行虚拟化共享,这里的虚拟化共享指的是:假设有4个计算节点,通过一个PCIE SWITCH模块连接一个网

络设备,PCIE SWITCH模块能够将该网络设备虚拟为相同的四份,分别共享至4个计算节点。

[0030] 作为优选地,PCIE卡组包括若干个以太网卡以及若干个光纤网卡(Fibre Channel HBA,FC网卡);

[0031] 相应的,与以太网卡对应的网络设备为以太网交换机;与光纤网卡对应的网络设备为光纤交换机。

[0032] 当然,每个PCIE卡组内包括的PCIE卡的类型以及各种类型的PCIE卡的数量本发明不限定,每个PCIE SWITCH模块连接的PCIE卡组内的PCIE卡的类型及数量可以相同也可以不同,本发明对此也不做限定。

[0033] 可以理解的是,PCIE SWITCH模块具有虚拟化共享的功能。例如若有10个计算节点,目前的情况下需要在每个计算节点内分别设置1个以太网卡以及1个光纤网卡,故共需10个以太网卡以及10个光纤网卡,而本发明中只需令一个PCIE SWITCH模块连接1个以太网卡以及1个光纤网卡即可,减少了需要设置的PCIE卡的数量,降低了成本;并且,现有技术中每个计算节点内的以太网卡以及光纤网卡均需连接相应的交换机,假设有一个以太网交换机以及一个光纤交换机的话,接线数量为20;而本发明中各个计算节点分别连接PCIE SWITCH模块,接线数量为10,PCIE SWITCH模块连接2个PCIE卡,接线数量为2,可见,本发明的结构减小了接线数量,并且简化了接线的复杂程度,便于进行管理;同时,上述优点在计算节点、PCIE卡或网络设备数量较多时更为明显。

[0034] 作为优选地,该系统包括两个PCIE SWITCH模块。

[0035] 可以理解的是,设置两个PCIE SWITCH模块是为了实现网络设备冗余的目的,当一组网络设备故障时,还可通过另一个PCIE SWITCH模块将网络设备共享至计算节点。当然,该系统也可包括两个以上PCIE SWITCH模块,本发明对此不限定。另外,不同的PCIE SWITCH模块连接的PCIE卡组以及网络设备可以相同(即作为冗余)也可以不同(PCIE卡的数量过多,一个PCIE SWITCH模块不够的情况)。

[0036] 参见图2所示,图2为本发明提供的一种多主机网络交换系统的结构示意图;图2中计算节点数量为8,PCIE SWITCH模块的数量为2,每个PCIE SWITCH模块连接了2个以太网卡(2*10G LAN)以及两个光纤网卡(2*SG FC)。当然,本发明不限定计算节点以及PCIE SWITCH模块的数量。

[0037] 作为优选地,参见图3所示,图3为本发明提供的一种多主机网络交换系统的具体实施例的结构示意图。该系统还包括:

[0038] 分别与各个计算节点一一对应的若干个PCIE retimer模块,每个PCIE retimer模块一端通过PCIE总线与对应的计算节点相连,另一端分别通过PCIE总线连接各个PCIE SWITCH模块的上行端口;

[0039] PCIE retimer模块用于对计算节点发送的PCIE信号进行增强处理。

[0040] 可以理解的是,由于PCIE总线传输距离有限,而PCIE SWITCH模块的一般设置的距离计算节点较远,故需要对PCIE信号进行增强,从而保证信号的传输可靠性。

[0041] 其中,PCIE总线具体为PCIE3.0x8总线。也可以为其他类型的PCIE总线,本发明对此不做限定。

[0042] 作为优选地,计算节点、PCIE retimer模块以及PCIE SWITCH模块设置于服务器的主板上,PCIE卡组设置于子板上。

[0043] 可以理解的是,将PCIE卡组独立于主板之外,即可以灵活的依据自身需要设置PCIE的卡槽数量以及类型,大大提高了配置灵活性。当然,将PCIE卡组设置于主板上也可。

[0044] 本发明提供了一种多主机网络交换系统,将PCIE卡组设置于计算节点的外部,各个计算节点分别通过PCIE SWITCH模块连接PCIE卡组;PCIE SWITCH模块通过PCIE卡组连接相应的网络设备,并将连接的网络设备虚拟化共享至各个计算节点。即在本发明中,将PCIE卡组单独设置,独立于计算节点之外,设置PCIE卡槽时不再受到计算节点的体积等因素的限制,大大提高了系统配置的灵活性。并且,本发明通过PCIE SWITCH模块实现了将一组PCIE卡组共享至各个计算节点,即每个PCIE SWITCH模块均能实现将与其连接的几个网络设备共享至各个计算节点的操作,在此过程中,仅需各个计算节点与该PCIE SWITCH模块相连即可,故减少了需要设置的PCIE卡的数量,降低了成本,同时大大减少所接线缆的数量,便于进行管理。

[0045] 需要说明的是,在本说明书中,诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来,而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同要素。

[0046] 对所公开的实施例的上述说明,使本领域专业技术人员能够实现或使用本发明。对这些实施例的多种修改对本领域的专业技术人员来说将是显而易见的,本文中所定义的一般原理可以在不脱离本发明的精神或范围的情况下,在其他实施例中实现。因此,本发明将不会被限制于本文所示的这些实施例,而是要符合与本文所公开的原理和新颖特点相一致的最宽的范围。

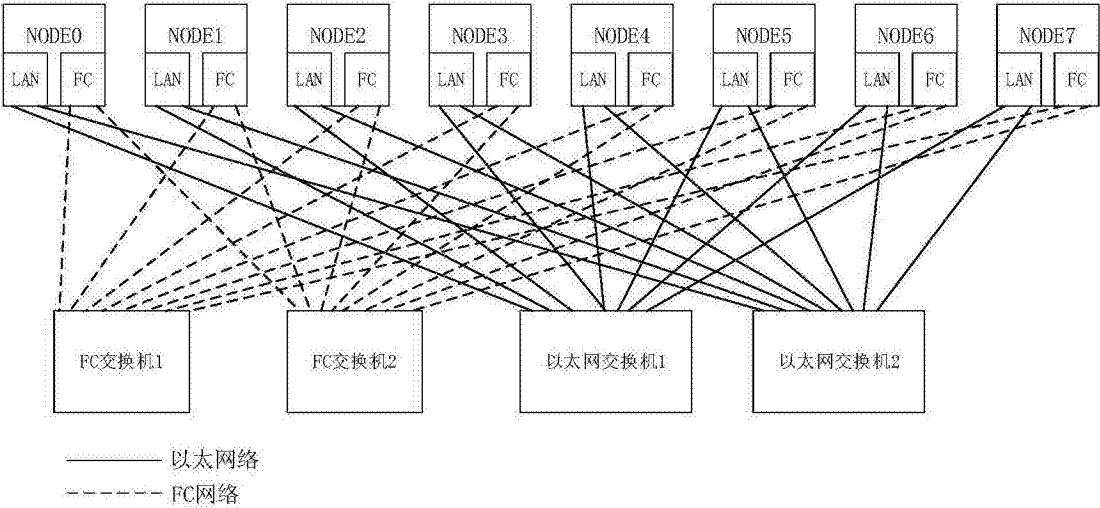


图1

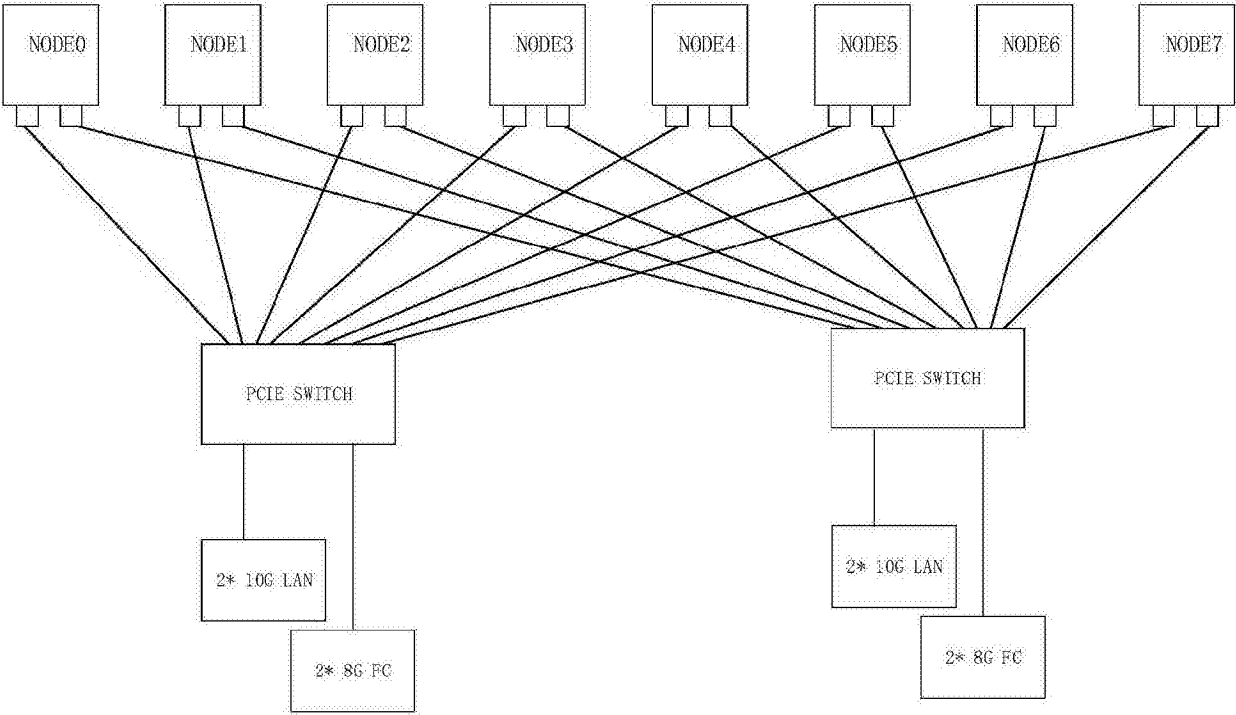


图2

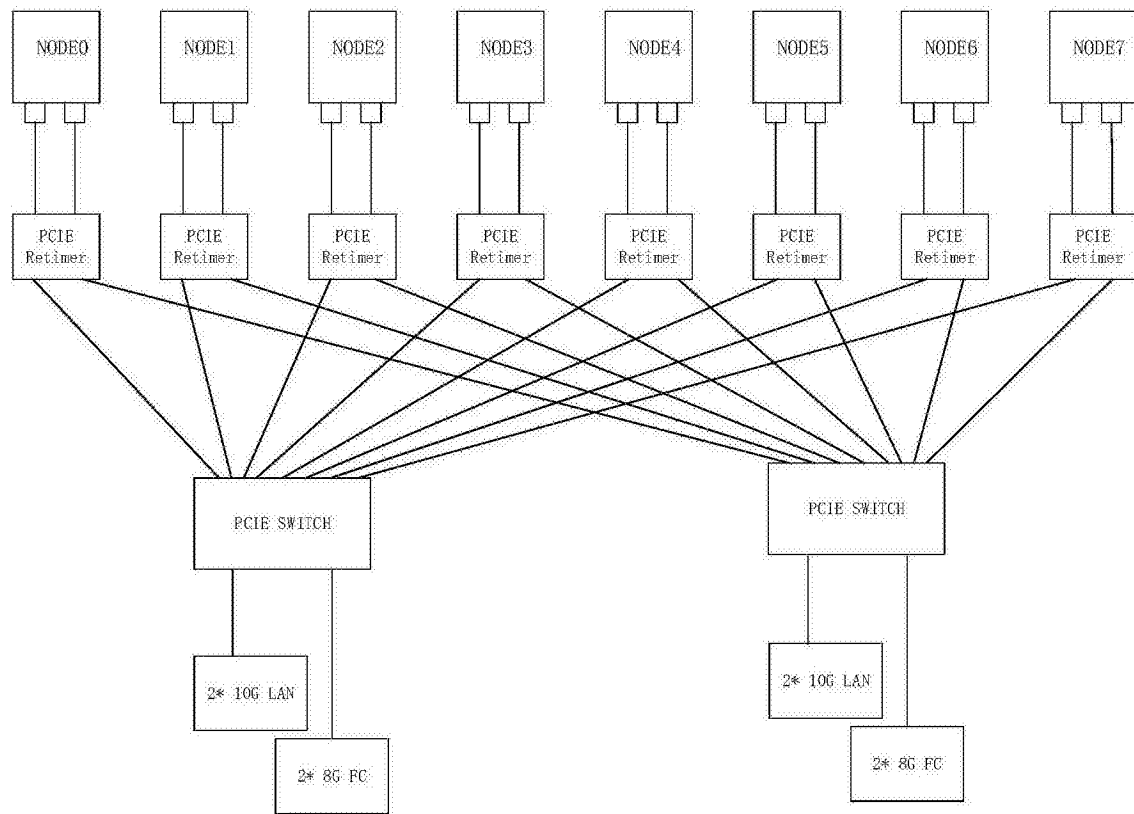


图3