

一种集群及通信方法

5 本申请要求于 2023 年 9 月 19 日提交国家知识产权局、申请号为 202311211303.3、申请名称为“一种集群及通信方法”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请实施例涉及云技术领域，尤其涉及一种集群及通信方法。

10 背景技术

在云服务系统中，云厂商在系统中的物理服务器上加入一块卸载卡，从而将物理服务器中处理器的部分任务（功能）卸载至该卡上完成处理，从而减轻物理服务器中处理器的负担，提高整个云服务系统的性能。

15 目前的云服务系统包含多个物理服务器，对于多个物理服务器中的任意一个物理服务器而言，该物理服务器不仅包含中央处理器、内存、网卡以及硬盘等一系列必要的硬件资源，还额外设置有卸载卡，该卸载卡具有专属的处理器以及存储器，故该卸载卡可代替中央处理器完成资源调度、网络转发、存储前端等一系列的任务，故该物理服务器的中央处理器可以处理更多额外的任务。

20 由于云服务系统中存在任务的局部峰值这一情况，也就是说，系统中的某些物理服务器所需处理大量的任务，而另一些物理服务器则仅需处理少量的任务，这样会导致某些服务器上的卸载卡的利用率高，而另一些物理服务器上的卸载卡的利用率则较低，造成了资源浪费。

发明内容

本申请实施例提供了一种集群及通信方法，可将各种资源以池化的方式呈现，并使得各个资源池解耦，从而提高资源的利用率。

25 本申请实施例的第一方面提供了一种集群，该集群包含数据处理器池以及不同类型的多个计算资源池，其中，数据处理器池包含多个数据处理器，多个计算资源池中的任意一个计算资源池包含多个计算资源。

30 数据处理器池通过高速互联总线与多个计算资源池连接，也就是说，任意一个计算资源池与该集群的外部之间可通过高速互联总线、数据处理器池完成通信连接，多个计算资源池之间也可通过高速互联总线、数据处理器池完成通信连接，任意一个计算资源池的内部也可通过高速互联总线、数据处理器池完成通信连接。由此可见，数据处理器池可用于管理多个计算资源池与集群的外部之间的通信，多个计算资源池之间的通信以及任意一个计算资源池的内部的通信。

35 从上述集群可以看出：在该集群中，各种类型的计算资源以池化的方式进行部署，多个计算资源池之间实现解耦，可按该集群所需提供的云服务类型来进行配置，从而提高该集群的性能和资源利用率。在该集群中，数据处理器也以池化的方式进行部署，可构成数据处理器池，并作为该集群的通信调度中心，实现该池内各个数据处理器的集中分配和利用，从而提高各个数据处理器的利用率。

40 在一种可能实现的方式中，数据处理器池以及多个计算资源池设置于多个机柜中，用于设置数据处理器池的机柜与用于设置任意一个计算资源池的机柜为不同的机柜。前述实现方式中，数据处理器池以及多个计算资源池分别设置在多个机柜中。其中，用于设置数据处理器池的机柜与用于设置任意一个计算资源池的机柜通常为不同的机柜。同理，用于设置某个计算资源池的机柜和用于设置另一个计算资源池的机柜通常也为不同的机柜。

45 在一种可能实现的方式中，多个数据处理器包含通用处理单元以及多个加速处理单元，通用处理单元用于将多个加速处理单元中的至少一个加速处理单元，分配至多个计算资源池中存在通信需求的计算资源，分配至计算资源的至少一个加速处理单元用于满足计算资源的通信需求。前述实现方式中，当多个计算资源池中的任意一个计算资源存在通信需求时，通用处理单元可将多个加速处理单元中的至少一个加速处理单元分配至该计算资源，也就是建立该计算资源与分配至该计算资源的至少一个加速处理单

元之间的连接，故分配至该计算资源的至少一个加速处理单元可满足该计算资源的通信需求。在一种可能实现的方式中，多个数据处理器包含通用处理单元以及多个加速处理单元；通用处理单元用于构建多个计算资源池与集群的外部之间的第一通信路径，多个计算资源池之间的第一通信路径，以及任意一个计算资源池的内部的第一通信路径。多个加速处理单元用于构建多个计算资源池与集群的外部之间的第二通信路径，多个计算资源池之间的第二通信路径，以及任意一个计算资源池的内部的第二通信路径。前述实现方式中，多个数据处理器包含一个或多个通用处理单元以及多个加速处理单元，其中，一个或多个通用处理单元用于构建多个计算资源池中的任意一个计算资源与该集群的外部之间的第一通信路径，任意两个计算资源池中的计算资源之间的第一通信路径，以及任意一个计算资源池中的两个计算资源之间的第一通信路径。多个加速处理单元用于构建多个计算资源池中的任意一个计算资源与集群的外部之间的第二通信路径，任意两个计算资源池中的计算资源之间的第二通信路径，以及任意一个计算资源池中的两个计算资源之间的第二通信路径。

在一种可能实现的方式中，通用处理单元以及多个加速处理单元部署在数据处理器池的内部。前述实现方式中，在数据处理器池的多个数据处理器中，任意一个数据处理器内部具有一个通用处理单元和一个加速处理单元，在这种情况下，通用处理单元和加速处理单元之间是相互耦合的。

在一种可能实现的方式中，多个加速处理单元部署在数据处理器池的内部，通用处理单元部署在数据处理器池的外部，数据处理器池的外部为至少一个计算资源池的内部。前述实现方式中，在数据处理器池所包含的多个数据处理器中，任意一个数据处理器内部仅具有一个加速处理单元。而通用处理单元的数量通常为一个，该通用处理单元可拉远部署，也就是说，该通用处理单元不部署在数据处理器池的任意一个数据处理器中，而是部署在某个或某些计算资源池中，在这种情况下，通用处理单元和加速处理单元之间是解耦的。

在一种可能实现的方式中，多个计算资源池与集群的外部之间的第一通信路径，以多个计算资源池中的任意一个计算资源和集群的外部为端点，中间经过通用处理单元以及多个加速处理单元中的至少一个加速处理单元；多个计算资源池之间的第一通信路径，以任意两个计算资源池中的计算资源为端点，中间经过通用处理单元以及多个加速处理单元中的至少一个加速处理单元；任意一个计算资源池的内部的第一通信路径，以任意一个计算资源池中的两个计算资源为端点，中间经过通用处理单元以及多个加速处理单元中的至少一个加速处理单元。前述实现方式中，对于第一通信路径而言，由于起始点和结束点的不同，可具体分为多个计算资源池与集群的外部之间的第一通信路径，多个计算资源池之间的第一通信路径以及任意一个计算资源池的内部的第一通信路径这三类，但这三类第一通信路径中间均经过通用处理单元以及至少一个加速处理单元。

在一种可能实现的方式中，多个计算资源池与集群的外部之间的第二通信路径，以多个计算资源池中的任意一个计算资源和集群的外部为端点，中间经过多个加速处理单元中的至少一个加速处理单元；多个计算资源池之间的第二通信路径，以任意两个计算资源池中的计算资源为端点，中间经过多个加速处理单元中的至少一个加速处理单元；任意一个计算资源池的内部的第二通信路径，以任意一个计算资源池中的两个计算资源为端点，中间经过多个加速处理单元中的至少一个加速处理单元。前述实现方式中，对于第二通信路径而言，由于起始点和结束点的不同，可具体分为多个计算资源池与集群的外部之间的第二通信路径，多个计算资源池之间的第二通信路径以及任意一个计算资源池的内部的第二通信路径这三类，但这三类第二通信路径中间均仅经过至少一个加速处理单元。

在一种可能实现的方式中，通用处理单元通过令牌发放的方式，建立多个加速处理单元与多个计算资源池中的计算资源之间的连接，对于多个加速处理单元中的任意一个加速处理单元，该加速处理单元所能接收的令牌的数量等于与该加速处理单元所能连接的计算资源的数量。前述实现方式中，在某一轮次的令牌发放中，通用处理单元将令牌发放消息发送给首个加速处理单元，该加速处理单元基于该消息申请一个令牌，并与某个计算资源建立连接。接着，该加速处理单元可判断自身的令牌桶是否已满，若是，说明令牌全部领完，则通知通用处理单元，以使得通用处理单元将该加速处理单元从多个加速处理单元中剔除。然后，该加速处理单元可将令牌发放消息继续发送至下一个加速处理单元，直至回到通用处理单元。至此，则完成了该轮次的令牌发放，可开始下一轮次的令牌发放。那么，当通用处理单元完成针对所有加速处理单元的令牌发放后，多个加速处理单元中的任意一个加速处理单元则与至少一个计

算资源完成了连接。由此可见，对于多个加速处理单元中的任意一个加速处理单元，该加速处理单元所能接收的令牌的数量等于与该加速处理单元所能连接的计算资源的数量。

在一种可能实现的方式中，当加速处理单元所支持的高速互联总线与计算资源所支持的高速互联总线不同时，计算资源与加速处理单元之间通过计算资源所支持的高速互联总线、转换设备与加速处理单元所支持的高速互联总线完成连接。前述实现方式中，当加速处理单元所支持的高速互联总线与计算资源所支持的高速互联总线相同时，计算资源与加速处理单元之间可通过该高速互联总线连接。当加速处理单元所支持的高速互联总线与计算资源所支持的高速互联总线不同时，计算资源与加速处理单元之间则通过计算资源所支持的高速互联总线、转换设备与加速处理单元所支持的高速互联总线完成连接，其中，转换设备用于实现计算资源所支持的高速互联总线与加速处理单元所支持的高速互联总线之间的转换。

在一种可能实现的方式中，多个计算资源池包含以下至少两项：中央处理器（central processing unit, CPU）池、神经网络处理器（neural network processing unit, NPU）池、图形处理器（graphic processing unit, GPU）池、内存池以及硬盘（solid state disk, SSD）池，CPU 池包含多个 CPU，NPU 池包含多个 NPU，GPU 池包含多个 GPU，内存池包含多个内存，SSD 池包含多个 SSD。

在一种可能实现的方式中，任意一个计算资源池包含多个计算资源用于部署提供云服务的云实例，云实例包含以下任意一项：计算资源、虚拟机、容器以及微虚拟机。

本申请实施例的第二方面提供了一种通信方法，该方法通过集群实现，集群包含数据处理器池以及不同类型的多个计算资源池，数据处理器池包含多个数据处理器，多个计算资源池中的任意一个计算资源池包含多个计算资源，数据处理器池通过高速互联总线与多个计算资源池连接；该方法包括：数据处理器池管理多个计算资源池与集群的外部之间的通信，多个计算资源池之间的通信以及任意一个计算资源池的内部的通信。

从上述方法可以看出：在该集群中，各种类型的计算资源以池化的方式进行部署，多个计算资源池之间实现解耦，可按该集群所需提供的云服务类型来进行配置，从而提高该集群的性能和资源利用率。在该集群中，数据处理器也以池化的方式进行部署，可构成数据处理器池，并作为该集群的通信调度中心，实现该池内各个数据处理器的集中分配和利用，从而提高各个数据处理器的利用率。

在一种可能实现的方式中，数据处理器池以及多个计算资源池设置于多个机柜中，用于设置数据处理器池的机柜与用于设置任意一个计算资源池的机柜为不同的机柜。

在一种可能实现的方式中，多个数据处理器包含通用处理单元以及多个加速处理单元，方法还包括：通用处理单元将多个加速处理单元中的至少一个加速处理单元，分配至多个计算资源池中存在通信需求的计算资源，以使得分配至计算资源的至少一个加速处理单元满足计算资源的通信需求。

在一种可能实现的方式中，多个数据处理器包含通用处理单元以及多个加速处理单元；数据处理器池管理多个计算资源池与集群的外部之间的通信，多个计算资源池之间的通信以及任意一个计算资源池的内部的通信包括：通用处理单元构建多个计算资源池与集群的外部之间的第一通信路径，多个计算资源池之间的第一通信路径，以及任意一个计算资源池的内部的第一通信路径。多个加速处理单元构建多个计算资源池与集群的外部之间的第二通信路径，多个计算资源池之间的第二通信路径，以及任意一个计算资源池的内部的第二通信路径。

在一种可能实现的方式中，通用处理单元以及多个加速处理单元部署在数据处理器池的内部。

在一种可能实现的方式中，多个加速处理单元部署在数据处理器池的内部，通用处理单元部署在数据处理器池的外部，数据处理器池的外部为至少一个计算资源池的内部。

在一种可能实现的方式中，多个计算资源池与集群的外部之间的第一通信路径，以多个计算资源池中的任意一个计算资源和集群的外部为端点，中间经过通用处理单元以及多个加速处理单元中的至少一个加速处理单元；多个计算资源池之间的第一通信路径，以任意两个计算资源池中的计算资源为端点，中间经过通用处理单元以及多个加速处理单元中的至少一个加速处理单元；任意一个计算资源池的内部的第一通信路径，以任意一个计算资源池中的两个计算资源为端点，中间经过通用处理单元以及多个加速处理单元中的至少一个加速处理单元。

在一种可能实现的方式中，多个计算资源池与集群的外部之间的第二通信路径，以多个计算资源池

中的任意一个计算资源和集群的外部为端点，中间经过多个加速处理单元中的至少一个加速处理单元；多个计算资源池之间的第二通信路径，以任意两个计算资源池中的计算资源为端点，中间经过多个加速处理单元中的至少一个加速处理单元；任意一个计算资源池的内部的第二通信路径，以任意一个计算资源池中的两个计算资源为端点，中间经过多个加速处理单元中的至少一个加速处理单元。

5 在一种可能实现的方式中，该方法还包括：通用处理单元通过令牌发放的方式，建立多个加速处理单元与多个计算资源池中的计算资源之间的连接，对于多个加速处理单元中的任意一个加速处理单元，该加速处理单元所能接收的令牌的数量等于与该加速处理单元所能连接的计算资源的数量。

在一种可能实现的方式中，当加速处理单元所支持的高速互联总线与计算资源所支持的高速互联总线不同时，计算资源与加速处理单元之间通过计算资源所支持的高速互联总线、转换设备与加速处理单元所支持的高速互联总线完成连接。

10 在一种可能实现的方式中，多个计算资源池包含以下至少两项：CPU池、NPU池、GPU池、内存池以及SSD池，CPU池包含多个CPU，NPU池包含多个NPU，GPU池包含多个GPU，内存池包含多个内存，SSD池包含多个SSD。

在一种可能实现的方式中，任意一个计算资源池包含多个计算资源用于部署提供云服务的云实例，云实例包含以下任意一项：计算资源、虚拟机、容器以及微虚拟机。

15 本申请实施例的第三方面提供了一种云服务系统，云服务系统包含多个如第一方面或第一方面中任意一种可能实现的方式所述的集群。

本申请实施例的第四方面提供了一种计算设备集群，计算设备集群包括至少一个计算设备，每个计算设备包括处理器以及存储器；存储器用于存储指令；处理器用于根据指令，以使得计算设备集群执行第二方面或第二方面中任意一种可能实现的方式所述的方法。

20 本申请实施例的第五方面提供了一种计算机存储介质，计算机存储介质存储有一个或多个指令，指令在由一个或多个计算机执行时使得一个或多个计算机实施第二方面或第二方面中任意一种可能实现的方式所述的方法。

本申请实施例的第六方面提供了一种计算机程序产品，计算机程序产品存储有指令，指令在由计算机执行时，使得计算机实施第二方面或第二方面中任意一种可能实现的方式所述的方法。

25 本申请实施例提供的该集群中，各种类型的计算资源以池化的方式进行部署，多个计算资源池之间实现解耦，可按该集群所需提供的云服务类型来进行配置，从而提高该集群的性能和资源利用率。在该集群中，数据处理器也以池化的方式进行部署，可构成数据处理器池，并作为该集群的通信调度中心，实现该池内各个数据处理器的集中分配和利用，从而提高各个数据处理器的利用率。

30 附图说明

图1为本申请实施例提供的云服务系统的一个结构示意图；

图2为本申请实施例提供的集群的一个结构示意图；

图3为本申请实施例提供的集群的另一结构示意图；

35 图4为本申请实施例提供的集群的另一结构示意图；

图5为本申请实施例提供的集群的另一结构示意图；

图6为本申请实施例提供的令牌发放的一个结构示意图；

图7为本申请实施例提供的令牌发放的一个流程示意图；

图8为本申请实施例提供的集群的另一结构示意图；

40 图9为本申请实施例提供的通信方法的一个示意图；

图10为本申请实施例提供的计算设备的一个结构示意图；

图11为本申请实施例提供的计算设备集群的一个结构示意图；

图12为本申请实施例提供的计算机集群中的计算机设备通过网络连接的一个示意图。

45 具体实施方式

本申请实施例提供了一种集群及通信方法，可将各种资源以池化的方式呈现，并使得各个资源池解

耦，从而提高资源的利用率。

本申请的说明书和权利要求书及上述附图中的术语“第一”、“第二”等是用于区别类似的对象，而不必用于描述特定的顺序或先后次序。应该理解这样使用的术语在适当情况下可以互换，这仅仅是描述本申请的实施例中对相同属性的对象在描述时所采用的区分方式。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，以便包含一系列单元的过程、方法、系统、产品或设备不必限于那些单元，而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它单元。

在云服务系统中，云厂商在系统中的物理服务器上加入一块卸载卡，从而将物理服务器中处理器的部分任务（功能）卸载至该卡上完成处理，从而减轻物理服务器中处理器的负担，提高整个云服务系统的性能。

目前的云服务系统包含多个物理服务器，对于多个物理服务器中的任意一个物理服务器而言，该物理服务器不仅包含中央处理器、内存、网卡以及硬盘等一系列必要的硬件资源，还额外设置有卸载卡，该卸载卡具有专属的处理器以及存储器，故该卸载卡可代替中央处理器完成资源调度、网络转发、存储前端等一系列的任务，故该物理服务器的中央处理器可以处理更多额外的任务。

由于云服务系统中存在任务的局部峰值这一情况，也就是说，系统中的某些物理服务器所需处理大量的任务，而另一些物理服务器则仅需处理少量的任务，这样会导致某些服务器上的卸载卡的利用率高，而另一些物理服务器上的卸载卡的利用率则较低，造成了资源浪费。

进一步地，当需要卸载卡额外负担新增的任务（功能）时，受限于卸载卡自身的资源，往往难以拓展卸载卡的新功能，进而限制了整个云服务系统在各种新的场景中的应用。

进一步地，若某个物理服务器用于提供人工智能（Artificial Intelligence, AI）场景下的云服务，AI 场景中的神经网络模型在训练时往往有独立的数据通信平面，无法直接使用物理服务器上的卸载卡，需要额外增设新的卸载卡供模型训练单独使用，导致成本剧增。

进一步地，受限于工艺，物理服务器上的卸载卡的规格无法跟随业界进行实时更新，也就是说，物理服务器上的卸载卡无法快速替换，这样会影响整个云服务系统的性能。

为了解决上述问题，本申请实施例提供了一种云服务系统，如图 1 所示（图 1 为本申请实施例提供的云服务系统的一个结构示意图），该云服务系统包含多个集群（pod），多个集群之间可通过某种网络（例如，数据中心网络（data center network, DCN），数据中心网络可以为以太网等等）连接。在这多个集群中，一个集群可对外提供一种或多种类型的云服务。比如，某个集群可对外提供云上的 AI 服务，又如，某个集群可对外提供传统的云计算服务，又如，某个集群可对外提供传统的云存储服务等等。

对于多个集群中的任意一个集群而言，如图 2 所示（图 2 为本申请实施例提供的集群的一个结构示意图），该集群也可以称为超节点（hypernode），该集群可包含柜级（rack）的多个硬件池，这个多个硬件池包括：数据处理器（data processing unit, DPU）池以及不同类型的多个计算资源池（也可以称为算力资源池）。下面对数据处理器以及计算资源池分别进行介绍：

对于不同类型的多个计算资源池中的任意一个计算资源池而言，该计算资源池可包含同种类型的多个计算资源。例如，该计算资源池可以为中央处理器（central processing unit, CPU）池，CPU 池则包含多个 CPU。又如，该计算资源池可以为神经网络处理器（neural network processing unit, NPU）池，NPU 池则包含多个 NPU。又如，该计算资源池可以为图形处理器（graphic processing unit, GPU），GPU 池则包含多个 GPU。又如，该计算资源池可以为张量处理单元（tensor processing unit, TPU），TPU 池则包含多个 TPU。又如，该计算资源池可以为内存池，内存池则包含多个内存。又如，该计算资源池可以为固态硬盘（solid state disk, SSD）池，SSD 池包含多个 SSD 等等。由此可见，该集群所包含的多个计算资源池，可以是前述至少一种类型的计算资源池的任意组合。需要说明的是，该集群中所包含的计算资源池的主要类型，与该集群所能提供的云服务的类型相关联，例如，该集群若用于提供 AI 服务，该集群包含的多个计算资源池，主要以 CPU 池、NPU 池和 GPU 池为主。又如，该集群若用于提供通用负载的云计算服务，该集群包含的多个计算资源池，主要以 CPU 池为主。又如，该集群若用于提供云存储服务，该集群包含的多个计算资源池，主要以 SSD 池为主等等。

数据处理器池包含多个数据处理器，数据处理器可以软件定义架构（software defined infrastructure, SDI）芯片，专用集成电路（application specific integrated circuit, ASIC）芯

片,现场可编程门阵列(field programmable gate array, FPGA)芯片、或可编程逻辑器件(programmable logic device, PLD)芯片等各种形式呈现。需要说明的是,在该集群中,数据处理器池可作为多个计算资源池的卸载卡池,可用于统筹管理多个计算资源池与该集群的外部(例如,另一集群,或者,云服务系统中所具有的传统的数据中心、可用区等等)之间的通信,任意两个计算资源池之间的通信,以及任意一个资源池的内部通信。因此,数据处理器池可以视为该集群的出口,用于接入该集群的外部,且数据处理器池还可以视为多个计算资源池之间的转接口,用于连通多个计算资源池,且数据处理器池还可以视为任意一个计算资源池所包含的多个计算资源之间的转接口,用于连通该计算资源池所包含的多个计算资源。

值得注意的是,数据处理器池可通过高速互联总线(可包含高速互联交换机以及高速互连通信连接线等等)与多个计算资源池连接,也就是说,任意一个计算资源池与该集群的外部之间可通过高速互联总线、数据处理器池完成通信连接,多个计算资源池之间也可通过高速互联总线、数据处理器池完成通信连接,任意一个计算资源池的内部也可通过高速互联总线、数据处理器池完成通信连接。由此可见,数据处理器池基于高速互联总线,可统筹管理多个计算资源池与集群的外部之间的通信,任意两个计算资源池之间的通信以及任意一个计算资源池的内部通信。其中,高速互联总线可以为以下任意一种:快捷外围部件互连标准(peripheral component interconnect express, PCIe)总线,或扩展工业标准结构(extended industry standard architecture, EISA)总线、统一总线(unified bus, Ubus 或 UB)、计算机快速链接(compute express link, CXL)、缓存一致互联协议(cache coherent interconnect for accelerators, CCIX)等等。

还值得注意的是,对于任意一个计算资源池所包含的任意一个计算资源,该计算资源可用于部署提供云服务的一个或多个云实例,这些云实例包含以下任意一项:计算资源(也就是裸机)、虚拟机(virtual machine, VM)、容器(docker)以及微虚拟机(microVM)。由此可见,任意一个计算资源池均可部署有一个云实例集群,该云实例集群可用于对外提供某种类型的云服务。

还值得注意的是,从物理硬件层面上而言,数据处理器池以及多个计算资源池分别设置在多个机柜(rack)中。一般地,用于设置数据处理器池的机柜(也就是数据处理器池所包含的多个数据处理器所在的机柜)与用于设置任意一个计算资源池的机柜(也就是该计算资源池所包含的多个计算资源所在的机柜)通常为不同的机柜。同理,用于设置某个计算资源池的机柜(也就是该计算资源池所包含的多个计算资源所在的机柜)和用于设置另一个计算资源池的机柜(也就是该计算资源池所包含的多个计算资源所在的机柜)通常也为不同的机柜。由此可见,数据处理器池以及多个计算资源池之间是存在物理隔离的,且该隔离是柜级的。

例如,如图3所示(图3为本申请实施例提供的集群的另一结构示意图),设云服务系统中包含集群1、集群2、集群3、...、集群p等p个集群。集群1包含SDI卡池1, CPU池1以及NPU池1, SDI卡池1包含相互连接的m个SDI卡, CPU池1包含相互连接的k个CPU, NPU池1包含相互连接的h个NPU,且m个SDI卡、k个CPU以及h个NPU均接入UB交换机,也就是说,SDI卡池1通过UB交换机与CPU池1、NPU池1连接, CPU池1和NPU池1之间也通过UB交换机连接, CPU池1内的CPU之间也通过UB交换机连接, NPU池1内的CPU之间也通过UB交换机连接。其中, p、m、k、h均大于1,且k与h的和等于n。由此可见,SDI卡与xPU(CPU+NPU)之间的配比为m:n。当然,集群2也可以包含SDI卡池2, CPU池2, NPU池2以及GPU池2等等,集群3至集群p也是如此,此处不再赘述。

基于图3可知,SDI卡池1、SDI卡池2、...、SDI卡池p可接入DCN,故在集群1中, CPU池1中的CPU以及NPU池1中的NPU可通过SDI卡池1中的SDI卡,与集群2、集群3、...、集群p以及系统中的虚拟私有云(virtual private cloud),也就是集群1的外部完成通信。进一步地,在集群1中, CPU池1中的CPU可通过SDI卡池1中的SDI卡,与NPU池1中的NPU完成通信。进一步地,在集群1中, CPU池1中的CPU之间可通过SDI卡池1中的SDI卡完成通信,同理, NPU池1中的NPU之间也可通过SDI卡池1中的SDI卡完成通信。当然,集群2至集群p也是如此,此处不再赘述。

具体地,在该集群中,数据处理器池所包含的多个数据处理器可通过以下方式呈现:

多个数据处理器包含一个或多个通用处理单元(service processors unit, SPU)以及多个加速处理单元(network processor unit, NPU),需要说明的是,数据处理器器的数量通常等于加速处理单元的

数量，数据处理器的数量通常大于或等于通用处理单元的数量，也就是说，当通用处理单元的数量为一个时，该通用处理单元用于统一管理多个加速处理单元，当通用处理单元的数量为多个时，一个通用处理单元用于管理多个加速处理单元中的一个加速处理单元。

其中，一个或多个通用处理单元用于构建多个计算资源池中的任意一个计算资源与该集群的外部之间的第一通信路径（也可以称为慢通信路径），任意两个计算资源池中的计算资源之间的第一通信路径，以及任意一个计算资源池中的两个计算资源之间的第一通信路径。多个加速处理单元用于构建多个计算资源池中的任意一个计算资源与集群的外部之间的第二通信路径（也可以称为快通信路径），任意两个计算资源池中的计算资源之间的第二通信路径，以及任意一个计算资源池中的两个计算资源之间的第二通信路径。

10 更具体地，在该集群中，通用处理单元以及加速处理单元可通过以下多种方式进行设置：

（1）通用处理单元和加速处理单元之间相互耦合。在这种情况下，在数据处理器池所包含的多个数据处理器中，任意一个数据处理器的内部具有一个通用处理单元和一个加速处理单元。需要说明的是，对于任意一个数据处理器而言，该数据处理器内的通用处理单元以及加速处理单元可集成在同一个芯片中，也可以分别部署在两个不同的芯片中，此处不做限制。

15 依旧如上述例子，如图 4 所示（图 4 为本申请实施例提供的集群的另一结构示意图，图 4 是在图 3 的基础上绘制得到的），在集群 1 中，对于 SDI 卡池 1 所包含的 m 个 SDI 卡而言，每个 SDI 卡均具有 1 个 SPU 和 1 个 NPU，1 个 SPU 用于管理相应的 1 个 NPU。

（2）通用处理单元和加速处理单元之间解耦。在这种情况下，在数据处理器池所包含的多个数据处理器中，任意一个数据处理器的内部仅具有一个加速处理单元。而通用处理单元的数量通常为一个，该通用处理单元可拉远部署，也就是说，该通用处理单元不部署在数据处理器池的任意一个数据处理器中，而是部署在某个或某些计算资源池中，例如，可以将某个计算资源池中的某个计算资源直接作为通用处理单元等等。

20 依旧如上述例子，如图 5 所示（图 5 为本申请实施例提供的集群的另一结构示意图，图 5 是在图 3 的基础上绘制得到的），在集群 1 中，对于 SDI 卡池 1 所包含的 m 个 SDI 卡而言，每个 SDI 卡均具有 1 个 NPU。值得注意的是，SPU 并未部署在 SDI 卡池 1 中，而是从 CPU 池 1 中选择其中一个 CPU 直接作为 SPU，该 SPU 用于统一管理 m 个 NPU。

25 更具体地，在该集群中，通用处理单元和加速处理单元可通过以下方式构建第一通信路径和第二通信路径：

（1）设存在待发送至某个计算资源池中的某个计算资源的报文，某个加速处理单元接收到该报文时，可先从自身的流表中判断是否存在与该报文相匹配的流表项，若不存在，该加速处理单元则将该报文发送至管理该加速处理单元的通用处理单元。然后，该通用处理单元可从自身的流表中，确定与该报文相匹配的流表项，并将该流表项下发至该加速处理单元，以使得加速处理单元基于该流表项将该报文发送至该计算资源，并将该流表项更新至自身的流表，故后续再有发送至该计算资源的新报文，加速处理单元可直接基于该流表项将新报文发送至该计算资源。基于前述过程可知，该过程用于传输该报文的通信路径，该通信路径的起始点到该通信路径的结束点之间，经过了通用处理单元以及至少一个加速处理单元（在前述过程中，该加速处理单元在将该报文发送至该计算资源的过程中，可能还经过其余加速处理单元），故该通信路径可视为第一通信路径。由于起始点和结束点的不同，第一通信路径可分为以下三种：

（1.1）若某个第一通信路径的起始点为该集群的外部，结束点为多个计算资源池中的任意一个计算资源，且中间经过通用处理单元以及至少一个加速处理单元，则该第一通信路径可视为该计算资源与该集群的外部之间的第一通信路径。依旧如图 5 所示的例子，若 CPU 池 2（位于集群 2）中某个 CPU 上的 VM2 需要将某个报文发送至 CPU 池 1（位于集群 1）中某个 CPU 上的 VM1，该报文由 SDI 卡池 1 中的某个 NPU 接收到后，由于该 NPU 自身的流表不存在与该报文相匹配的流表项，故该 NPU 将报文发送至 SPU，SPU 向该 NPU 下发与该报文相匹配的流表项后，该 NPU 可更新自身的流表，并基于该流表项将该报文发送至 VM1。由此可见，用于传输该报文的通信路径依次经过 VM2（所在的 CPU）、该 NPU、SPU 以及 VM1（所在的 CPU）。

(1.2) 若某个第一通信路径的起始点为多个计算资源池中某个计算资源池中的某个计算资源, 结束点为多个计算资源池中另一个计算资源池中的某个计算资源, 且中间经过通用处理单元以及至少一个加速处理单元, 则该第一通信路径可视为这两个计算资源池中的计算资源之间的第一通信路径。依旧如图 5 所示的例子, 若 CPU 池 1 中某个 CPU 上的 VM3 需要将某个报文发送至 NPU 池 1 中某个 NPU 上的 VM4, 该报文由 SDI 卡池 1 中的某个 NPU 接收到后, 由于该 NPU 自身的流表不存在与该报文相匹配的流表项, 故该 NPU 将报文发送至 SPU, SPU 向该 NPU 下发与该报文相匹配的流表项后, 该 NPU 可更新自身的流表, 并基于该流表项将该报文发送至 VM4。由此可见, 用于传输该报文的通信路径依次经过 VM3(所在的 CPU)、该 NPU、SPU 以及 VM4(所在的 CPU)。

(1.3) 若某个第一通信路径的起始点为多个计算资源池中某个计算资源池中的某个计算资源, 结束点为该计算资源池中的另一个计算资源, 且中间经过通用处理单元以及至少一个加速处理单元, 则该第一通信路径可视为该计算资源池中的两个计算资源之间的第一通信路径。依旧如图 5 所示的例子, 若 CPU 池 1 中某个 CPU 上的 VM5 需要将某个报文发送至 CPU 池 1 中另一 CPU 上的 VM6, 该报文由 SDI 卡池 1 中的某个 NPU 接收到后, 由于该 NPU 自身的流表不存在与该报文相匹配的流表项, 故该 NPU 将报文发送至 SPU, SPU 向该 NPU 下发与该报文相匹配的流表项后, 该 NPU 可更新自身的流表, 并基于该流表项将该报文发送至 VM6。由此可见, 用于传输该报文的通信路径依次经过 VM4(所在的 CPU)、该 NPU、SPU 以及 VM6(所在的 CPU)。

(2) 设存在待发送至某个计算资源池中的某个计算资源的报文, 某个加速处理单元接收到该报文时, 可先从自身的流表中判断是否存在与该报文相匹配的流表项, 若存在, 该加速处理单元则直接基于该流表项将该报文发送至该计算资源。基于前述过程可知, 该过程用于传输该报文的通信路径, 该通信路径的起始点到该通信路径的结束点之间, 未经过通用处理单元, 仅经过了至少一个加速处理单元(在前述过程中, 该加速处理单元在将该报文发送至该计算资源的过程中, 可能还经过其余加速处理单元), 故该通信路径可视为第二通信路径。由于起始点和结束点的不同, 第二通信路径可分为以下三种:

(2.1) 若某个第二通信路径的起始点为该集群的外部, 结束点为多个计算资源池中的任意一个计算资源, 且中间经过至少一个加速处理单元, 则该第二通信路径可视为该计算资源与该集群的外部之间的第二通信路径。依旧如图 5 所示的例子, 若 CPU 池 2(位于集群 2)中某个 CPU 上的 VM2 需要将某个报文发送至 CPU 池 1(位于集群 1)中某个 CPU 上的 VM1, 该报文由 SDI 卡池 1 中的某个 NPU 接收到后, 由于该 NPU 自身的流表存在与该报文相匹配的流表项, 故该 NPU 基于该流表项将该报文发送至 VM1。由此可见, 用于传输该报文的通信路径依次经过 VM2(所在的 CPU)、该 NPU 以及 VM1(所在的 CPU)。

(2.2) 若某个第二通信路径的起始点为多个计算资源池中某个计算资源池中的某个计算资源, 结束点为多个计算资源池中另一个计算资源池中的某个计算资源, 且中间经过至少一个加速处理单元, 则该第二通信路径可视为这两个计算资源池中的计算资源之间的第二通信路径。依旧如图 5 所示的例子, 若 CPU 池 1 中某个 CPU 上的 VM3 需要将某个报文发送至 NPU 池 1 中某个 NPU 上的 VM4, 该报文由 SDI 卡池 1 中的某个 NPU 接收到后, 由于该 NPU 自身的流表存在与该报文相匹配的流表项, 故该 NPU 基于该流表项将该报文发送至 VM4。由此可见, 用于传输该报文的通信路径依次经过 VM3(所在的 CPU)、该 NPU 以及 VM4(所在的 CPU)。

(2.3) 若某个第二通信路径的起始点为多个计算资源池中某个计算资源池中的某个计算资源, 结束点为该计算资源池中的另一个计算资源, 且中间经过至少一个加速处理单元, 则该第二通信路径可视为该计算资源池中的两个计算资源之间的第二通信路径。依旧如图 5 所示的例子, 若 CPU 池 1 中某个 CPU 上的 VM5 需要将某个报文发送至 CPU 池 1 中另一 CPU 上的 VM6, 该报文由 SDI 卡池 1 中的某个 NPU 接收到后, 由于该 NPU 自身的流表存在与该报文相匹配的流表项, 故该 NPU 基于该流表项将该报文发送至 VM6。由此可见, 用于传输该报文的通信路径依次经过 VM4(所在的 CPU)、该 NPU 以及 VM6(所在的 CPU)。

应理解, 本申请实施例中, 仅以该集群中的某个计算资源为通信路径的结束点, 该集群的外部或该集群中的另一计算资源为通信路径的起始点为例进行示意性介绍。在实际应用中, 二者也可以调换过来, 也就是说, 该集群中的某个计算资源也可以为通信路径的起始点, 而该集群的外部或该集群中的另一计算资源为通信路径的结束点, 此处不做赘述。

更具体地, 在该集群中, 通用处理单元还可通过以下方式管理多个加速处理单元:

当多个计算资源池中的任意一个计算资源存在通信需求时（例如，通用处理单元确定该计算资源已启动，可视为该计算资源存在通信需求，通用处理单元可主动为该计算资源分配加速处理单元。又如，该计算资源可向通用处理单元通知其存在通信需求，该计算资源可向通用处理单元申请其所需的加速处理单元等等），通用处理单元可将多个加速处理单元中的至少一个加速处理单元分配至该计算资源，也就是建立该计算资源与分配至该计算资源的至少一个加速处理单元之间的连接，故分配至该计算资源的至少一个加速处理单元可满足该计算资源的通信需求。

示例性地，通用处理单元可通过令牌发放的方式，建立多个加速处理单元与多个计算资源池中的计算资源之间的连接。其中，令牌发放的流程如图 6 和图 7 所示（图 6 为本申请实施例提供的令牌发放的一个结构示意图，图 7 为本申请实施例提供的令牌发放的一个流程示意图），设该集群具有一个通用处理单元，且通用处理单元作为令牌发放模块，各个加速处理单元作为令牌接收模块，其中，各个加速处理单元具有一个令牌桶，令牌桶的容量是有限的，也就是说，各个加速处理单元所能接收的令牌是有限的。

由于令牌发放的流程是多轮次的，且每一轮次的过程是类似的，下文以其中某个轮次进行介绍。在某一轮次的令牌发放中，通用处理单元将令牌发放消息发送给首个加速处理单元，该加速处理单元基于该消息申请一个令牌，并与某个计算资源建立连接。接着，该加速处理单元可判断自身的令牌桶是否已满，若是，说明令牌全部领完，则通知通用处理单元，以使得通用处理单元将该加速处理单元从多个加速处理单元中剔除。然后，该加速处理单元可将令牌发放消息继续发送至下一个加速处理单元，直至回到通用处理单元。至此，则完成了该轮次的令牌发放，可开始下一轮次的令牌发放。那么，当通用处理单元完成针对所有加速处理单元的令牌发放后，多个加速处理单元中的任意一个加速处理单元则与至少一个计算资源完成了连接。由此可见，对于多个加速处理单元中的任意一个加速处理单元，该加速处理单元所能接收的令牌的数量等于与该加速处理单元所能连接的计算资源的数量。

更具体地，在该集群中，加速处理单元与计算资源之间可通过以下多种方式连接：

（1）当加速处理单元所支持的高速互联总线与计算资源所支持的高速互联总线相同时，计算资源与加速处理单元之间可通过该高速互联总线连接。例如，如图 4 和图 5 所示的例子。

（2）当加速处理单元所支持的高速互联总线与计算资源所支持的高速互联总线不同时，计算资源与加速处理单元之间则通过计算资源所支持的高速互联总线、转换设备与加速处理单元所支持的高速互联总线完成连接，其中，转换设备用于实现计算资源所支持的高速互联总线与加速处理单元所支持的高速互联总线之间的转换。例如，如图 8 所示（图 8 为本申请实施例提供的集群的另一结构示意图，图 8 是在图 3 的基础上绘制得到的），在集群 1 中，对于 SDI 卡池 1 所包含的 m 个 SDI 卡而言，每个 SDI 卡均具有 1 个 NPU。值得注意的是，SDI 卡池 1 中的 m 个 NPU 所支持的高速互连总线为 UB，而 CPU 池 1 中的 k 个 CPU 所支持的高速互连总线为 PCIe，故这 k 个 CPU 均通过转接卡接入 UB 交换机中，以与 m 个 NPU 实现通信。

本申请实施例提供了设置与云服务系统中的集群，该集群具备以下多种优势：

（1）在该集群中，各种类型的计算资源以池化的方式进行部署，多个计算资源池之间实现解耦，可按该集群所需提供的云服务类型来进行配置，从而提高该集群的性能和资源利用率。在该集群中，数据处理器也以池化的方式进行部署，可构成数据处理器池，并作为该集群的通信调度中心，实现该池内各个数据处理器的集中分配和利用，从而提高各个数据处理器的利用率。

（2）在该集群中，数据处理器池和多个计算资源池之间通过高速互联总线连接，也就是说，多个计算资源池与该集群的外部之间，多个计算资源池之间，任意一个计算资源池的内部均可通过高速互联总线连接，当高速互连总线能够提供足够的带宽，可以提高该集群对外和对内的通信速度，并降低该集群所需的成本。

（3）在该集群中，计算资源可实现乐高化组合，比如，CPU、NPU、GPU 之间可实现灵活配比，且在软件层面上可实现统一管理。

（4）若该集群用于对外提供云上的 AI 服务，由于该集群包含数据处理器池，数据处理器池中具有多个数据处理器，提供了足够强大的数据处理能力，且高速互连总线的存在，也提供了足够量的带宽，故该集群的数据处理器池可支持 AI 模型训练时所需的大量模型参数的传输，提升了模型训练时的可靠

性，可减少因网络导致的模型训练中断等情况的出现。

(5) 在该集群中，数据处理器所包含的通用处理单元和加速处理单元可解耦部署，即分开部署，进一步释放通用处理单元和加速处理单元所占据的芯片面积，实现加速处理单元的池化部署以及通用处理单元的集中部署，具备灵活 scale out 的能力。

5 (6) 在该集群中，可设置有包含多个内存的内存池，其余类型计算资源池中的计算资源，比如，CPU、NPU 以及 GPU 等等，可使用内存池中的更多内存，不再受限于某个物理服务器中 CPU、NPU 以及 GPU 等仅能使用该物理服务器中的内存，从而构成内存共享（借用）场景，且具备极低的时延。

(7) 该集群可以提供一个软硬件垂直整合平台，也就是说，该集群中的各个池形成了柜级的资源集合，在抽象层统一支持计算，异构，存储等柜级资源的管理以及流量出口。

10 (8) 在该集群中，可设置有包含多个数据处理器的数据处理器池，以此作为柜级的流量出口，实现分布式流量的资源管理调度方案。其中，数据处理器和计算资源可实现 m:n 映射的池化管理，使得该集群中所有计算资源能够使用的带宽提升很多倍。此外，这样的部署也可以使得物理服务器的形态简化，多个数据处理器可在不同类型的计算资源之间实现超分复用，从而降低成本。

15 (9) 在该集群中，数据处理器池被拉远部署，“队头阻塞”(head-of-line blocking, HOL blocking)、拥塞等关键问题得到有效解决。

(10) 在该集群中，数据处理器和计算资源之间的配置可以为 m:n，当计算资源和数据处理器在处理用户的业务时，无论是计算资源还是数据处理器出现故障，都可以快速启动主备互换，实现业务转移，用户对此的感知是秒级的，可提高用户体验。

20 (11) 在该集群中，通用处理单元可集合在一个芯片上，相较于之前的一卡内集成一个通用处理单元和一个加速处理单元的情况，不仅可减少通用处理单元所占据的芯片面积，还可提高通用处理单元的利用率。

(12) 在该集群中，交换机可协同实现在网计算(all reduce, scatter-gather)，数据库加速(DB filter, Table Scan\ 正则表达式)以及哈希加速等等。

25 (13) 在该集群中，为当前数据面静态分配的核，可统一按照资源池管理，从而提升计算资源的利用率，降低计算资源的占用率。

(14) 在该集群中，各种类型的计算资源可以兼容传统的基于 PCIe 的 virtio，nvme 等类型，同时也支持基于新的高速互联总线的全新生态，有利于系统在各种场景中落地。

(15) 在该集群中，若计算资源和数据处理器所支持的高速互连总线不同，二者之间可以通过转接设备，归一到同一高速互连总线方案，有利于系统应用在更多场景中。

30 (16) 在该集群中，组网的 xpu-xpu 之间支持内存借用，统一编址，在高速互连总线的支持下，能实现超大带宽的内存热迁移。

(17) 该集群支持基于 vfio/vdpa 的设备直通的内存免 pin 方案，支持内存超分复用等等。

35 以上是对本申请实施例提供的集群所进行的详细说明，以下将对本申请实施例提供的通信方法进行介绍。图 9 为本申请实施例提供的通信方法的一个示意图，如图 9 所示，该方法通过如图 2 所示的集群实现，该集群包含数据处理器池以及不同类型的多个计算资源池，数据处理器池包含多个数据处理器，多个计算资源池中的任意一个计算资源池包含多个计算资源，数据处理器池通过高速互联总线与多个计算资源池连接。该方法包括：

901、数据处理器池管理多个计算资源池与集群的外部之间的通信，多个计算资源池之间的通信以及任意一个计算资源池的内部通信。

40 在一种可能实现的方式中，多个数据处理器包含通用处理单元以及多个加速处理单元；数据处理器池管理多个计算资源池与集群的外部之间的通信，多个计算资源池之间的通信以及任意一个计算资源池的内部通信包括：通用处理单元构建多个计算资源池与集群的外部之间的第一通信路径，多个计算资源池之间的第一通信路径，以及任意一个计算资源池的内部的第一通信路径。多个加速处理单元构建多个计算资源池与集群的外部之间的第二通信路径，多个计算资源池之间的第二通信路径，以及任意一个
45 计算资源池的内部第二通信路径。

在一种可能实现的方式中，数据处理器池以及多个计算资源池设置于多个机柜中，用于设置数据处

理器池的机柜与用于设置任意一个计算资源池的机柜为不同的机柜。

在一种可能实现的方式中，多个数据处理器包含通用处理单元以及多个加速处理单元，方法还包括：通用处理单元将多个加速处理单元中的至少一个加速处理单元，分配至多个计算资源池中存在通信需求的计算资源，以使得分配至计算资源的至少一个加速处理单元满足计算资源的通信需求。

5 在一种可能实现的方式中，通用处理单元以及多个加速处理单元部署在数据处理器池的内部。

在一种可能实现的方式中，多个加速处理单元部署在数据处理器池的内部，通用处理单元部署在数据处理器池的外部，数据处理器池的外部为至少一个计算资源池的内部。

在一种可能实现的方式中，多个计算资源池与集群的外部之间的第一通信路径，以多个计算资源池中的任意一个计算资源和集群的外部为端点，中间经过通用处理单元以及多个加速处理单元中的至少一个加速处理单元；多个计算资源池之间的第一通信路径，以任意两个计算资源池中的计算资源为端点，中间经过通用处理单元以及多个加速处理单元中的至少一个加速处理单元；任意一个计算资源池的内部的第一通信路径，以任意一个计算资源池中的两个计算资源为端点，中间经过通用处理单元以及多个加速处理单元中的至少一个加速处理单元。

在一种可能实现的方式中，多个计算资源池与集群的外部之间的第二通信路径，以多个计算资源池中的任意一个计算资源和集群的外部为端点，中间经过多个加速处理单元中的至少一个加速处理单元；多个计算资源池之间的第二通信路径，以任意两个计算资源池中的计算资源为端点，中间经过多个加速处理单元中的至少一个加速处理单元；任意一个计算资源池的内部的第二通信路径，以任意一个计算资源池中的两个计算资源为端点，中间经过多个加速处理单元中的至少一个加速处理单元。

在一种可能实现的方式中，该方法还包括：通用处理单元通过令牌发放的方式，建立多个加速处理单元与多个计算资源池中的计算资源之间的连接，对于多个加速处理单元中的任意一个加速处理单元，该加速处理单元所能接收的令牌的数量等于与该加速处理单元所能连接的计算资源的数量。

在一种可能实现的方式中，当加速处理单元所支持的高速互联总线与计算资源所支持的高速互联总线不同时，计算资源与加速处理单元之间通过计算资源所支持的高速互联总线、转换设备与加速处理单元所支持的高速互联总线完成连接。

25 在一种可能实现的方式中，多个计算资源池包含以下至少两项：CPU 池、NPU 池、GPU 池、内存池以及 SSD 池，CPU 池包含多个 CPU，NPU 池包含多个 NPU，GPU 池包含多个 GPU，内存池包含多个内存，SSD 池包含多个 SSD。

在一种可能实现的方式中，任意一个计算资源池包含多个计算资源用于部署提供云服务的云实例，云实例包含以下任意一项：计算资源、虚拟机、容器以及微虚拟机。

30 需要说明的是，上述方法中各个步骤的实现过程等内容，由于与本申请装置实施例基于同一构思，其带来的技术效果与本申请方法实施例相同，具体内容可参考本申请实施例前述所示的装置实施例中的叙述，此处不再赘述。

请参阅图 10，图 10 为本申请实施例提供的计算设备的一个结构示意图。如图 10 所示，该计算设备 1000 包括：处理器 1001、存储器 1002、通信接口 1003 和总线 1004，处理器 1001、存储器 1002 与通信接口 1003 通过总线（图中未标注）耦合。存储器 1002 存储有指令，当存储器 1002 中的执行指令被执行时，计算设备 1000 执行上述方法实施例中的通信方法。

计算设备 1000 可以是配置成实施以上方法的一个或多个集成电路，例如：一个或多个特定集成电路（application specific integrated circuit, ASIC），或，一个或多个微处理器（digital signal processor, DSP），或，一个或者多个现场可编程门阵列（field programmable gate array, FPGA），或这些集成电路形式中至少两种的组合。再如，当装置中的单元可以通过处理元件调度程序的形式实现时，该处理元件可以是通用处理器，例如中央处理器（central processing unit, CPU）或其它可以调用程序的处理器。再如，这些单元可以集成在一起，以片上系统（system-on-a-chip, SOC）的形式实现。

45 处理器 1001 可以是中央处理单元（central processing unit, CPU），还可以是其它通用处理器、数字信号处理器（digital signal processor, DSP）、专用集成电路（application specific integrated circuit, ASIC）、现场可编程门阵列（field programmable gate array, FPGA）或者其它可编程逻辑

器件、晶体管逻辑器件，硬件部件或者其任意组合。通用处理器可以是微处理器，也可以是任何常规的处理

存储器 1002 可以是易失性存储器或非易失性存储器，或可包括易失性和非易失性存储器两者。其中，非易失性存储器可以是只读存储器（read-only memory, ROM）、可编程只读存储器（programmable ROM, PROM）、可擦除可编程只读存储器（erasable PROM, EPROM）、电可擦除可编程只读存储器（electrically EPROM, EEPROM）或闪存。易失性存储器可以是随机存取存储器（random access memory, RAM），其用作外部高速缓存。通过示例性但不是限制性说明，许多形式的 RAM 可用，例如静态随机存取存储器（static RAM, SRAM）、动态随机存取存储器（DRAM）、同步动态随机存取存储器（synchronous DRAM, SDRAM）、双倍数据速率同步动态随机存取存储器（double data rate SDRAM, DDR SDRAM）、增强型同步动态随机存取存储器（enhanced SDRAM, ESDRAM）、同步连接动态随机存取存储器（synchlink DRAM, SLD RAM）和直接内存总线随机存取存储器（direct rambus RAM, DR RAM）。

存储器 1002 中存储有可执行的程序代码，处理器 1001 执行该可执行的程序代码以分别实现前述通用处理单元和加速处理单元等模块的功能，从而实现上述通信方法。也即，存储器 1002 上存有用于执行上述通信方法的指令。

通信接口 1003 使用例如但不限于网络接口卡、收发器一类的收发模块，来实现计算设备 1000 与其他设备或通信网络之间的通信。

总线 1004 除包括数据总线之外，还可以包括电源总线、控制总线和状态信号总线等。总线可以是快捷外围部件互连标准（peripheral component interconnect express, PCIe）总线，或扩展工业标准结构（extended industry standard architecture, EISA）总线、统一总线（unified bus, Ubus 或 UB）、计算机快速链接（compute express link, CXL）、缓存一致互联协议（cache coherent interconnect for accelerators, CCIX）等。总线可以分为地址总线、数据总线、控制总线等。

请参阅图 11，图 11 为本申请实施例提供的计算设备集群的一个结构示意图。如图 11 所示，该计算设备集群 1100 包括至少一台计算设备 1000。

如图 11 所示，所述计算设备集群 1100 包括至少一个计算设备 1000。计算设备集群 1100 中的一个或多个计算设备 1000 中的存储器 1002 中可以存有相同的用于执行上述通信方法的指令。

在一些可能的实现方式中，该计算设备集群 1100 中的一个或多个计算设备 1000 的存储器 1002 中也可以分别存有用于执行上述通信方法的部分指令。换言之，一个或多个计算设备 1000 的组合可以共同执行用于执行上述通信方法。

需要说明的是，计算设备集群 1100 中的不同的计算设备 1000 中的存储器 1002 可以存储不同的指令，分别用于执行上述数据处理器的一部分功能。也即，不同的计算设备 1000 中的存储器 1002 存储的指令可以实现通用处理单元以及加速处理单元等等模块中的一个或多个模块的功能。

在一些可能的实现方式中，计算设备集群 1100 中的一个或多个计算设备 1000 可以通过网络连接。其中，所述网络可以是广域网或局域网等等。

请参阅图 12，图 12 为本申请实施例提供的计算机集群中的计算机设备通过网络连接的一个示意图。如图 12 所示，两个计算设备 1000A 和 1000B 之间通过网络进行连接。具体地，通过各个计算设备中的通信接口与所述网络进行连接。

在一种可能的实现方式中，计算设备 1000A 中的存储器中存有执行通用处理单元等等模块的功能的指令。同时，计算设备 1000B 中的存储器中存有执行加速处理单元等等模块的功能的指令。

应理解，图 12 中示出的计算设备 1000A 的功能也可以由多个计算设备完成。同样，计算设备 1000B 的功能也可以由多个计算设备完成。

本申请实施例还涉及一种计算机存储介质，该计算机可读存储介质中存储有用于进行信号处理的程序，当其在计算机上运行时，使得计算机执行如图 9 所示实施例中的通信方法。

本申请实施例还涉及一种计算机程序产品，该计算机程序产品存储有指令，该指令在由计算机执行时使得计算机执行如图 9 所示实施例中的通信方法。

所属领域的技术人员可以清楚地了解到，为描述的方便和简洁，上述描述的系统、装置和单元的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

在本申请所提供的几个实施例中，应该理解到，所揭露的系统，装置和方法，可以通过其它的方式实现。例如，以上所描述的装置实施例仅仅是示意性的，例如，所述单元的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个单元或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口，装置或单元的间接耦合或通信连接，可以是电性，机械或其它的形式。

所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

另外，在本申请各个实施例中的各功能单元可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现，也可以采用软件功能单元的形式实现。

所述集成的单元如果以软件功能单元的形式实现并作为独立的产品销售或使用时，可以存储在一个计算机可读取存储介质中。基于这样的理解，本申请的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的全部或部分可以以软件产品的形式体现出来，该计算机软件产品存储在一个存储介质中，包括若干指令用以使得一台计算机设备（可以是个人计算机，服务器，或者网络设备等）执行本申请各个实施例所述方法的全部或部分步骤。而前述的存储介质包括：U 盘、移动硬盘、只读存储器（ROM， read-only memory）、随机存取存储器（RAM， random access memory）、磁碟或者光盘等各种可以存储程序代码的介质。

20

权 利 要 求 书

1. 一种集群，其特征在于，所述集群包含数据处理器池以及不同类型的多个计算资源池，所述数据处理器池包含多个数据处理器，所述多个计算资源池中的任意一个计算资源池包含多个计算资源；

所述数据处理器池通过高速互联总线与所述多个计算资源池连接；

5 所述数据处理器池用于管理所述多个计算资源池与所述集群的外部之间的通信，所述多个计算资源池之间的通信以及所述任意一个计算资源池的内部的通信。

2. 根据权利要求 1 所述的集群，其特征在于，所述数据处理器池以及所述多个计算资源池设置于多个机柜中，用于设置所述数据处理器池的机柜与用于设置所述任意一个计算资源池的机柜为不同的机柜。

3. 根据权利要求 2 所述的集群，其特征在于，所述多个数据处理器包含通用处理单元以及多个加速处理单元，所述通用处理单元用于将所述多个加速处理单元中的至少一个加速处理单元，分配至所述多个计算资源池中存在通信需求的计算资源，分配至所述计算资源的至少一个加速处理单元用于满足所述计算资源的通信需求。

4. 根据权利要求 3 所述的集群，其特征在于，所述通用处理单元用于构建所述多个计算资源池与所述集群的外部之间的第一通信路径，所述多个计算资源池之间的第一通信路径，以及所述任意一个计算资源池的内部的通信路径；

所述多个加速处理单元用于构建所述多个计算资源池与所述集群的外部之间的第二通信路径，所述多个计算资源池之间的第二通信路径，以及所述任意一个计算资源池的内部的第二通信路径。

5. 根据权利要求 3 或 4 所述的集群，其特征在于，所述通用处理单元以及所述多个加速处理单元部署在所述数据处理器池的内部。

6. 根据权利要求 3 或 4 所述的集群，其特征在于，所述多个加速处理单元部署在所述数据处理器池的内部，所述通用处理单元部署在所述数据处理器池的外部，所述数据处理器池的外部为至少一个计算资源池的内部。

7. 根据权利要求 3 至 6 任意一项所述的集群，其特征在于，所述多个计算资源池与所述集群的外部之间的第一通信路径，以所述多个计算资源池中的任意一个计算资源和所述集群的外部为端点，中间经过所述通用处理单元以及所述多个加速处理单元中的至少一个加速处理单元；

所述多个计算资源池之间的第一通信路径，以任意两个计算资源池中的计算资源为端点，中间经过所述通用处理单元以及所述多个加速处理单元中的至少一个加速处理单元；

所述任意一个计算资源池的内部的通信路径，以所述任意一个计算资源池中的两个计算资源为端点，中间经过所述通用处理单元以及所述多个加速处理单元中的至少一个加速处理单元。

8. 根据权利要求 3 至 6 任意一项所述的集群，其特征在于，所述多个计算资源池与所述集群的外部之间的第二通信路径，以所述多个计算资源池中的任意一个计算资源和所述集群的外部为端点，中间经过所述多个加速处理单元中的至少一个加速处理单元；

所述多个计算资源池之间的第二通信路径，以任意两个计算资源池中的计算资源为端点，中间经过所述多个加速处理单元中的至少一个加速处理单元；

所述任意一个计算资源池的内部的第二通信路径，以所述任意一个计算资源池中的两个计算资源为端点，中间经过所述多个加速处理单元中的至少一个加速处理单元。

9. 根据权利要求 2 至 8 任意一项所述的集群，其特征在于，当加速处理单元所支持的高速互联总线与计算资源所支持的高速互联总线不同时，计算资源与加速处理单元之间通过计算资源所支持的高速互联总线、转换设备与加速处理单元所支持的高速互联总线完成连接。

10. 根据权利要求 1 至 9 任意一项所述的集群，其特征在于，所述多个计算资源池包含以下至少两项：中央处理器 CPU 池、神经网络处理器 NPU 池、图形处理器 GPU 池、内存池以及硬盘 SSD 池，CPU 池包含多个 CPU，NPU 池包含多个 NPU，GPU 池包含多个 GPU，内存池包含多个内存，SSD 池包含多个 SSD。

11. 根据权利要求 1 至 10 任意一项所述的集群，其特征在于，所述任意一个计算资源池包含多个计算资源用于部署提供云服务的云实例，所述云实例包含以下任意一项：计算资源、虚拟机、容器以及微虚拟机。

12. 一种通信方法，其特征在于，所述方法通过集群实现，所述集群包含数据处理器池以及不同类型的多个计算资源池，所述数据处理器池包含多个数据处理器，所述多个计算资源池中的任意一个计算资源池包含多个计算资源，所述数据处理器池通过高速互联总线与所述多个计算资源池连接；

所述方法包括：

5 所述数据处理器池管理所述多个计算资源池与所述集群的外部之间的通信，所述多个计算资源池之间的通信以及所述任意一个计算资源池的内部的通信。

13. 根据权利要求 12 所述的方法，其特征在于，所述数据处理器池以及所述多个计算资源池设置于多个机柜中，用于设置所述数据处理器池的机柜与用于设置所述任意一个计算资源池的机柜为不同的机柜。

10 14. 根据权利要求 13 所述的方法，其特征在于，所述多个数据处理器包含通用处理单元以及多个加速处理单元，所述方法还包括：

所述通用处理单元将所述多个加速处理单元中的至少一个加速处理单元，分配至所述多个计算资源池中存在通信需求的计算资源，以使得分配至所述计算资源的至少一个加速处理单元满足所述计算资源的通信需求。

15 15. 根据权利要求 14 所述的方法，其特征在于，所述多个数据处理器包含通用处理单元以及多个加速处理单元；所述数据处理器池管理所述多个计算资源池与所述集群的外部之间的通信，所述多个计算资源池之间的通信以及所述任意一个计算资源池的内部的通信包括：

所述通用处理单元构建所述多个计算资源池与所述集群的外部之间的第一通信路径，所述多个计算资源池之间的第一通信路径，以及所述任意一个计算资源池的内部的第一通信路径；

20 所述多个加速处理单元构建所述多个计算资源池与所述集群的外部之间的第二通信路径，所述多个计算资源池之间的第二通信路径，以及所述任意一个计算资源池的第二通信路径。

16. 根据权利要求 14 或 15 所述的方法，其特征在于，所述通用处理单元以及所述多个加速处理单元部署在所述数据处理器池的内部。

25 17. 根据权利要求 14 或 15 所述的方法，其特征在于，所述多个加速处理单元部署在所述数据处理器池的内部，所述通用处理单元部署在所述数据处理器池的外部，所述数据处理器池的外部为至少一个计算资源池的内部。

18. 根据权利要求 14 至 17 任意一项所述的方法，其特征在于，所述多个计算资源池与所述集群的外部之间的第一通信路径，以所述多个计算资源池中的任意一个计算资源和所述集群的外部为端点，中间经过所述通用处理单元以及所述多个加速处理单元中的至少一个加速处理单元；

30 所述多个计算资源池之间的第一通信路径，以任意两个计算资源池中的计算资源为端点，中间经过所述通用处理单元以及所述多个加速处理单元中的至少一个加速处理单元；

所述任意一个计算资源池的内部的第一通信路径，以所述任意一个计算资源池中的两个计算资源为端点，中间经过所述通用处理单元以及所述多个加速处理单元中的至少一个加速处理单元。

35 19. 根据权利要求 14 至 18 任意一项所述的方法，其特征在于，所述多个计算资源池与所述集群的外部之间的第二通信路径，以所述多个计算资源池中的任意一个计算资源和所述集群的外部为端点，中间经过所述多个加速处理单元中的至少一个加速处理单元；

所述多个计算资源池之间的第二通信路径，以任意两个计算资源池中的计算资源为端点，中间经过所述多个加速处理单元中的至少一个加速处理单元；

40 所述任意一个计算资源池的第二通信路径，以所述任意一个计算资源池中的两个计算资源为端点，中间经过所述多个加速处理单元中的至少一个加速处理单元。

20. 根据权利要求 13 至 19 任意一项所述的方法，其特征在于，当加速处理单元所支持的高速互联总线与计算资源所支持的高速互联总线不同时，计算资源与加速处理单元之间通过计算资源所支持的高速互联总线、转换设备与加速处理单元所支持的高速互联总线完成连接。

45 21. 根据权利要求 12 至 20 任意一项所述的方法，其特征在于，所述多个计算资源池包含以下至少两项：CPU 池、NPU 池、GPU 池、内存池以及 SSD 池，CPU 池包含多个 CPU，NPU 池包含多个 NPU，GPU 池包含多个 GPU，内存池包含多个内存，SSD 池包含多个 SSD。

22. 根据权利要求 12 至 21 任意一项所述的方法，其特征在于，所述任意一个计算资源池包含多个计算资源用于部署提供云服务的云实例，所述云实例包含以下任意一项：计算资源、虚拟机、容器以及微虚拟机。

5 23. 一种云服务系统，其特征在于，所述云服务系统包含多个如权利要求 1 至 11 任意一项所述的集群。

24. 一种计算设备集群，其特征在于，所述计算设备集群包括至少一个计算设备，每个计算设备包括处理器以及存储器：

所述存储器用于存储指令；

10 所述处理器用于根据所述指令，以使得所述计算设备集群执行权利要求 12 至 22 中任一项所述的方法。

25. 一种计算机存储介质，其特征在于，所述计算机存储介质存储有一个或多个指令，所述指令在由一个或多个计算机执行时使得所述一个或多个计算机实施权利要求 12 至 22 任一所述的方法。

26. 一种计算机程序产品，其特征在于，所述计算机程序产品存储有指令，所述指令在由计算机执行时，使得所述计算机实施权利要求 12 至 22 任意一项所述的方法。

15