(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 116257471 A
(43)申请公布日 2023.06.13

(21)申请号 202111511024.X

(22)申请日 2021.12.10

(71)申请人 华为技术有限公司
地址 518129 广东省深圳市龙岗区坂田华为总部办公楼

(72)发明人 陈晓雨

(74)专利代理机构 北京中博世达专利商标代理有限公司 11274
专利代理师 申健

(51)Int.Cl.
G06F 13/28(2006.01)
G06F 9/54(2006.01)

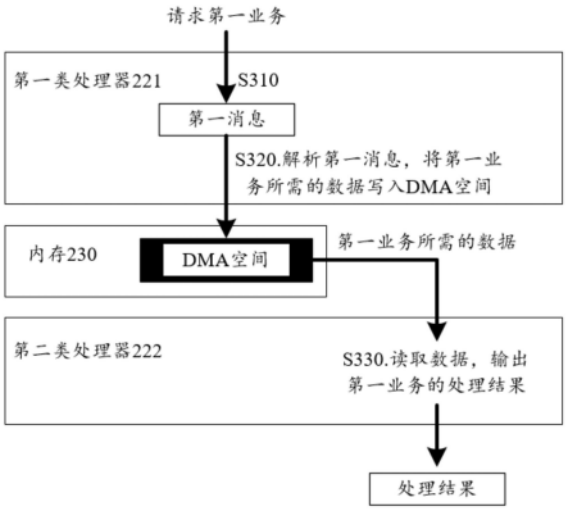
权利要求书3页 说明书17页 附图3页

(54)发明名称

一种业务处理方法及装置

(57)摘要

一种业务处理方法及装置，涉及计算机技术领域。该方法包括：首先，服务器包括的第一类处理器获取用于请求第一业务的第一消息；其次，第一类处理器解析第一消息，将第一业务所需的数据写入DMA空间，该DMA空间是由服务器的内存基于DMA技术提供的存储空间；最后，服务器中性能优于第一类处理器的第二类处理器从DMA空间中读取前述数据，输出第一业务的处理结果。由于第一类处理器无需向第二类处理器发送中断信号，避免第二类处理器的业务中断，降低业务的处理时延。且，DMA空间可以直接被硬件访问，如直接被第一类处理器和第二类处理器进行数据访问，降低地址转换导致的处理时延，进一步降低服务器处理业务的整体时延，提高服务器处理业务的性能。



CN 116257471 A

1.一种业务处理方法,其特征在于,所述方法由服务器执行,所述服务器包括第一类处理器和第二类处理器,所述第二类处理器的性能优于所述第一类处理器,所述方法包括:

所述第一类处理器获取用于请求第一业务的第一消息;

所述第一类处理器解析所述第一消息,将所述第一业务所需的数据写入直接内存访问DMA空间;所述DMA空间是由所述服务器的内存基于DMA技术提供的存储空间;

所述第二类处理器从所述DMA空间中读取所述数据,输出所述第一业务的处理结果。

2.根据权利要求1所述的方法,其特征在于,

所述第一类处理器获取用于请求第一业务的第一消息,包括:

所述第一类处理器从所述服务器的网卡中读取所述第一消息;或者,

所述第一类处理器接收所述网卡发送的所述第一消息。

3.根据权利要求1或2所述的方法,其特征在于,所述方法还包括:

所述第一类处理器对所述内存进行初始化,确定所述DMA空间;或者,

所述第二类处理器对所述内存进行初始化,确定所述DMA空间。

4.根据权利要求1-3中任一项所述的方法,其特征在于,

所述第一业务包括一个或多个任务;

一个任务的信号量增加线程和信号量减少线程挂载于所述第二类处理器的同一个处理器。

5.根据权利要求1-4中任一项所述的方法,其特征在于,所述第一业务所需的数据包括所述第一业务的业务数据,和处理所述业务数据的函数;

所述第一类处理器解析所述第一消息,将所述第一业务所需的数据写入DMA空间,包括:

所述第一类处理器依据所述第一消息确定函数标识,所述函数标识用于指示执行所述第一业务所需的函数列表,所述函数列表包括一个或多个函数;

所述第一类处理器将所述函数列表和所述业务数据写入所述DMA空间。

6.根据权利要求5所述的方法,其特征在于,所述第二类处理器包括一个或多个处理器;

所述第一类处理器将所述函数列表和所述业务数据写入所述DMA空间,包括:

所述第一类处理器从所述第二类处理器中确定一个或多个可用处理器,所述可用处理器为所述第二类处理器中符合调度需求的处理器;

所述第一类处理器将所述函数列表和所述业务数据写入所述DMA空间中的就绪队列ReadyQ,一个ReadyQ与一个可用处理器关联,所述一个ReadyQ用于存储所述一个可用处理器的待处理任务的信息。

7.根据权利要求6所述的方法,其特征在于,所述服务器中存储有状态信息,所述状态信息用于指示所述一个或多个处理器的线程标识ID,和挂载在所述处理器中的空闲用户态线程的数量;

所述调度需求是指处理器的线程ID为非空、且挂载在所述处理器中的空闲用户态线程的数量达到第一阈值。

8.根据权利要求1-7中任一项所述的方法,其特征在于,所述方法还包括:

所述第一类处理器确定所述第二类处理器中可用处理器的数量;

若所述可用处理器的数量小于或等于第二阈值,所述第一类处理器启动所述第二类处理器中除所述可用处理器外的一个或多个其他处理器。

9.根据权利要求1所述的方法,其特征在于,所述第一业务所需的数据包括所述第一业务的业务数据,和处理所述业务数据的函数;

所述第二类处理器从所述DMA空间中读取所述数据,输出所述第一业务的处理结果,包括:

所述第二类处理器依据所述DMA空间中的用户态线程调度策略,确定所述第一业务的调度计划;所述用户态线程调度策略包含多个函数的执行顺序,以及所述DMA空间提供给所述第二类处理器中任一处理器的存储资源阈值;

所述第二类处理器依据所述调度计划调度所述处理所述业务数据的函数,对所述业务数据进行处理,输出所述第一业务的处理结果。

10.一种业务处理装置,其特征在于,所述装置应用于服务器,所述服务器包括第一类处理器和第二类处理器,所述第二类处理器的性能优于所述第一类处理器,所述装置包括:应用于所述第一类处理器的硬件调度模块,和应用于所述第二类处理器的业务处理模块;

所述硬件调度模块,用于获取用于请求第一业务的第一消息;

所述硬件调度模块,还用于解析所述第一消息,将所述第一业务所需的数据写入直接内存访问DMA空间;所述DMA空间是由服务器的内存基于DMA技术提供的存储空间;

所述业务处理模块,用于从所述DMA空间中读取所述数据,输出所述第一业务的处理结果。

11.根据权利要求10所述的装置,其特征在于,

所述硬件调度模块,具体用于从所述服务器的网卡中读取所述第一消息;或者,

所述硬件调度模块,具体用于接收所述网卡发送的所述第一消息。

12.根据权利要求10或11所述的装置,其特征在于,

所述硬件调度模块,还用于对所述内存进行初始化,确定所述DMA空间;或者,

所述业务处理模块,还用于对所述内存进行初始化,确定所述DMA空间。

13.根据权利要求10-12中任一项所述的装置,其特征在于,

所述第一业务包括一个或多个任务;

一个任务的信号量增加线程和信号量减少线程挂载于所述第二类处理器的同一个处理器。

14.根据权利要求10-13中任一项所述的装置,其特征在于,所述第一业务所需的数据包括所述第一业务的业务数据,和处理所述业务数据的函数;

所述硬件调度模块,具体用于依据所述第一消息确定函数标识,所述函数标识用于指示执行所述第一业务所需的函数列表,所述函数列表包括一个或多个函数;

所述硬件调度模块,还具体用于将所述函数列表和所述业务数据写入所述DMA空间。

15.根据权利要求14所述的装置,其特征在于,所述第二类处理器包括一个或多个处理器;

所述硬件调度模块,具体用于从所述第二类处理器中确定一个或多个可用处理器,所述可用处理器为所述第二类处理器中符合调度需求的处理器;

所述硬件调度模块,还具体用于将所述函数列表和所述业务数据写入所述DMA空间中

的就绪队列ReadyQ，一个ReadyQ与一个可用处理器关联，所述一个ReadyQ用于存储所述一个可用处理器的待处理任务的信息。

16.根据权利要求15所述的装置，其特征在于，所述服务器中存储有状态信息，所述状态信息用于指示所述一个或多个处理器的线程标识ID，和挂载在所述处理器中的空闲用户态线程的数量；

所述调度需求是指处理器的线程ID为非空、且挂载在所述处理器中的空闲用户态线程的数量达到第一阈值。

17.根据权利要求10-16中任一项所述的装置，其特征在于，

所述硬件调度模块，还用于：确定所述第二类处理器中可用处理器的数量；若所述可用处理器的数量小于或等于第二阈值，启动所述第二类处理器中除所述可用处理器外的一个或多个其他处理器。

18.根据权利要求10-17所述的装置，其特征在于，所述第一业务所需的数据包括所述第一业务的业务数据，和处理所述业务数据的函数；

所述业务处理模块，具体用于依据所述DMA空间中的用户态线程调度策略，确定所述第一业务的调度计划；所述用户态线程调度策略包含多个函数的执行顺序，以及所述DMA空间提供给所述第二类处理器中任一处理器的存储资源阈值；

所述业务处理模块，还具体用于依据所述调度计划调度所述处理所述业务数据的函数，对所述业务数据进行处理，输出所述第一业务的处理结果。

19.一种服务器，其特征在于，所述服务器包括：内存、第一类处理器和第二类处理器，所述第二类处理器的性能优于所述第一类处理器；

所述第一类处理器，用于获取用于请求第一业务的第一消息；

所述第一类处理器，还用于解析所述第一消息，将所述第一业务所需的数据写入直接内存访问DMA空间；所述DMA空间是由所述内存基于DMA技术提供的存储空间；

所述第二类处理器，用于从所述DMA空间中读取所述数据，输出所述第一业务的处理结果。

20.一种计算机可读存储介质，其特征在于，所述存储介质中存储有计算机程序或指令，当所述计算机程序或指令被服务器执行时，实现权利要求1至9中任一项所述的方法。

一种业务处理方法及装置

技术领域

[0001] 本申请涉及计算机技术领域,尤其涉及一种业务处理方法及装置。

背景技术

[0002] 远程过程调用(remote procedure call,RPC)协议是一种源服务器通过网络从目的服务器上请求服务,且源服务器不需要了解目的服务器的底层网络的技术的协议。如多个服务器基于RPC协议进行数据的处理,如数据访问或数据计算等业务。通常,源服务器向目的服务器发起RPC请求,目的服务器在接收到该RPC请求后执行RPC请求指示的业务。而目的服务器为获取执行前述任务所需的数据,一般会产生中断信号以中断本地的操作系统(operation system,OS)业务,以确定前述数据的物理地址,进而,目的服务器依据该物理地址指示的数据完成RPC请求指示的业务;待目的服务器执行完前述业务后,再次执行本地的OS业务,即本地的OS业务因RPC请求指示的业务产生中断,导致目的服务器的业务处理时延较高。因此,提供一种更有效的业务处理方法成为目前亟待解决的问题。

发明内容

[0003] 本申请实施例提供一种业务处理方法及装置,解决了服务器的业务处理时延较高的问题。

[0004] 本申请采用如下技术方案。

[0005] 第一方面,本申请提供了一种业务处理方法,该业务处理方法应用于服务器,或支持实现该服务器实现业务处理方法的物理设备,例如该物理设备包括芯片系统。该服务器包括第一类处理器,以及性能优于第一类处理器的第二类处理器。示例的,该业务处理方法由前述的服务器执行,该业务处理方法包括:首先,第一类处理器获取用于请求第一业务的第一消息;其次,第一类处理器解析第一消息,将第一业务所需的数据写入直接内存访问(direct memory access,DMA)空间,该DMA空间是由服务器的内存基于DMA技术提供的存储空间;最后,第二类处理器从DMA空间中读取前述数据,输出第一业务的处理结果。

[0006] 在本实施例中,由于第一类处理器将第一业务所需的数据写入DMA空间的过程中,未向第二类处理器发送中断信号,因此,第二类处理器无需中断OS业务,减少了第一类处理器调度第二类处理器的时间,从而降低了第二类处理器执行第一业务的处理时延。另外,由于第一业务所需的数据被写入DMA空间,该DMA空间可以直接被硬件进行数据访问,如第一类处理器和第二类处理器,因此,在第一业务的处理过程中,第二类处理器无需对该第一业务所需数据的物理地址进行转换,避免了物理地址转换导致的业务处理时延,这进一步减少了服务器处理业务的整体时延,提高了服务器处理业务的性能。

[0007] 在一种可选的实现方式中,第一类处理器获取用于请求第一业务的第一消息,包括:第一类处理器从服务器的网卡中读取第一消息。由第一类处理器从网卡中读取接收到的第一消息,避免了网卡主动发送第一消息的过程,降低了网卡在服务器内部的收发过程,提高了网卡对远端进行数据收发的效率,从而,提升了服务器在单位时间内的数据吞吐量,

从而提高了服务器的业务处理效率。

[0008] 或者,第一类处理器接收网卡发送的第一消息。由网卡依据硬件逻辑向第一类处理器发送第一消息,避免了第一类处理器对网卡进行多次轮询,减少了第一类处理器的计算资源消耗,提高了第一类处理器处理第一业务过程中,解析第一消息的效率,进而,提高了服务器处理第一业务的效率。

[0009] 在另一种可选的实现方式中,该业务处理方法还包括:第一类处理器对内存进行初始化,确定DMA空间。由第一类处理器确定DMA空间,可以避免第二类处理器消耗计算资源执行该DMA空间的确定过程,提高第二类处理器222执行其他业务的效率。

[0010] 或者,第二类处理器对内存进行初始化,确定DMA空间。由第二类处理器确定DMA空间,可以避免在第一类处理器调度第二类处理器执行业务的过程中,还要消耗计算资源执行该DMA空间的确定过程,进而,提高第一类处理器221的调度效率。

[0011] 另外,DMA空间也可以是第一类处理器和第二类处理器协同实现的。

[0012] 在另一种可选的实现方式中,前述的第一业务包括一个或多个任务。一个任务的信号量增加线程和信号量减少线程挂载于第二类处理器的同一个处理器。如此,该任务不需要由第二类处理器包含的两个处理器进行跨核协同,减少了该任务执行时所需的处理器通信时间,从而降低了该任务的处理时延,进而,提高了服务器执行包含该任务的业务的执行效率。

[0013] 在另一种可选的实现方式中,第一业务所需的数据包括第一业务的业务数据,和处理业务数据的函数。第一类处理器解析第一消息,将第一业务所需的数据写入DMA空间,包括:①,第一类处理器依据第一消息确定函数标识,该函数标识用于指示执行第一业务所需的函数列表,该函数列表包括一个或多个函数。②,第一类处理器将函数列表和业务数据写入DMA空间。

[0014] 在另一种可选的实现方式中,第二类处理器包括一个或多个处理器。第一类处理器将函数列表和业务数据写入DMA空间,包括:第一类处理器从第二类处理器中确定一个或多个可用处理器,并将函数列表和业务数据写入DMA空间中的就绪队列(Ready queue,ReadyQ)。前述的可用处理器为第二类处理器中符合调度需求的处理器,且,一个ReadyQ与一个可用处理器关联,该一个ReadyQ用于存储一个可用处理器的待处理任务的信息。

[0015] 在一种可能的示例中,服务器中存储有状态信息,该状态信息用于指示一个或多个处理器的线程标识(identifier,ID),和挂载在处理器中的空闲用户态线程的数量。前述的调度需求是指处理器的线程ID为非空、且挂载在处理器中的空闲用户态线程的数量达到第一阈值。在本实施例中,第一类处理器可以依据第二类处理器中各线程ID是否为空,对第二类处理器包含的处理器进行调度,避免了第二类处理器启用轮询(Poll)线程来查询待处理任务,避免了在服务器中无第二类处理器的待处理任务时,第二类处理器启用Poll线程导致的空转,降低了第二类处理器的计算资源消耗,提高了第二类处理器执行其他业务的效率,从而提升了服务器执行业务的整体性能。

[0016] 在另一种可选的实现方式中,该业务处理方法还包括:第一类处理器确定第二类处理器中可用处理器的数量;若可用处理器的数量小于或等于第二阈值,第一类处理器启动第二类处理器中除可用处理器外的一个或多个其他处理器。由第一类处理器对第二类处理器中可用处理器的数量进行监控(或监听),使得第二类处理器中的可用处理器的数量达

到第二阈值，提高服务器执行业务的总体性能。

[0017] 在另一种可选的实现方式中，第一业务所需的数据包括第一业务的业务数据，和处理业务数据的函数。第二类处理器从DMA空间中读取数据，输出第一业务的处理结果，包括：第二类处理器依据DMA空间中的用户态线程调度策略，确定第一业务的调度计划，进而，第二类处理器依据该调度计划调度处理业务数据的函数，对业务数据进行处理，输出第一业务的处理结果。前述的用户态线程调度策略包含多个函数的执行顺序，以及DMA空间提供给第二类处理器中任一处理器的存储资源阈值。

[0018] 在本实施例中，第二类处理器可以依据该用户态线程的调度策略，确定多个函数的执行顺序，避免了第二类处理器中的多个处理器在执行第一业务的过程中发生混乱，提高了服务器执行第一业务的性能。

[0019] 第二方面，本申请提供了一种业务处理装置，所述业务处理装置包括用于执行第一方面或第一方面任一种可能设计中的业务处理方法的各个模块。

[0020] 示例的，该业务处理装置应用于服务器，该服务器包括第一类处理器和第二类处理器，第二类处理器的性能优于第一类处理器，该业务处理装置包括：应用于第一类处理器的硬件调度模块，和应用于第二类处理器的业务处理模块。

[0021] 硬件调度模块，用于获取用于请求第一业务的第一消息。

[0022] 硬件调度模块，还用于解析第一消息，将第一业务所需的数据写入DMA空间。该DMA空间是由服务器的内存基于DMA技术提供的存储空间。

[0023] 业务处理模块，用于从DMA空间中读取数据，输出第一业务的处理结果。

[0024] 另外，当业务处理装置用于实现上述第一方面中任一方面的方法实施例时，有益效果可以参见第一方面中任一方面的描述，此处不再赘述。所述业务处理装置具有实现上述第一方面中任一方面的方法实例中行为的功能。所述功能可以通过硬件实现，也可以通过硬件执行相应的软件实现。所述硬件或软件包括一个或多个与上述功能相对应的模块。

[0025] 第三方面，本申请提供了一种服务器，该服务器包括：内存、第一类处理器和第二类处理器，该第二类处理器的性能优于第一类处理器。

[0026] 第一类处理器，用于获取用于请求第一业务的第一消息。

[0027] 第一类处理器，还用于解析第一消息，将第一业务所需的数据写入DMA空间。该DMA空间是由内存基于DMA技术提供的存储空间。

[0028] 第二类处理器，用于从DMA空间中读取数据，输出第一业务的处理结果。

[0029] 另外，该服务器还可以通过逻辑电路或执行代码指令用于实现第一方面和第一方面中任一种可能实现方式中的方法。当服务器用于实现上述第一方面中任一方面的方法实施例时，有益效果可以参见第一方面中任一方面的描述，此处不再赘述。

[0030] 第四方面，本申请提供一种计算机可读存储介质，包括：计算机软件指令；当计算机软件指令在计算设备中运行时，使得计算设备执行如第一方面或第一方面任意一种可能的实现方式中所述方法的操作步骤。

[0031] 第五方面，本申请提供一种计算机程序产品，当计算机程序产品在计算机上运行时，使得计算设备执行如第一方面或第一方面任意一种可能的实现方式中所述方法的操作步骤。

[0032] 第六方面，本申请提供一种芯片，包括控制电路和接口电路，接口电路用于接收来

自计算设备之外的其它计算设备的信号并传输至处理器，或将来自处理器的信号发送给计算设备之外的其它计算设备，处理器通过逻辑电路或执行代码指令用于实现上述第一方面或第一方面中任一种可能实现方式中方法的操作步骤。

[0033] 第七方面，本申请提供一种计算设备，该计算设备包括存储器和至少两个不同的处理器，存储器用于存储一组计算机指令；当处理器执行所述一组计算机指令时，执行第一方面或第一方面任一种可能实现方式中的业务处理方法的操作步骤。

[0034] 本申请在上述各方面提供的实现方式的基础上，还可以进行进一步组合以提供更多实现方式。

附图说明

[0035] 图1为本申请提供的一种通信系统的应用场景图；

[0036] 图2为本申请提供的一种服务器的结构示意图；

[0037] 图3为本申请提供的一种业务处理方法的流程示意图一；

[0038] 图4为本申请提供的一种业务处理方法的流程示意图二；

[0039] 图5为本申请提供的一种业务处理方法的流程示意图三；

[0040] 图6为本申请提供的一种业务处理装置的结构示意图。

具体实施方式

[0041] 本申请提供一种业务处理方法，该业务处理方法由服务器执行，该服务器包括第一类处理器，以及性能优于第一类处理器的第二类处理器，该业务处理方法包括：首先，第一类处理器获取用于请求第一业务的第一消息；其次，第一类处理器解析第一消息，将第一业务所需的数据写入DMA空间，该DMA空间是由服务器的内存基于DMA技术提供的存储空间；最后，第二类处理器从DMA空间中读取前述数据，输出第一业务的处理结果。

[0042] 在本实施例中，由于第一类处理器将第一业务所需的数据写入DMA空间的过程中，未向第二类处理器发送中断信号，因此，第二类处理器无需中断OS业务，减少了第一类处理器调度第二类处理器的时间，从而降低了第二类处理器执行第一业务的处理时延。另外，由于第一业务所需的数据被写入DMA空间，该DMA空间可以直接被硬件进行数据访问，如第一类处理器和第二类处理器，因此，在第一业务的处理过程中，第二类处理器无需对该第一业务所需数据的物理地址进行转换，避免了物理地址转换导致的业务处理时延，这进一步减少了服务器处理业务的整体时延，提高了服务器处理业务的性能。

[0043] 为了下述各实施例的描述清楚简洁，首先给出相关技术的简要介绍。

[0044] 图1为本申请提供的一种通信系统的应用场景图，该通信系统包括计算机集群(computer cluster)110和客户端120，计算机集群110可以通过网络130与客户端120进行通信，网络130可以是因特网，或其他网络(如以太网)。网络130可以包括一个或多个网络设备，网络设备可以是路由器或交换机等。

[0045] 客户端120可以是运行有应用程序的计算机，该运行有应用程序的计算机可以是物理机，也可以是虚拟机。

[0046] 例如，若该运行有应用程序的计算机为物理机，该物理机可以是服务器或终端(terminal)。其中，终端也可以称为终端设备、用户设备(user equipment，UE)、移动台

(mobile station,MS)、移动终端(mobile terminal,MT)等。终端可以是手机、平板电脑、笔记本电脑、桌面电脑、台式计算机、虚拟现实(virtual reality,VR)终端设备、增强现实(augmented reality,AR)终端设备、工业控制中的无线终端、无人驾驶中的无线终端、远程手术(remote medical surgery)中的无线终端、智能电网(smart grid)中的无线终端、运输安全(transportation safety)中的无线终端、智慧城市(smart city)中的无线终端、智慧家庭(smart home)中的无线终端等等。本申请的实施例对客户端120所采用的具体技术和具体设备形态不做限定。

[0047] 又如,该运行有应用程序的计算机为虚拟机,该客户端120可以是运行在计算机集群110中任意一个或多个服务器上的软件模块。

[0048] 计算机集群110是指通过局域网或互联网连接的计算机集合,通常用于执行大型任务(也可以称为作业(job)),这里的作业通常是需要较多计算资源并行处理的大型作业,本实施例不限定作业的性质和数量。一个作业可能包含多个计算任务,这些任务可以分配给多个计算资源执行。大多数任务是并发或并行执行的,而有一些任务则需要依赖于其他任务所产生的数据。计算机集群中每台服务器使用相同的硬件和相同的操作系统;或者,可以根据业务需求,计算机集群的服务器中采用不同的硬件和不同的操作系统。由计算机集群110包含的多个服务器协同执行任务,避免了该任务仅由一个服务器执行导致的处理时延较大,且该任务可以由多个服务器并发执行,提高了任务的执行效率。

[0049] 如图1所示,计算机集群110包括多个服务器,例如服务器111～服务器114,各个服务器可以提供计算资源。就一台服务器来说,它可以包含多个处理器或处理器核,每个处理器或者处理核可以是一个计算资源,因此一台服务器可以提供多个计算资源。例如,服务器可以是指客户端(client)服务器或服务端(server)服务器。

[0050] 计算机集群110可以处理多种类型的作业。本实施例对任务的数量,以及可以并行执行的任务的数据都没有予以限制。示例的,上述的作业是指虚拟机的热迁移、服务器之间的主从备份、或者基于RPC的远端操作(如数据访问、RPC操作以及相应的回包等)等,如数据备份。该作业可以由客户端生成并提交给计算器集群执行,比如提交给计算机集群中的一个或者多个服务器执行。具体的,本申请不限定提交业务的位置。

[0051] 在图1中,作业可以从客户端120通过网络130提交给计算机集群110。比如作业可以提交给服务器111,在作业提交给服务器111的情况下,服务器111可以用于管理计算机集群110中所有的服务器,以完成该作业所包括的一个或多个任务,如调度其他服务器的计算资源或存储资源等。又比如,作业也可以提交给计算机集群110中的其他服务器,本实施例不限定提交作业的位置。

[0052] 值得注意的是,图1只是示意图,该通信系统中还可以包括其他设备,在图1中未画出,本申请的实施例对该系统中包括的服务器(计算设备)、客户端的数量和类型不做限定。示例的,计算机集群110还可以包括更多或更少的计算设备,如计算机集群110包括两个计算设备,一个计算设备用于实现上述服务器111和服务器112的功能,另一个计算设备用于实现上述服务器113和服务器114的功能。

[0053] 示例性的,图1中的任一个服务器或客户端120可通过图2所示的服务器200来实现。图2为本申请提供的一种服务器200的结构示意图,该服务器200包括基板管理控制器(baseboard management controller,BMC)210、第一类处理器221、第二类处理器222、内存

230、硬盘240和网卡250。

[0054] 基板管理控制器210,可以对设备进行固件升级,对设备的运行状态进行管理以及排除故障等。处理器可通过外围器件互联(Peripheral Component Interconnect express,PCIe)总线、通用串行总线(universal serial bus,USB),或者集成电路总线(Inter-Integrated Circuit,I2C)等总线访问基板管理控制器210。基板管理控制器210还可以和至少一个传感器相连。通过传感器获取计算机设备的状态数据,其中状态数据包括:温度数据,电流数据、电压数据等等。在本申请中不对状态数据的类型做具体限制。基板管理控制器210通过PCIe总线或者其他类型的总线和处理器通信,例如,将获取到的状态数据,传递给处理器进行处理。基板管理控制器210也可以对存储器中的程序代码进行维护,包括升级或恢复等等。基板管理控制器210还可以对服务器200内的电源电路或时钟电路进行控制等。总之,基板管理控制器210可以通过以上方式实现对服务器200的管理。然而,基板管理控制器210只是一个可选设备。在一些实施方式中,处理器可以直接和传感器通信,从而对计算机设备直接进行管理和维护。

[0055] 值得说明的是,服务器中器件的连接方式除了采用上述PCIe总线、USB总线、I2C总线外,还可以通过扩展工业标准结构(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。总线还可以分为地址总线、数据总线、控制总线等。

[0056] 处理器通过双倍速率(double data rate,DDR)总线和内存230相连。这里,不同的内存230可能采用不同的数据总线与处理器通信,因此DDR总线也可以替换为其他类型的数据总线,本申请实施例不对总线类型进行限定。

[0057] 另外,服务器200还包括各种输入输出(input/output,I/O)设备,处理器可以通过PCIe总线访问这些I/O设备。

[0058] 处理器(processor)是服务器200的运算核心和控制核心,如图2所示出的第一类处理器221和第二类处理器222。处理器中可以包括一个或多个处理器核(core),如图2所示出的核221A和核222A。

[0059] 处理器可以是一块超大规模的集成电路。在处理器中安装有操作系统和其他软件程序,从而处理器能够实现对内存230及各种PCIe设备的访问。可以理解的是,在本发明实施例中,处理器可以是中央处理器(central processing unit,CPU),可以是其他特定集成电路(application specific integrated circuit,ASIC)。处理器还可以是其他通用处理器、数字信号处理器(digital signal processing,DSP)、现场可编程门阵列(field programmable gate array,FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。实际应用中,服务器200也可以包括多个处理器。

[0060] 在第一种可能的示例中,第一类处理器221和第二类处理器222是以处理器的性能划分的,如第二类处理器222的性能优于第一类处理器221。

[0061] 在一种可能的情形中,性能是指处理器执行业务的效率。则第二类处理器222的性能优于第一类处理器221是指,第二类处理器222执行业务的效率高于第一类处理器221执行前述业务的效率,如该业务是指整点运算业务、或浮点运算业务等。

[0062] 在另一种可能的情形中,处理器可以提供计算资源、存储资源和传输资源(或称网

络资源),上述的性能包括:存储资源的数据读写速度,计算资源在单位时间内的浮点运算次数和整点运算次数,以及传输资源的带宽中至少一种,其中,该带宽为单位时间内传输数据的最大值。

[0063] 在第二种可能的示例中,第一类处理器221和第二类处理器222是以处理器的架构类型划分的,如第一类处理器221为精简指令集机器(advanced reduced instruction set computer machine,ARM)架构的处理器,第二类处理器222为X86架构的处理器。

[0064] 值得注意的是,上述两种示例仅为本实施例提供的第一类处理器221和第二类处理器222的可能区分方式,不应理解为对服务器200或对本申请的限定。

[0065] 在本实施例中,第一类处理器221和第二类处理器222均可以执行远端请求的全部业务。该远端是指与服务器220通信的其他服务器或客户端,还可以是指与服务器220进行通信的应用程序(application,APP)等。

[0066] 在一些情形中,第一类处理器221可以执行远端请求的一部分业务,且第一类处理器221还可以调度第二类处理器执行远端请求的另一部分业务。另外,第一类处理器221也可以调度第二类处理器执行远端请求的所有业务,本申请对此不予限定。

[0067] 可选的,第一类处理器221可以称为异构单元或多线程小核等,第二类处理器222可以称为业务单元或高性能大核等。

[0068] 内存230,也称为主存(main memory)。内存230通常用来存放操作系统中各种正在运行的软件、输入和输出数据以及与外存交换的信息等。为了提高处理器的访问速度,内存230需要具备访问速度快的优点。在传统的计算机系统架构中,通常采用动态随机存取存储器(dynamic random access memory,DRAM)作为内存230。处理器能够通过内存控制器高速访问内存230,对内存230中的任意一个存储单元进行读操作和写操作。除了DRAM之外,内存230还可以是其他随机存取存储器,例如静态随机存取存储器(static random access memory,SRAM)等。本实施例不对内存230的数量和类型进行限定。此外,可对内存230进行配置使其具有保电功能。保电功能是指系统发生掉电又重新上电时,存储器中存储的数据也不会丢失。具有保电功能的内存230被称为非易失性存储器。

[0069] 在本实施例中,处理器可以对内存230进行管理,如处理器可以基于DMA技术将内存230的至少部分存储空间设置为直接内存访问空间(DMA空间),该DMA空间可以保存处理器执行业务所需的业务数据或函数等信息。

[0070] 在第一种可能的情形中,第一类处理器221对内存230进行初始化,获取前述的DMA空间。由第一类处理器221确定DMA空间,可以避免第二类处理器222消耗计算资源执行该DMA空间的确定过程,提高第二类处理器222执行其他业务的效率。

[0071] 在第二种可能的情形中,第二类处理器222对内存230进行初始化,获取前述的DMA空间。由第二类处理器222确定DMA空间,可以避免在第一类处理器221调度第二类处理器222执行业务的过程中,还要消耗计算资源执行该DMA空间的确定过程,进而,提高第一类处理器221的调度效率。

[0072] 以上两种可能的情形仅为本实施例提供的两种可能的示例,在另一些可能的情形中,该DMA空间还可以是由其他处理器或计算设备对该内存230进行处理后获得的,本申请对此不予限定。

[0073] 在一种可选的实现方式中,处理器可以将内存230的部分或全部存储空间进行虚

拟化获得一个存储资源池，以便存储服务器200在运行过程中所需的其他数据。如内存230包括多个内存页(page)，内存页是内存230的数据I/O操作的最小单位，内存页也称为数据读写的原子单位。每个内存页对应内存230的一段存储地址空间，如一个内存页可以用于存储4千字节(kilo bytes,KB)的数据，则该内存页对应4KB的存储地址空间。另外，一个内存页也可以对应更大或更小的存储地址空间，如2KB或8KB等。

[0074] 在本实施例中，处理器可以将内存230的一部分存储空间设置为DMA空间，将内存230的另一份存储空间设置为存储资源池，本申请对此不予限定。

[0075] I/O设备是指可以进行数据传输的硬件，也可以理解为与I/O接口对接的设备。常见的I/O设备有网卡、打印机、键盘、鼠标等，如I/O设备可以是图2所示出的网卡250。所有的外存也可以作为I/O设备，如硬盘、软盘、光盘等。处理器可通过PCIe总线访问各个I/O设备。需要说明的是，PCIe总线只是其中的一个示例，可以被替换为其他总线，例如UB总线或CXL等。

[0076] 如图2所示，网卡250包括处理器251、存储器252和通信接口253。在一些可能的示例中，包含处理单元和网络适配器(network interface card,NIC)的网卡也被称为智能网卡(intelligent NIC,iNIC)。

[0077] 处理器251是指具有处理能力的处理器，例如数据处理单元(data processing unit,DPU)。DPU具有CPU的通用性和可编程性，但更具有专用性，可以在网络数据包，存储请求或分析请求上高效运行。DPU通过较大程度的并行性(需要处理大量请求)与CPU区别开来。可选的，这里的DPU也可以替换成图形处理单元(graphics processing unit,GPU)、嵌入式神经网络处理器(neural-network processing units,NPU)等处理芯片。

[0078] 存储器252可以是指与处理器251直接交换数据的内部存储器，它可以随时读写数据，而且速度很快，作为操作系统或其他正在运行中的程序的临时数据存储器。存储器252包括至少两种存储器，例如存储器252既可以是随机存取存储器，也可以是ROM。举例来说，随机存取存储器是DRAM，或者存储级存储器(storage class memory,SCM)。DRAM是一种半导体存储器，与大部分RAM一样，属于一种易失性存储器(volatile memory)设备。SCM是一种同时结合传统储存装置与存储器特性的复合型储存技术，存储级存储器能够提供比硬盘更快速的读写速度，但存取速度上比DRAM慢，在成本上也比DRAM更为便宜。然而，DRAM和SCM在本实施例中只是示例性的说明，存储器252还可以包括其他随机存取存储器，例如SRAM等。而对于只读存储器，举例来说，可以是PROM、EPROM等。另外，存储器252还可以是双列直插式存储器模块或双线存储器模块(dual in-line memory module,DIMM)，即由DRAM组成的模块，还可以是固态硬盘(solid state disk,SSD)。实际应用中，网卡250中可配置多个存储器252，以及不同类型的存储器252。本实施例不对存储器252的数量和类型进行限定。此外，可对存储器252进行配置使其具有保电功能。保电功能是指系统发生掉电又重新上电时，存储器252中存储的数据也不会丢失。具有保电功能的存储器被称为非易失性存储器。

[0079] 在一种可能的情形中，存储器252中存储有软件程序，处理器251运行存储器252中的软件程序可实现数据访问(如数据同步)、计算业务或者虚拟机迁移的管理等。

[0080] 通信接口253用于实现的服务器200与其他设备进行通信的网络接口卡，例如，网络适配器253可以实现从并行到串行的数据转换、数据包的装配和拆装、网络存取控制、数据缓存和网络信号中的一种或多种功能。

[0081] 下面将结合附图对本实施例提供的业务处理方法的具体实现方式进行详细描述。

[0082] 如图3所示,图3为本申请提供的一种业务处理方法的流程示意图一,该业务处理方法可以应用于图1所示出的通信系统,该业务处理方法可以由服务器执行,如图1所示出的服务器111～服务器114中的任一个,或者客户端120,或者图2所示出的服务器200,这里以服务器200来执行本实施例的业务处理方法为例进行说明。

[0083] 如图3所示,本实施例提供的业务处理方法包括以下步骤。

[0084] S310,第一类处理器221获取用于请求第一业务的第一消息。

[0085] 示例的,该第一业务是指由客户端应用(client APP)发起的业务,该client APP是指大数据应用、人工智能应用或云计算应用等。例如,第一业务是指RPC请求指示的业务,如远程直接内存操作(remote direct memory operation,RDMO);又如,第一业务是指基于远程直接内存访问(remote direct memory access,RDMA)技术的单边RDMA操作,如RDMAread操作和RDMAwrite操作,只需要本端明确信息的源地址和目的地址,远端应用不必感知此次通信,数据的读或存都通过远端的网卡完成,再由远端网卡封装成消息返回到本端。

[0086] 可选的,第一业务包括一个或多个任务,这里的任务是指数据访问或基于数据的计算任务等。

[0087] 该第一消息是指携带有第一业务的标识或指令的信息。示例的,该第一消息可以是服务器200的网卡250从外部设备接收到的。如网卡250接收和发送数据通常采用消息队列的方式,消息队列包括一组队列对(queue pair,QP),QP包括发送队列和接收队列,如网卡250中用于发送数据的消息队列是发送队列(send queue,SQ),网卡250中用于接收数据的消息队列是接收队列(receive queue,RQ)。另外,网卡250中还维护有完成队列(completion queue,CQ)。消息队列是多个服务器之间进行通信所采用的连接方式,例如,多个服务器之间可以利用传输控制协议/网络之间互连协议(transmission control protocol/internet protocol,TCP/IP)建立多条连接,每条连接都有一个接收队列和发送队列,该接收队列和发送队列用于传输该连接的数据。消息队列中存储有工作队列元素(work queue element,WQE),WQE中存储有指向发送队列、或接收队列数据的地址和长度的信息,数据的长度可以由数据的地址和偏移量来确定,WQE中指示了数据的地址和偏移量的信息又被称为散集(scatter/gather,SG)信息。如果一组数据包括多段的数据,一段数据的SG信息包括该数据的地址和偏移量,则WQE中关于该组数据的多个SG信息也可称为SG的链,或称散集列表(scatter/gather list,SGL)。例如,发送队列中的WQE为发送队列条目(send queue entry,SQE)、接收队列中的WQE为接收队列条目(receive queue entry,RQE),完成队列中的WQE为完成队列条目(completion queue entry,CQE)。关于SQ、RQ和CQ更详尽的描述可以参考通常技术的相关内容,此处不予赘述。

[0088] 以下给出两种可行的情形,来说明第一类处理器221获取第一消息的具体实现过程。

[0089] 在第一种可行的情形中,第一类处理器221从服务器200的网卡250中读取第一消息。

[0090] 例如,第一类处理器221中挂载有轮询(Poll)线程,该Poll线程可以从网卡250维护的CQ中读取CQE,并确定与该CQE关联的RQE,进而,从RQE指示的地址中读取网卡250获取

的第一消息。由第一类处理器221从网卡250中读取接收到的第一消息，避免了网卡250主动发送第一消息的过程，降低了网卡250在服务器200内部的收发过程，提高了网卡250对远端进行数据收发的效率，从而，提升了服务器200在单位时间内的数据吞吐量，从而提高了服务器的业务处理效率。

[0091] 在第二种可行的情形中，第一类处理器221接收网卡250发送的第一消息。例如，网卡250中设置有硬件逻辑，网卡250基于该硬件逻辑主动将第一消息发送到第一类处理器221。如此，由网卡250依据硬件逻辑向第一类处理器221发送第一消息，避免了第一类处理器221对CQ进行多次轮询，减少了第一类处理器221的计算资源消耗，提高了第一类处理器221处理第一业务过程中，解析第一消息的效率，进而，提高了服务器200处理第一业务的效率。

[0092] 以上两种可行的情形仅为本实施例提供的示例，不应理解为对本申请的限定。

[0093] 请继续参见图3，本实施例提供的业务处理方法还包括以下步骤。

[0094] S320，第一类处理器221解析第一消息，将第一业务所需的数据写入DMA空间。

[0095] 可选的，第一业务所需的数据包括第一业务所需的业务数据，以及处理该业务数据所需的函数。其中，前述的函数包括执行函数和管理函数中至少一种。执行函数包括但不限于字符串操作函数、数学函数、进程操作函数、信号处理函数、环境变量函数和字符检测函数等；管理函数包括但不限于内存管理函数、文件权限控制函数和接口处理函数等。

[0096] 其中，该DMA空间是由服务器200的内存230基于DMA技术提供的存储空间。第一类处理器221和第二类处理器222可以对该DMA空间进行直接访问，不需要进行地址转换等操作，减少了处理器对DMA空间进行数据访问的时延，进而，提高了第一业务的处理效率。

[0097] 针对于上述的第一类处理器221将第一业务所需的数据写入DMA空间，在图3的基础上，本实施例提供一种可能的具体实现方式，如图4所示，图4为本申请提供的一种业务处理方法的流程示意图二，第一类处理器221中挂载有多个线程：硬件调度线程、并发控制线程、一个或多个执行线程，硬件线程和/或轮询线程。

[0098] 其中，硬件调度线程用于解析第一消息，以及将第一业务所需的数据写入DMA空间；并发控制线程用于将第一业务划分为一个或多个任务；执行线程用于执行硬件调度线程分配的一个或多个任务。

[0099] 示例的，并发控制线程将第一业务划分为一个或多个任务，硬件调度线程对业务包含的任务进行区分，小粒度任务由第一类处理器221包含的执行线程处理，大粒度任务由第二类处理器222执行。如此，在不影响原本大粒度任务的执行效率的前提下，将小粒度任务交给执行线程处理，充分利用了第一类处理器221中执行线程的高并发，也避免了第一类处理器221中单个处理器的处理性能较差，影响大粒度任务的执行效率，从而提升了服务器200执行业务的整体性能。

[0100] 硬件线程用于接收网卡250依据硬件逻辑发送的CQ中的数据或指令，如上述的第一消息。轮询(Poll)线程用于从网卡250维护的CQ中读取前述的第一消息。

[0101] 在本实施例中，第一类处理器221中可以同时挂载有硬件线程和轮询线程，并由硬件线程或轮询线程获取前述的第一消息。但在另一些可能的情形中，第一类处理器221中可以仅挂载有硬件线程和轮询线程中的一个线程，并由挂载在第一类处理器221的线程获取前述的第一消息。

[0102] 另外,第一类处理器221中还维护有CPU调度策略,该CPU调度策略指示了第一类处理器221中可供使用的处理器,以及该可供使用的处理器的剩余可用资源(如计算资源、存储资源和传输资源)等;另外,该CPU调度策略还指示了第二类处理器222中可供第一类处理器221调度的处理器,以及可供调度的处理器的剩余可用资源等信息。

[0103] 值得注意的是,CPU调度策略中还可以记录有其他处理器相关的信息,本申请对此不予限定。

[0104] 如图4所示,存储区域包括前述的DMA空间,以及网卡维护的消息队列(QP1),该QP1包括SQ、RQ和CQ,其中,SQ用于存储服务器200将向外发送的数据。RQ用于存储服务器200接收的远端数据或请求,CQ用于记录SQ或RQ中已完成的工作队列元素的信息。

[0105] 图4的存储区域中仅示出了一个QP,但在一些示例中,存储区域中可以包括多个QP,每个QP中均包含SQ、RQ和CQ,由网卡250维护的这多个RQ可以映射到一个门铃寄存器,当第一类处理器221从网卡250中读取数据时均可通过该门铃寄存器,避免一个RQ关联一个门铃寄存器时,RQ中未存储有待读取数据所导致的Poll线程空转。另外,若由网卡250通过前述的门铃寄存器主动向第一类处理器221发送数据,可以避免第一类处理器221的Poll线程消耗第一类处理器221的计算资源,提高第一类处理器221执行业务的效率,从而提高服务器220的业务处理性能。

[0106] 在硬件实现上,存储区域包括的存储空间可以是由服务器200的内存230提供的。

[0107] 在一种可选的情况中,存储区域包括的存储空间也可以是由内存230和网卡250提供的。如,内存230基于DMA技术为DMA空间提供存储空间,网卡250包括的存储器252为消息队列提供存储空间。

[0108] 另外,图4所示出的存储区域中还可以存储有其他数据,本申请对此不予限定。

[0109] 请继续参见图4,DMA空间中存储有一个或多个处理器的就绪队列(ReadyQ)、CPU上下文,以及用户态线程调度策略。这里的处理器是指可对该DMA空间进行数据访问的任一个处理器,如第一类处理器221包括的处理器,或第二类处理器222包括的处理器。

[0110] 一个ReadyQ关联一个处理器,该ReadyQ用于存储与该ReadyQ关联的处理器的待处理任务的信息。待处理任务的信息可以包括,但不限于:执行该待处理任务所需的业务数据,以及函数等。在一些情况下,ReadyQ也可以称为该处理器的CPU运行队列。

[0111] 一个CPU上下文关联一个处理器,该CPU上下文至少包含处理器中的线程ID,如图4所示出的CPU1上下文和CPU2上下文。由于该CPU上下文包含了该处理器中所有线程的线程ID,因为,该CPU上下文实际上还记录了该处理器中线程的数量。另外,线程ID仅为本实施例为说明CPU上下文中记录处理器的线程的示例性说明,在一些情形中,线程ID也可以是指一个标签(tag),或者一串特殊的唯一字符串,该唯一字符串用于区分处理器中不同的线程。另外,CPU上下文中还可以记录有与该CPU上下文关联的处理器的其他信息,本申请对此不予限定。

[0112] 用户态线程调度策略包含多个函数的执行顺序,如服务器200可以支持实现N个函数,该用户态线程调度策略可以包含这N个函数的执行顺序(或执行逻辑),或者,N个函数中M个函数(M<N)的执行顺序(或执行逻辑)。在本实施例中,第二类处理器可以依据该用户态线程的调度策略,确定多个函数的执行顺序,避免了第二类处理器中的多个处理器在执行第一业务的过程中发生混乱,提高了服务器执行第一业务的性能。

[0113] 用户态线程调度策略还包含DMA空间提供给第二类处理器222中任一处理器的存储资源阈值。如20百万字节(mega bytes,MB)、20KB或其他等。如此,避免了第二类处理器222中任一处理器无限制的调度DMA空间所提供的存储资源,导致其他任务的处理时延增加,降低了服务器执行第一业务的整体时延。值得注意的是,该用户态线程调度策略可以是服务器200预设的,也可以是用户或客户端通过数据库匹配导入到前述的DMA空间的,本申请对此不予限定。

[0114] 如图4所示,S320中第一类处理器221将第一业务所需的数据写入DMA空间的过程,可以包括以下的S321和S322。

[0115] S321,第一类处理器221依据第一消息确定函数标识。

[0116] 例如,第一类处理器221可以将第一消息与函数库进行匹配,确定与第一业务关联的函数标识。

[0117] 该函数标识用于指示执行第一业务所需的函数列表,该函数列表包括一个或多个函数。例如,函数列表包括一个执行函数。又如,函数列表包括一个管理函数和多个执行函数。函数列表所包括的函数可以存储在函数库中。

[0118] 本申请实施例中,该函数库还可以维护有执行其他业务(如第二业务)所需的函数。

[0119] 在一种可能的情形中,函数库存储在前述的DMA空间中。在另一种可能的情形中,函数库存储在存储区域中除前述的DMA空间的其他存储空间中。在又一种可能的情形中,函数库存储在第一类处理器221和第二类处理器222可直接访问的数据库中,该数据库的存储功能可以由服务器200包括的硬盘240提供,也可以是由与服务器200通信的其他存储设备提供,本申请对数据库的位置不做限定。

[0120] 以上函数库的三种存储情形仅为本实施例提供的示例,不应理解为对本申请的限定。例如,函数库也可以是由一个分布式存储系统维护的。

[0121] 可选的,若第一消息中指示第一业务需并发(parallel)处理,则并发调度线程可以依据服务器提供给第一业务的并发度,确定执行第一业务所需的多个函数,比如并发执行第一业务所需的一个管理函数和多个执行函数。服务器提供给第一业务的并发度是指是服务器中可供第一业务所使用的处理器的最大数量。

[0122] S322,第一类处理器221根据S321中确定的函数标识,将函数标识所标识的函数列表和业务数据写入DMA空间。

[0123] 示例的,第一类处理器221从第二类处理器222中确定一个或多个可用处理器,并将函数列表和业务数据写入DMA空间中与前述可用处理器关联的ReadyQ。

[0124] 其中,前述的可用处理器为第二类处理器222中符合调度需求的处理器。

[0125] 可选的,服务器中存储有状态信息,该状态信息用于指示一个或多个处理器的线程ID,和挂载在处理器中的空闲用户态线程的数量。示例的,该状态信息可以包括前述的CPU上下文。在另一种可能的示例中,状态信息还可以包括前述的CPU调度策略。

[0126] 例如,调度需求是指处理器中的线程ID为非空、且挂载在处理器中的空闲用户态线程的数量达到第一阈值,如该第一阈值为1、2或其他等。其中,处理器中的线程ID为非空是指该处理器可供第一类处理器221进行调度,挂载在处理器中的空闲用户态线程的数量达到第一阈值是指,该处理器具有空闲的用户态线程可以用于执行第一业务的相关任务。

[0127] 在本实施例中，第一类处理器221可以依据第二类处理器222中各线程ID是否为空，对第二类处理器222包含的处理器进行调度，避免了第二类处理器222启用Poll线程来查询待处理任务，避免了在服务器200中无第二类处理器222的待处理任务时，第二类处理器222启用Poll线程导致的空转，降低了第二类处理器222的计算资源消耗，提高了第二类处理器222执行其他业务的效率，从而提升了服务器200执行业务的整体性能。

[0128] 在一种可选的实现方式中，第一类处理器221还可以确定第二类处理器中可用处理器的数量，若可用处理器的数量小于或等于第二阈值(如5个)，第一类处理器221启动第二类处理器222中除可用处理器外的一个或多个其他处理器。

[0129] 示例的，若状态信息包含第二类处理器222中10个处理器的线程ID，该10个线程ID中具有3个状态为非空的线程ID、且这3个线程ID对应的处理器中空闲用户态线程达到第一阈值，则该10个处理器包含3个可用处理器，假设第二阈值为5个，则第一类处理器221可以启动该10个处理器包含的其他2个处理器的OS业务线程，以使该其他2个处理器可以提供给第一类处理器221进行调度，从而，提升服务器200执行业务的整体效率。

[0130] 请继续参见图3，本实施例提供的业务处理方法还包括以下步骤。

[0131] S330，第二类处理器222从DMA空间中读取数据，输出第一业务的处理结果。

[0132] 示例的，第二类处理器222依据DMA空间中的用户态线程调度策略：确定第一业务的调度计划，以及依据该调度计划调度处理业务数据的函数，对业务数据进行处理，输出第一业务的处理结果。

[0133] 如图4所示，第二类处理器222输出处理结果可以包括：第二类处理器222将处理结果写入存储区域包括的发送队列(SQ)，并由网卡250发送该SQ中存储的处理结果。由存储区域维护的SQ来对第一业务的处理结果进行存储，并由服务器包括的网卡或其他通信接口来发送该处理结果，减少第一业务的处理结果被服务器向外发送的次数，从而节省服务器的传输资源。

[0134] 在本实施例中，由于第一类处理器将第一业务所需的数据写入DMA空间的过程中，未向第二类处理器发送中断信号，因此，第二类处理器无需中断OS业务，减少了第一类处理器调度第二类处理器的时间，从而降低了第二类处理器执行第一业务的处理时延。另外，由于第一业务所需的数据被写入DMA空间，该DMA空间可以直接被硬件进行数据访问，如第一类处理器和第二类处理器，因此，在第一业务的处理过程中，第二类处理器无需对该第一业务所需数据的地址进行转换，避免了地址转换导致的业务处理时延，这进一步减少了服务器执行第一业务的整体时延，提高了服务器处理业务的性能。

[0135] 可选的，若第一业务是单边RDMA操作，由服务器200作为远端时，为完成该单边RDMA操作，第二类处理器200需启动与单边RDMA操作关联的信号量增加(sem up)线程和信号量减少(sem down)线程，sem up线程是指对第一业务中任务的信号量(sem)进行叠加，若叠加获得的信号量大于阈值(如零)，则继续该任务；若叠加获得的信号量小于或等于阈值(如零)，则唤醒阻塞在该信号量上的进程。sem down线程是指对第一业务中任务的信号量(sem)进行叠减，若叠减获得的信号量小于阈值(如零)，则继续该任务；若叠减获得的信号量大于阈值(如零)，则进程执行完毕。关于sem up线程的更多信息可以参照通常技术中信号量释放(V操作)的相关内容，关于sem up线程的更多信息可以参照通常技术中信号量通过(P操作)的相关内容，此处不予赘述。

[0136] 示例的，一个任务的sem up线程和sem down线程挂载于第二类处理器222的同一个处理器。如此，该任务不需要由第二类处理器222包含的两个处理器进行跨核协同，减少了该任务执行时所需的处理器通信时间，从而降低了该任务的处理时延，进而，提高了服务器执行包含该任务的业务的执行效率。

[0137] 值得注意的是，本文的上述实施例是以单个服务器获取到请求业务的消息来说明的，但是一些可能的情形中，本申请提供的业务处理方法还可以应用于更多的场景，如图5所示，图5为本申请提供的一种业务处理方法的流程示意图三，客户端服务器510和服务端服务器520可以实现上述实施例中任一服务器或客户端的功能，关于客户端服务器510和服务端服务器520的硬件实现，如网卡511、处理器512、网卡521、第一类处理器522、内存523和第二类处理器524，可以参考图2中服务器200的相关内容，此处不予赘述。

[0138] 在一种可能的示例中，客户端服务器510和服务端服务器520位于同一个计算机集群，如客户端服务器510为图1所示出的服务器111，服务端服务器520为图1所示出的服务器112。

[0139] 在另一种可能的示例中，客户端服务器510为图1所示出的客户端120，服务端服务器520为图1所示出的服务器112。

[0140] 在又一种可能的示例中，客户端服务器510和服务端服务器520与计算机集群并无关联，客户端服务器510和服务端服务器520均为独立的服务器。

[0141] 以上三种示例仅为本实施例对客户端服务器510和服务端服务器520的物理位置的说明，不应理解为对本申请的限定。

[0142] 在一些情形中，客户端服务器510可以称为client端，服务端服务器520可以称为server端。

[0143] 如图5所示，本实施例提供的业务处理方法包括以下内容1～9。

[0144] 内容1，处理器512敲门铃，以使网卡511发起RDMO请求。

[0145] 内容2，网卡511向网卡521发送RDMO请求。

[0146] 内容3，网卡521敲门铃，以向第一类处理器522转发RDMO请求。

[0147] 内容4，第一类处理器521对该RDMO请求进行解析，并进行硬件调度。

[0148] 硬件调度的过程可以包括，但不限于：第一类处理器522将RDMO请求的任务交由第一类处理器522或第二类处理器524执行。关于RDMO请求的解析过程以及第一类处理器521更具体的实现方式，可以参考上述S320的相关内容，此处不予赘述。

[0149] 内容5，第一类处理器521将RDMO请求所需的数据(业务数据和函数列表)写入内存523提供的DMA空间。

[0150] 内容6，第二类处理器524从内存523提供的DMA空间读取RDMO请求所需的数据，并输出RDMO请求的处理结果。关于第二类处理器524从DMA空间读取数据和输出处理结果的过程可以参考上述S330的相关内容，此处不予赘述。

[0151] 内容7，第二类处理器524向网卡521发送RDMO请求的处理结果。

[0152] 内容8，网卡521向网卡511发送RDMO请求的处理结果。

[0153] 内容9，网卡511向处理器512发送RDMO请求的处理结果。

[0154] 值得注意的是，图5是以服务端服务器520实现本申请提供的业务处理方法进行说明的，但在一些可能的情形中，client端(客户端服务器510)在接收到server端(服务端服

务器520)的回包后,也可以实现本申请提供的业务处理方法。如前述的回包是指server端执行RDMO请求的确定信息(acknowledgment,ACK)。

[0155] 综上,本申请提供的业务处理方法,服务器的内存基于DMA技术提供有DMA空间,该DMA空间可以被不同的硬件(如第一类处理器和第二类处理器)直接访问,进而触发第二类处理器读取业务的数据并输出处理结果。由于第一类处理器将第一业务所需的数据写入DMA空间的过程中,未向第二类处理器发送中断信号,因此,第二类处理器无需中断OS业务,减少了第一类处理器调度第二类处理器的时间,从而降低了第二类处理器执行第一业务的处理时延。

[0156] 其次,在第一业务的处理过程中,第二类处理器无需对该第一业务所需数据的地址进行转换,避免了地址转换导致的业务处理时延,这进一步减少了服务器执行第一业务的整体时延,提高了服务器处理业务的性能。

[0157] 另外,第一类处理器还可以将业务拆分为多个任务,并调度第一类处理器和第二类处理器包括的多个处理器,以执行多个任务,从而获得业务的处理结果,降低了处理器的业务处理时延。

[0158] 以及,在第二类处理器输出处理结果的过程中,多个可用处理器可以将各自获得的数据填充到SQ,以便服务器将这些数据一起发送到其他设备,这减少了服务器进行数据通信的资源消耗。

[0159] 最后,由于业务的执行过程中,第二类处理器可以从DMA空间直接获取到业务的数据,进而由挂载于处理器的用户态线程执行相应的任务,这避免了处理器中内核态和用户态的切换,减少了处理器的执行任务的处理时延,提升了服务器执行业务的处理效率。

[0160] 可以理解的是,为了实现上述实施例中功能,服务器包括了执行各个功能相应的硬件结构和/或软件模块。本领域技术人员应该很容易意识到,结合本申请中所公开的实施例描述的各示例的单元及方法步骤,本申请能够以硬件或硬件和计算机软件相结合的形式来实现。某个功能究竟以硬件还是计算机软件驱动硬件的方式来执行,取决于技术方案的特定应用场景和设计约束条件。

[0161] 图6为本申请提供的一种业务处理装置的结构示意图。业务处理装置可以用于实现上述方法实施例中服务器的功能,因此也能实现上述方法实施例所具备的有益效果。在本申请的实施例中,该业务处理装置可以是如图1所示的任一个服务器,也可以是如图1所示的客户端120,或者是图2所示出的服务器200,或者是图5所示出的客户端服务器510或服务端服务器520,还可以是应用于服务器的模块(如芯片)。

[0162] 如图6所示,业务处理装置600包括硬件调度模块610、业务处理模块620和存储模块630。业务处理装置600用于实现上述方法实施例中服务器的功能。在一种可能的示例中,硬件调度模块610应用于上述的第一类处理器,业务处理模块620应用于第二类处理器。该业务处理装置600用于实现上述业务处理方法的具体过程包括以下内容1～3。

[0163] 内容1,硬件调度模块610用于获取用于请求第一业务的第一消息。

[0164] 内容2,硬件调度模块还用于解析第一消息,将第一业务所需的数据写入DMA空间。该DMA空间是由服务器的内存基于DMA技术提供的存储空间。如图6所示,该DMA空间可以由存储模块630进行管理。

[0165] 内容3,业务处理模块630用于从DMA空间中读取数据,输出第一业务的处理结果。

[0166] 另外，当业务处理装置用于实现上述的方法实施例时，有益效果可以参见前述实施例的相关描述，此处不再赘述。

[0167] 另外，图6所示出的存储模块630还可以用于存储上述的用户态线程调度策略、状态信息等。关于用户态线程调度策略、状态信息更详细的内容可以参考上述方法实施例中的相关阐述，此处不予赘述。

[0168] 当业务处理装置600用于实现图3所示的方法实施例时：硬件调度模块610用于执行S310和S320，业务处理模块620用于执行S330。

[0169] 当业务处理装置600用于实现图4所示的方法实施例时：硬件调度模块610用于执行S310、S321和S322，业务处理模块620用于执行S330。

[0170] 当业务处理装置600用于实现图5所示的方法实施例时：硬件调度模块610、业务处理模块620和存储模块630用于协同实现内容2～8。

[0171] 或者，当业务处理装置600用于实现前述客户端服务器510的功能时，硬件调度模块610、业务处理模块620和存储模块630用于协同实现内容1、内容2、内容8和内容9。

[0172] 应理解的是，本发明本申请实施例的业务处理装置600可以通过CPU实现，也可以通过ASIC实现，或可编程逻辑器件(programmable logic device，PLD)实现，上述PLD可以是复杂程序逻辑器件(complex programmable logical device，CPLD)、FPGA、通用阵列逻辑(generic array logic，GAL)或其任意组合。业务处理装置600通过软件实现图3至图5中任一所示的业务处理方法时，业务处理装置600及其各个模块也可以为软件模块。

[0173] 有关上述业务处理装置600更详细的描述可以直接参考上述图3～图5所示的实施例中相关描述直接得到，这里不加赘述。

[0174] 示例的，当业务处理装置600通过硬件实现时，该硬件可以是一种计算设备，如上述的服务器，或应用在服务器的处理器或芯片等，如该计算设备包括接口电路和控制电路。

[0175] 接口电路用于接收来自计算设备之外的其它设备的信号并传输至控制电路，或将来自控制电路的信号发送给计算设备之外的其它设备。

[0176] 控制电路通过逻辑电路或执行代码指令用于实现上述实施例中任一种可能实现方式的方法。有益效果可以参见上述实施例中任一方面的描述，此处不再赘述。

[0177] 应理解，根据本申请实施例的服务器可对应于申请实施例中的业务处理装置600，并可以对应于执行根据本申请实施例的方法图3～图5中的相应主体，并且业务处理装置600中的各个模块的上述和其它操作和/或功能分别为了实现图3至图5中的各个方法的相应流程，为了简洁，在此不再赘述。

[0178] 可以理解的是，本申请的实施例中的处理器可以是CPU、NPU或GPU，还可以是其它通用处理器、DSP、ASIC、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器。

[0179] 本申请的实施例中的方法步骤可以通过硬件的方式来实现，也可以由处理器执行软件指令的方式来实现。软件指令可以由相应的软件模块组成，软件模块可以被存放于随机存取存储器(random access memory，RAM)、闪存、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)、寄存器、硬盘、移动硬盘、CD-ROM或者本领域熟知的任何其它形式的存储介质中。一种示例性的存储介质耦合至

处理器，从而使处理器能够从该存储介质读取信息，且可向该存储介质写入信息。当然，存储介质也可以是处理器的组成部分。处理器和存储介质可以位于ASIC中。另外，该ASIC可以位于网络设备或终端设备中。当然，处理器和存储介质也可以作为分立组件存在于网络设备或终端设备中。

[0180] 在上述实施例中，可以全部或部分地通过软件、硬件、固件或者其任意组合来实现。当使用软件实现时，可以全部或部分地以计算机程序产品的形式实现。所述计算机程序产品包括一个或多个计算机程序或指令。在计算机上加载和执行所述计算机程序或指令时，全部或部分地执行本申请实施例所述的流程或功能。所述计算机可以是通用计算机、专用计算机、计算机网络、网络设备、用户设备或者其它可编程装置。所述计算机程序或指令可以存储在计算机可读存储介质中，或者从一个计算机可读存储介质向另一个计算机可读存储介质传输，例如，所述计算机程序或指令可以从一个网站站点、计算机、服务器或数据中心通过有线或无线方式向另一个网站站点、计算机、服务器或数据中心进行传输。所述计算机可读存储介质可以是计算机能够存取的任何可用介质或者是集成一个或多个可用介质的服务器、数据中心等数据存储设备。所述可用介质可以是磁性介质，例如，软盘、硬盘、磁带；也可以是光介质，例如，数字视频光盘(digital video disc，DVD)；还可以是半导体介质，例如，固态硬盘(solid state drive，SSD)。

[0181] 在本申请的各个实施例中，如果没有特殊说明以及逻辑冲突，不同的实施例之间的术语和/或描述具有一致性、且可以相互引用，不同的实施例中的技术特征根据其内在的逻辑关系可以组合形成新的实施例。在本申请的实施例中涉及的各种数字编号仅为描述方便进行的区分，并不用来限制本申请的实施例的范围。上述各过程的序号的大小并不意味着执行顺序的先后，各过程的执行顺序应以其功能和内在逻辑确定。

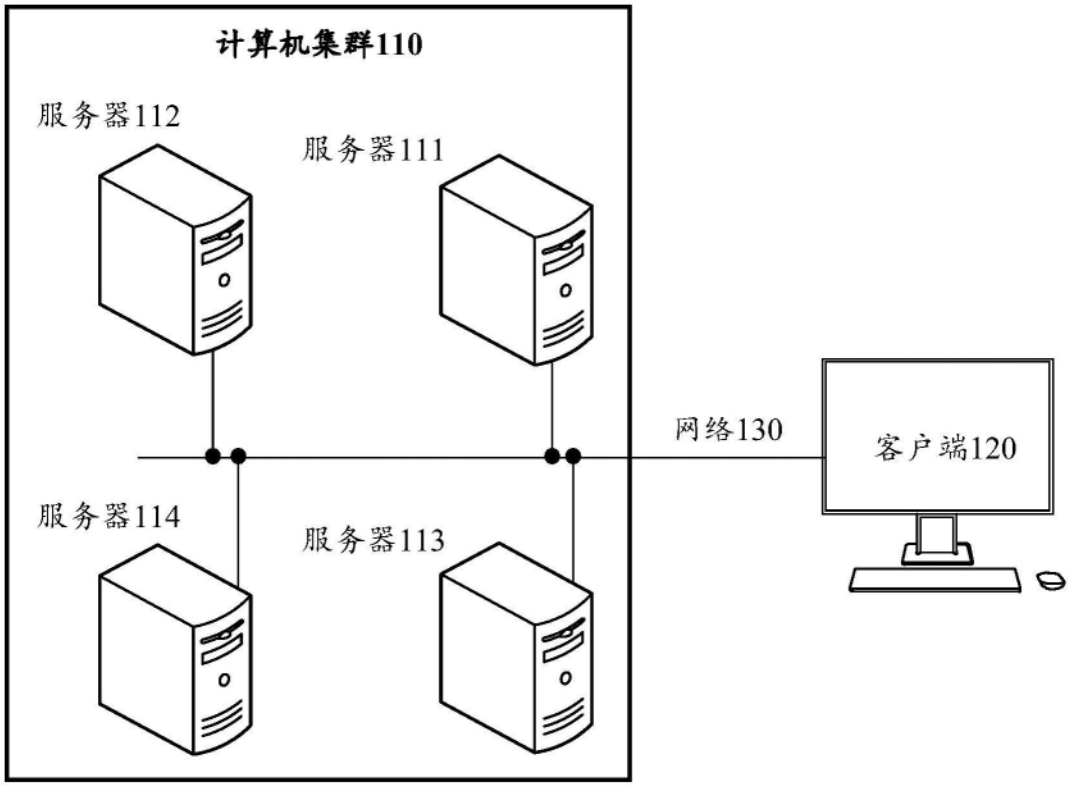

计算机集群110
服务器112
服务器111
服务器114
服务器113
网络130
客户端120

图1

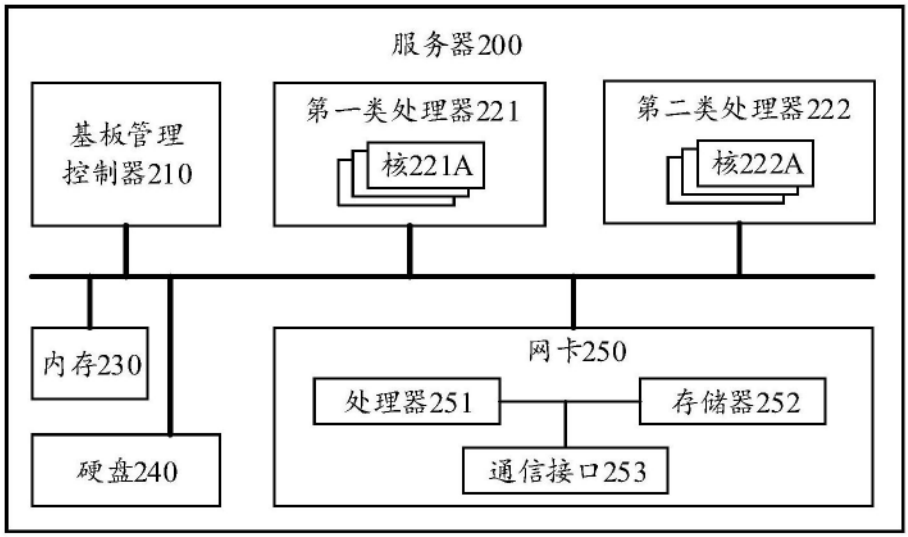

服务器200
基板管理
控制器210
第一类处理器221
核221A
第二类处理器222
核222A
内存230
硬盘240
网卡250
处理器251
存储器252
通信接口253

图2

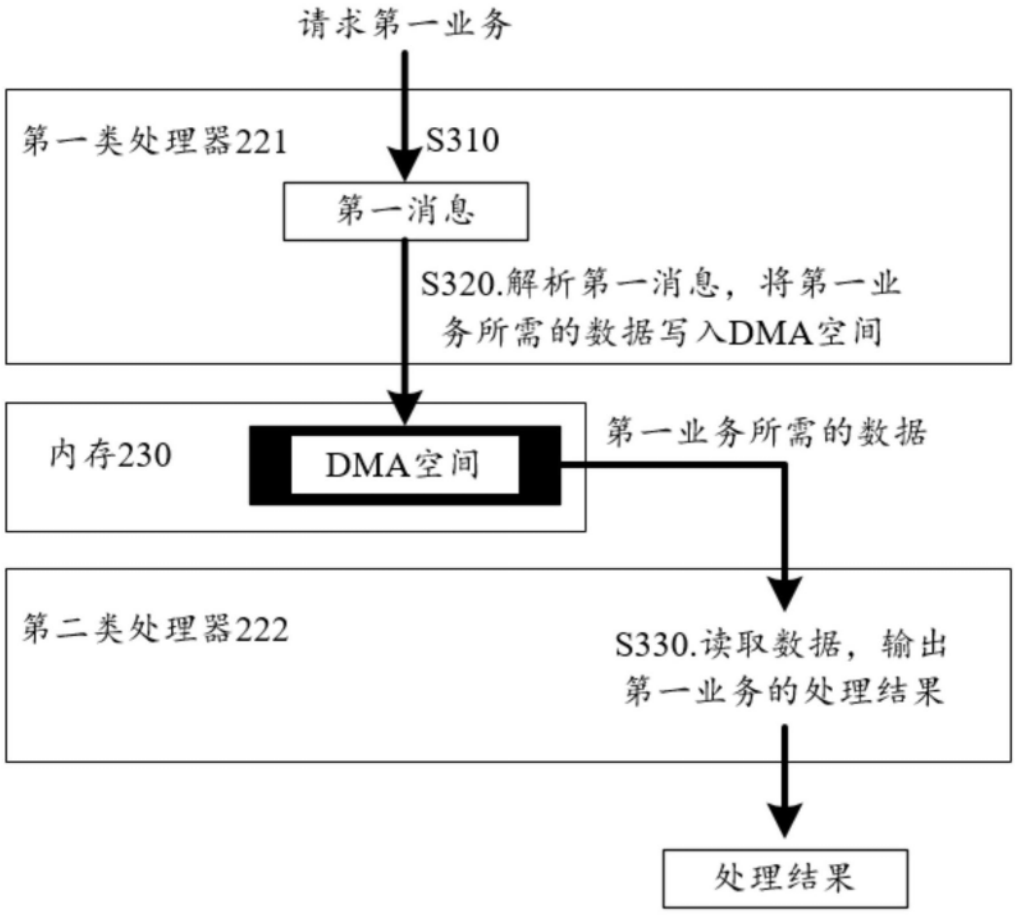

请求第一业务
第一类处理器221
S310
第一消息
S320.解析第一消息，将第一业务所需的数据写入DMA空间
内存230
DMA空间
第一业务所需的数据
第二类处理器222
S330.读取数据，输出第一业务的处理结果
处理结果

图3

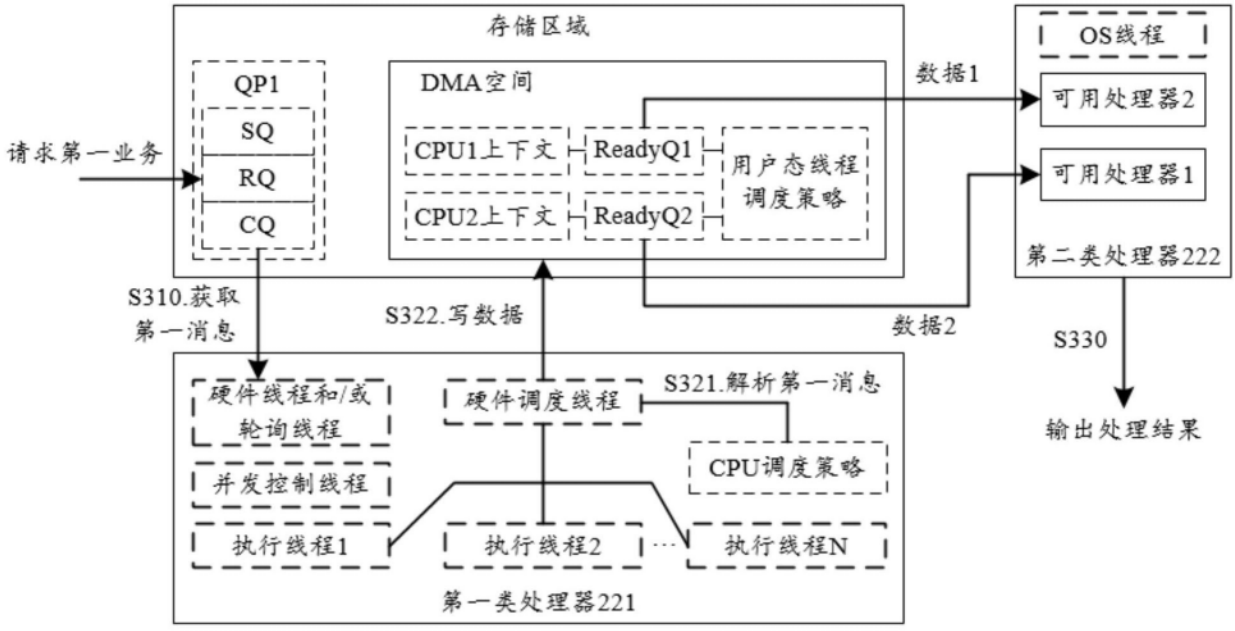

存储区域
QP1
SQ
RQ
CQ
DMA空间
CPU1上下文
ReadyQ1
CPU2上下文
ReadyQ2
用户态线程调度策略
请求第一业务
数据1
OS线程
可用处理器2
可用处理器1
第二类处理器222
S310.获取第一消息
S322.写数据
数据2
S330
输出处理结果
硬件线程和/或轮询线程
硬件调度线程
S321.解析第一消息
CPU调度策略
并发控制线程
执行线程1
执行线程2
执行线程N
第一类处理器221

图4

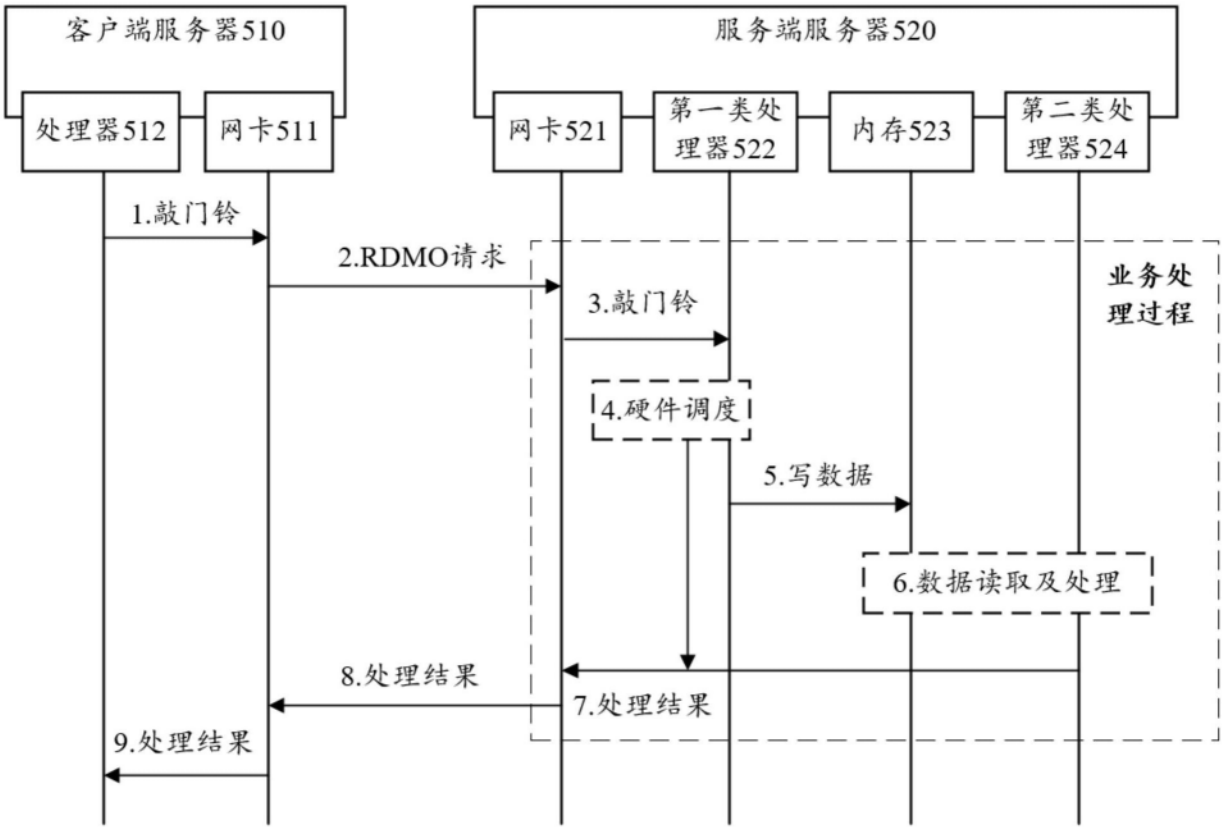

客户端服务器510
处理器512
网卡511
服务端服务器520
网卡521
第一类处理器522
内存523
第二类处理器524
1.敲门铃
2.RDMO请求
3.敲门铃
业务处理过程
4.硬件调度
5.写数据
6.数据读取及处理
8.处理结果
7.处理结果
9.处理结果

图5

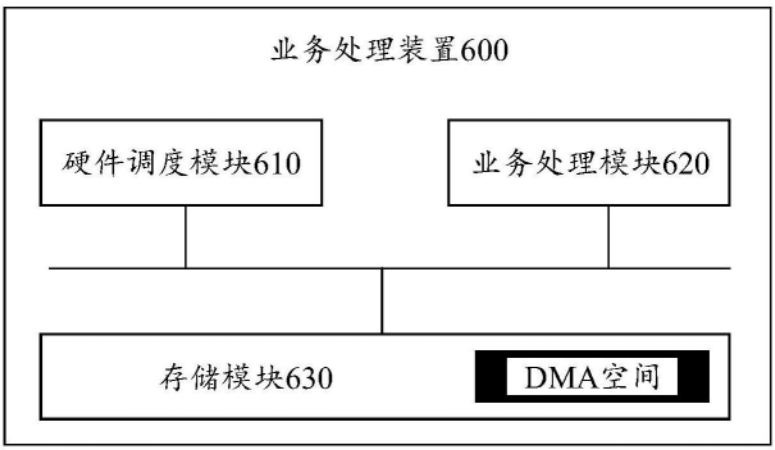

业务处理装置600
硬件调度模块610
业务处理模块620
存储模块630
DMA空间

图6