

存储设备、分布式存储系统以及数据处理方法

本申请要求于 2019 年 8 月 22 日提交的申请号为 201910779723.9、发明名称为“存储设备、AI 芯片、存储系统及 AI 计算方法”的中国专利申请的优先权，该中国专利申请要求于 2019 年 11 月 13 日提交的申请号为 201911109237.2、发明名称为“存储设备、分布式存储系统以及数据处理方法”的中国专利申请的优先权，两件专利申请的全部内容通过引用结合在本申请中。

技术领域

本申请涉及存储技术领域，特别涉及一种存储设备、分布式存储系统以及数据处理方法。

背景技术

随着存储技术以及人工智能（artificial intelligence, AI）技术的发展，存储设备中可以存储用于 AI 计算的业务数据，该业务数据例如可以是用于模型训练的样本集，比如说用于训练人脸识别模型的人脸图片集、用于训练语音识别模型的音频样本集、用于训练语义识别模型的样本文本等。

以人脸图片集为例，由于待计算的数据量庞大，单个服务器已经不能满足计算需求，因此目前主流架构为多设备的集群式架构。在该架构中，包括 AI 集群、存储集群和交换机。其中 AI 集群由多个 AI 服务器组成，用于进行 AI 计算；存储集群由多个存储设备组成，用于存储业务数据；交换机用于在 AI 服务器与存储设备之间转发业务数据。在 AI 计算的过程中，AI 服务器基于传输控制协议/因特网互联协议（transmission control protocol/internet protocol, TCP/IP）协议，与存储设备建立远程的网络连接。当 AI 服务器要获取 AI 计算所需的业务数据时，会通过交换机向存储设备发送数据获取请求，存储设备接收数据获取请求后，通过交换机向 AI 服务器发送存储的业务数据。然后，AI 服务器再将该业务数据加载至内存，进行 AI 计算。

在上述架构中，业务数据要从存储设备通过网络发送至交换机，再通过网络从交换机发送至 AI 服务器，才能让 AI 服务器得到业务数据来进行 AI 计算，可见 AI 计算的过程中获取业务数据的路径过长，造成获取业务数据的速度过慢，导致获取业务数据的效率低下。

发明内容

本申请实施例提供了一种存储设备、分布式存储系统以及数据处理方法，能够解决相关技术中业务数据获取速度过慢、获取效率低下的技术问题。所述技术方案如下：

第一方面，提供了一种存储设备，包括处理器、硬盘和 AI 装置，所述 AI 装置和所述处理器之间通过高速互联网络通信；所述处理器，用于接收业务数据，并将所述业务数据存储至所述硬盘中；所述 AI 装置，用于向所述处理器发送数据获取请求以获取所述业务数据，并对所述业务数据进行 AI 计算。

本实施例提供的存储设备，内部设置有 AI 装置，该存储设备既可以通过 AI 装置，提供 AI 计算的能力，又可以通过存储设备中的处理器和硬盘，提供业务数据存储的能力，从而实

现了存储与 AI 算力的融合。当需要进行 AI 计算时，业务数据在存储设备的内部通过高速互连网络进行传输，而无需经过外部网络的转发，因此极大地缩短了业务数据的传输路径，能够近距离地载入业务数据，从而提高了加载速度。可选地，所述数据获取请求包括第一数据获取请求；所述处理器，用于响应于所述第一数据获取请求，从所述硬盘中获取所述业务数据，并将所述业务数据发送给所述 AI 装置。通过该方式，提供了一种 AI 装置就近获取业务数据的方法，由于存储设备内部包含了 AI 装置、处理器和硬盘，当 AI 装置要获取业务数据时，通过向处理器发送数据获取请求，由存储设备的处理器从本地的硬盘中获取业务数据，发送给 AI 处理器，使得 AI 处理器能够从本地得到业务数据，免去了通过网络向远端存储设备请求业务数据带来的通信开销，从而节省了获取业务数据的时延。

可选地，所述数据获取请求包括第二数据获取请求；所述处理器，用于响应于所述第二数据获取请求，将所述业务数据的元数据发送给所述 AI 装置，所述元数据用于指示所述业务数据的地址；所述 AI 装置，用于当所述元数据指示所述业务数据位于所述存储设备本地时，向所述硬盘发送第一数据访问请求，所述第一数据访问请求包括所述元数据；所述硬盘，用于根据所述元数据获取所述业务数据，并通过 DMA 方式将所述业务数据写入所述 AI 装置。通过该方式，可以实现 AI 装置和硬盘之间的 DMA 直通，通过在 AI 装置和硬盘之间建立 DMA 通路，AI 装置和硬盘能够利用 DMA 通路进行快速业务数据交换，从而提高 AI 装置载入业务数据的速度，进而提高 AI 装置可同时处理的业务数据量，从而降低 AI 装置之间的传输 AI 参数的通信开销，加快 AI 训练的速度。

可选地，所述数据获取请求包括第三数据获取请求；所述处理器，用于响应于所述第三数据获取请求，将所述业务数据的元数据发送给所述 AI 装置，所述元数据用于指示所述业务数据的地址；所述 AI 装置，用于当所述元数据指示所述业务数据位于其他存储设备时，向所述其他存储设备发送第二数据访问请求，所述第二数据访问请求包括所述元数据。通过这种可选方式，AI 装置的 AI 内存与其他存储设备实现了 RDMA 直通，通过 AI 内存和其他存储设备进行快速业务数据交换，加快 AI 训练的速度。

可选地，所述存储设备还包括内存，所述处理器，还用于从所述内存中划分一段内存空间预留给所述 AI 装置使用。通过这种可选方式，AI 装置能够借用存储装置的内存来进行 AI 计算，从而扩大了 AI 装置的可用内存空间，让 AI 装置能够在更大的内存上进行 AI 计算，从而提高 AI 计算的效率。

可选地，所述 AI 装置包括 AI 处理器和 AI 内存；所述 AI 处理器用于当所述 AI 内存的可用容量达到预设阈值时，向所述处理器发送内存申请请求，所述 AI 内存的可用容量通过设定的批尺寸确定，所述内存申请请求用于请求所述处理器从所述内存中划分一段内存空间预留给所述 AI 装置使用。通过这种可选方式，AI 处理器可以利用内存的内存空间来进行训练，那么由于可用的内存空间更大，能够提高 AI 训练的批尺寸，从而提高 AI 装置一批可处理的业务数据量，降低不同 AI 装置之间交换 AI 参数的通信开销，提高 AI 训练的速度。经实验，如果仅通过 AI 内存进行 AI 训练，批尺寸最高设定为 256，而通过这种方式，可以将批尺寸设定为 32000，由此可见，批尺寸得到显著提升。

相关技术中，存储设备的内存的容量是固定的，这就导致存储业务数据时经常面临内存不足的问题。而通过这种可选方式，存储装置能够借用 AI 装置的 AI 内存来进行业务数据的读写，从而扩大了存储装置的可用内存空间，让存储装置能够在更大的内存上存储业务数据，

从而节省业务数据的读写时间，提高业务数据的读写效率。

可选地，所述 AI 装置包括 AI 处理器；所述 AI 处理器，用于将计算任务划分为至少两个子任务，将所述至少两个子任务中的第一子任务发送给所述处理器；所述处理器，还用于执行所述第一子任务，并将计算结果发送给所述 AI 处理器。通过这种可选方式，实现 AI 处理器和处理器的算力协同，AI 处理器能够通过借用存储设备自身的处理器的算力，来提高 AI 处理器的算力，从而加快了 AI 处理器处理 AI 计算的速度。

可选地，所述 AI 处理器，还用于在将计算任务划分为至少两个子任务之前确定所述 AI 处理器的算力不足。通过这种可选方式，AI 处理器能够自身的算力不足时，借用处理器的算力来处理 AI 计算，从而打破了 AI 训练过程中算力资源不足的瓶颈。

可选地，所述处理器，用于将计算任务划分为至少两个子任务，将所述至少两个子任务中的第二子任务发送给所述 AI 处理器；所述 AI 处理器，还用于执行所述第二子任务，并将计算结果发送给所述 AI 处理器。通过这种可选方式，实现 AI 处理器和处理器的算力协同，存储设备自身的处理器能够通过借用 AI 处理器的算力，来提高存储设备自身的处理器的算力，从而加快了存储设备自身的处理器处理业务数据的读写的速度。

所述处理器，还用于执行所述第一子任务，并将计算结果发送给所述 AI 处理器。

可选地，所述 AI 内存和所述存储设备的内存之间通过内存结构（memory fabric）通信。

可选地，所述 AI 内存和其他存储设备的 AI 内存通过 memory fabric 通信。

可选地，所述存储设备的内存和其他存储设备的内存通过 memory fabric 通信。

可选地，所述存储设备的内存和其他存储设备的 AI 内存通过 memory fabric 通信。

通过 memory fabric，可以实现同一存储设备中的内存与 AI 内存、不同存储设备的内存、不同存储设备的 AI 内存统一调度，从而提高存储系统调度内存资源和使用内存资源的效率。

可选地，所述 AI 装置和所述处理器之间通过高速串行计算机扩展总线标准（peripheral component interconnect express，简称：PCIe）总线通信；

可选地，所述处理器用于当所述内存的可用容量达到预设阈值时，向所述 AI 处理器发送内存申请请求，所述内存申请请求用于请求所述处理器从所述 AI 内存中划分一段内存空间预留给所述处理器使用。

可选地，所述 AI 装置还包括 AI 算力单元，所述 AI 算力单元，具体用于从所述 AI 内存中获取所述业务数据，并进行 AI 计算。可选地，所述 AI 处理器，还用于从所述 AI 内存中划分一段内存空间预留给所述存储设备的处理器使用。

可选地，将 AI 内存为第一层级，以内存为第二层级，对 AI 参数进行分层缓存，其中，第一层级的优先级高于第二层级的优先级。通过这一方式，实现了内存与 AI 内存之间的分层，会优先将 AI 参数缓存在 AI 内存中。

第二方面，提供了一种分布式存储系统，包括多个存储设备，所述多个存储设备包括第一存储设备；其中，所述第一存储设备包括第一处理器、第一硬盘和第一 AI 装置，所述第一 AI 装置和所述第一处理器之间通过高速互联网络通信；所述第一处理器，用于接收业务数据，并将所述业务数据存储至所述第一硬盘中；所述第一 AI 装置，用于向所述第一处理器发送数据获取请求以获取所述业务数据，并对所述业务数据进行 AI 计算。

本实施例提供的分布式存储系统，位于不同存储设备的两个 AI 装置，利用第一网络来交换 AI 参数；位于不同存储设备的两个存储装置，利用第二网络来交换业务数据，从而根据

AI 参数以及业务数据来协同进行 AI 计算，通过融合多个存储设备的存储能力和 AI 算力，能够提高系统整体的存储能力和 AI 算力。

可选地，所述第二存储设备用于通过第二网络将所述业务数据传输给所述第一存储设备；所述第二存储设备，还用于通过第一网络将 AI 参数传输给所述第一存储设备，所述 AI 参数用于对所述业务数据进行 AI 计算。通过部署第一网络和第二网络，AI 参数可以通过第一网络传输，而业务数据可以通过第二网络传输，因此，对于该系统而言，转发 AI 参数所使用的网络资源与转发其他数据所使用的网络资源能够相互分离，可以避免传递 AI 参数时占用存储设备原有的存储网络资源，从而避免 AI 计算的过程中因占用存储设备的网络带宽而导致存储设备的网络传输性能下降。并且，第一网络能够专用于转发 AI 相关的业务数据，因此基于该网络，可以避免对已有的业务数据中心的组网或存储设备集群产生影响。

可选地，所述第二存储设备，还用于通过所述第一网络将其他业务数据传输给所述第一存储设备。可选地，当所述第二网络的网络资源低于设定的存储网络资源阈值时，所述第一存储设备和所述第二存储设备之间通过所述第一网络传输其他业务数据。通过这种可选方式，为存储装置之间交换业务数据提供了一条新的通路，通过在第二网络的网络资源不足时，使用第一网络来交换业务数据，第一网络可以作为业务数据的一条新增的转发路径，那么业务数据既可以通过第二网络传输，也可以通过第一网络传输，因此提高了传输业务数据的网络带宽，从而缩短了交换业务数据的时延，提高了交换业务数据的速度，进而提高了进行 AI 计算的速度。

可选地，所述第二存储设备，还用于通过所述第二网络传输其他 AI 参数，所述其他 AI 参数用于对所述其他业务数据进行 AI 计算。

可选地，当第一网络的网络资源低于设定的 AI 网络资源阈值时，所述第一 AI 装置和所述第二 AI 装置之间通过所述第二网络传输其他 AI 参数。通过这种可选方式，为 AI 装置之间交换 AI 参数提供了一条新的通路，通过在第一网络的网络资源不足时，使用第二网络来交换 AI 参数，提高了传输 AI 参数的网络带宽，从而缩短了交换 AI 参数的时延，提高了交换 AI 参数的速度，进而提高了进行 AI 计算的速度。

可选地，所述第一 AI 装置包括第一 AI 处理器和第一 AI 内存，所述第二存储装置包括第二处理器，所述第二 AI 装置包括第二 AI 内存；所述第一 AI 处理器，用于向所述第二处理器发送第二网络的网络资源请求，并且向所述第二处理器发送内存 RDMA 访问请求；所述第一 AI 处理器，用于从所述第一内存中读取其他业务数据，通过所述第二网络将所述其他业务数据发送给所述第二 AI 装置；所述第二 AI 装置将所述其他业务数据写入所述第二内存。通过这种可选方式，实现了第一存储设备和第二存储设备的内存直通，第一存储设备的内存与第二存储设备的内存之间能够通过 RDMA 的方式来交换业务数据，如此，能够免去第一处理器与第二处理器的处理开销，让业务数据从第一内存直接到达第二内存，从而加快业务数据交换的速度，提高业务数据交换的效率。

可选地，该系统还包括管理装置；所述管理装置，用于接收第一作业请求，根据所述第一作业请求确定待训练的数据集的分布，所述数据集包括所述业务数据；当确定所述业务数据分布在所述第一存储设备时，向所述第一存储设备发送第一计算请求，所述第一计算请求用于请求所述第一 AI 装置对所述业务数据进行 AI 计算。通过这种可选方式，由于管理装置选择了业务数据所在的第一存储设备来提供 AI 计算，而第一存储设备通过内部的第一存储装

置,即可得到业务数据来进行AI计算,从而避免业务数据的跨存储设备移动,通过免去了访问其他存储设备以得到业务数据带来的时延,从而减少了获取业务数据的时延,也就提高了进行AI计算的速度。

可选地,所述第一AI装置,用于基于所述第一计算请求,从所述第一存储设备的第一存储装置获取所述业务数据,对所述业务数据进行AI计算,获得第一计算结果。可选地,所述管理装置向所述第一存储设备的第一AI装置发送第一计算请求之前,还用于确定所述第一存储设备的运行状态满足设定条件。通过这种可选方式,能够保证选中的第一存储设备当下未被占用,能够提供AI算力,避免选中已被占用的存储设备进行AI计算导致设备开销过大、AI计算任务无法及时完成的问题。

可选地,所述数据集还包括其他业务数据;所述管理装置,还用于当确定所述其他业务数据分布在所述多个存储设备中的第二存储设备中时,进一步确定所述第二存储设备的运行状态,当所述第二存储设备的运行状态不满足所述设定条件时,向所述第一存储设备发送第二计算请求,所述第一存储设备与所述第二存储设备之间的距离低于设定的距离阈值;所述第一AI装置,还用于基于所述第二计算请求,从所述第二存储设备中获取所述其他业务数据,对所述其他业务数据进行AI计算,获得第二计算结果。通过这种可选方式,如果业务数据所在的存储设备已被占用,管理装置能够选择与业务数据距离近的存储装置来提供AI计算,从而缩短业务数据的传输距离,减少业务数据的跨节点移动。

可选地,所述管理装置确定所述第二存储设备的运行状态是否满足设定条件之前,还用于接收第二作业请求,根据所述第二作业请求确定待训练的第二数据集的分布在第二存储设备。

第三方面,提供了一种数据处理方法,所述方法应用于分布式存储系统中,所述分布式存储系统包括多个存储设备;所述多个存储设备包括第一存储设备,该方法用于实现第二方面任意一种实现方式所提供的功能。

第四方面,提供了一种数据处理方法,所述方法应用于存储设备,其中,所述存储设备包括处理器、硬盘和AI装置,所述AI装置和所述处理器之间通过高速互连网络通信;该方法用于实现第一方面任意一种实现方式所提供的功能。

第五方面,提供了一种计算机可读存储介质,该存储介质中存储有至少一条指令,该指令由存储设备读取以使存储设备执行上述第四方面或第四方面任一种可选方式所提供的数据处理方法。

第六方面,提供了一种计算机可读存储介质,该存储介质中存储有至少一条指令,该指令由分布式存储系统读取以使所述分布式存储系统执行上述第三方面或第三方面任一种可选方式所提供的数据处理方法。

第七方面,提供了一种计算机程序产品,当计算机程序产品在存储设备上运行时,使得存储设备执行上述第四方面或第四方面任一种可选方式所提供的数据处理方法。

第八方面,提供了一种计算机程序产品,当计算机程序产品在分布式存储系统上运行时,使得分布式存储系统执行上述第三方面或第三方面任一种可选方式所提供的数据处理方法。

附图说明

图1是本申请实施例提供的一种分布式存储系统的架构图;

图 2 是本申请实施例提供的一种存储设备的结构示意图；
图 3 是本申请实施例提供的一种存储设备的结构示意图；
图 4 是本申请实施例提供的一种分布式存储系统的架构图；
图 5 是本申请实施例提供的一种存储设备的结构示意图；
图 6 是本申请实施例提供的一种存储设备的结构示意图；
图 7 是本申请实施例提供的一种数据处理方法的流程图；
图 8 是本申请实施例提供的一种数据处理方法的流程图；
图 9 是本申请实施例提供的一种数据处理方法的流程图；
图 10 是本申请实施例提供的一种数据传输流程的示意图；
图 11 是本申请实施例提供的一种数据传输流程的示意图；
图 12 是本申请实施例提供的一种数据处理方法的流程图；
图 13 是本申请实施例提供的一种数据传输流程的示意图；
图 14 是本申请实施例提供的一种数据处理方法的流程图；
图 15 是本申请实施例提供的一种数据处理方法的流程图；
图 16 是本申请实施例提供的一种数据处理方法的流程图；
图 17 是本申请实施例提供的一种数据处理方法的流程图；
图 18 是本申请实施例提供的一种数据处理方法的流程图；
图 19 是本申请实施例提供的一种数据处理方法的流程图；
图 20 是本申请实施例提供的一种数据传输流程的示意图；
图 21 是本申请实施例提供的一种数据处理方法的流程图；
图 22 是本申请实施例提供的一种数据处理方法的流程图；
图 23 是本申请实施例提供的一种数据传输流程的示意图；
图 24 是本申请实施例提供的一种数据处理方法的流程图；
图 25 是本申请实施例提供的一种数据传输流程的示意图；
图 26 是本申请实施例提供的一种数据处理方法的流程图；
图 27 是本申请实施例提供的一种数据传输流程的示意图；
图 28 是本申请实施例提供的一种数据处理方法的流程图；
图 29 是本申请实施例提供的一种数据传输流程的示意图；
图 30 是本申请实施例提供的一种数据处理方法的流程图；
图 31 是本申请实施例提供的一种数据处理方法的流程图；
图 32 是本申请实施例提供的一种数据处理方法的流程图；
图 33 是本申请实施例提供的一种数据处理方法的逻辑架构图。

具体实施方式

以下对本申请涉及的术语进行解释：

人工智能(artificial intelligence, AI)：是通过计算机或者计算机控制的机器，对人的智能进行模拟、延伸和扩展，来感知环境、获取知识并使用知识获得最佳结果的理论、方法、技术及应用系统。换句话说，人工智能是计算机科学的一个综合技术，它基于智能的实质，生产出一种新的能以人类智能相似的方式做出反应的智能机器。人工智能也就是研究各种智能

机器的设计原理与实现方法，使机器具有感知、推理与决策的功能。

通常来讲，AI的实现包括两个环节：训练以及推理。训练，是指通过大量标记过的样本训练出一个神经网络模型，使神经网络模型可以具有特定的功能。推理，也称预测或推断，是指利用训练好的神经网络模型，使用新的业务数据推理出各种结论。

高速串行计算机扩展总线标准（peripheral component interconnect express，简称：PCIe）总线：在外设部件互连标准（peripheral component interconnect，简称：PCI）总线的基础上发展的一种局部总线，用于将处理器与至少一个外部设备连接。符合PCIe总线标准的外部设备称为PCIe设备。PCIe总线上具有至少一个PCIe接口，每个PCIe接口的物理形态可以是一个插槽。每个PCIe接口用于连接一个PCIe设备。PCIe总线上的各个PCIe设备采用了串行互联方式，PCIe总线上的不同PCIe设备之间可以通过点对点的形式进行业务数据传输。PCIe协议中通常会兼容PCI协议相关的技术以及PCI设备。

缓存一致性系统（Huawei cache-coherent system，HCCS）：维护多端口（socket）之间业务数据一致性的协议标准。

AI参数：是指通过AI训练确定出的AI模型中的参数。通俗地说，AI模型可以视为一种函数，AI参数可以视为函数中的系数。例如，如果AI模型是神经网络，AI参数可以是神经网络中卷积核的权重。又例如，如果AI模型是支持向量机，AI参数可以是支持向量机中的支持向量；又例如，如果AI模型是线性回归模型或逻辑回归模型，AI参数可以是线性回归或逻辑回归模型中的系数。当然，列举的AI模型仅是示例，AI模型还可以是其他类型的模型，例如决策树模型、随机森林模型、置信网络、强化学习模型、迁移学习模型、归纳学习模型中的一项及其组合，相应地，AI参数也可以是其他类型的模型中的参数，本申请实施例并不限定AI参数和AI模型的具体类型。

AI参数的调整过程对于AI计算来讲是至关重要的。具体来讲，AI计算的过程中，通常会将数据集中的业务数据，输入至AI模型，AI模型会基于AI参数对业务数据进行推理预测，得到预测结果。根据预测结果与真实结果之间的误差，会对AI参数进行调整，使得根据调整后的AI参数来进行下一次推理预测时，误差得以减小。通过循环执行AI参数调整的过程，能够让AI参数通过调整来逐渐精确，当训练结束时，利用包含了精确的参数的AI模型，可以实现精确的推理预测，比如精确的将人脸图像中的人脸识别出来。通过实施本申请的一些实施例，可以极大的加速AI参数的调整过程。例如，通过执行下述图17实施例，可以通过借用存储设备的处理器的算力，来极大地提升AI算力，那么通过更大的AI算力，可以提高AI参数的计算速度。又如，通过执行下述图26实施例，可以通过借用存储后端网络资源（即第二网络的网络资源），来极大地提升AI计算可用的网络资源，那么通过更多的网络资源，可以加速AI参数的传输，从而提高AI参数调整更新的速度。

直接内存存取（direct memory access，简称：DMA）：一种在内存与外部设备之间传输数据的技术。通过DMA，可以无需通过中央处理器（central processing unit，CPU）的参与，而是由外部设备直接将数据写入内存或者访问内存中的数据。例如，外部设备可以向处理器申请内存空间，处理器会为外部设备分配内存缓存区（buffer），之后，外部设备可以直接将数据写入至该内存缓存区，或者从内存缓存区直接读取数据。这里所说的外部设备在DMA相关的协议中，可以称为DMA控制器。

在本申请的图9实施例中，硬盘可以是上面描述的外部设备，AI内存可以是上面描述的

内存, 硬盘可以通过上面描述的 DMA 的方式, 来绕开存储设备的处理器和 AI 处理器的参与, 直接访问 AI 内存, 从而加快访问 AI 内存的速度。

远程直接内存访问 (remote direct memory access, 简称: RDMA): 一种将数据从一台设备的内存直接传输至另一台设备的内存的技术。RDMA 提供了基于消息队列的点对点通信, 每个应用可以通过消息队列来获取自己的消息, 从而免去了 CPU 的介入。在 RDMA 相关的协议中, 执行 RDMA 操作 (RDMA verb) 的装置可以称为远程直接数据存取网络接口控制器 (RDMA network interface controller, 简称: RNIC)。RDMA 操作可以包括存储操作 (memory verb) 和消息操作 (messaging verb)。存储操作可以用于传输数据, 存储操作包括 RDMA 写操作、RDMA 读操作和原子操作。例如, RDMA 写操作的过程可以包括: 目的设备的 CPU 将内存区域的虚拟地址以及内存区域的权限信息发送给源设备的 CPU; 源设备的 CPU 将待写入的数据存储至源设备的内存区域; 源设备的 CPU 根据源设备的内存区域的虚拟地址、目的设备的内存区域的虚拟地址、目的设备的内存区域的权限信息, 生成 RDMA 写指令, 将 RDMA 写指令加入至 RNIC 的发送队列。之后, 源设备会通过门铃机制, 通知 RNIC 执行发送队列中的指令。之后, RNIC 会从发送队列中读取指令, 得到 RDMA 写指令, 根据 RDMA 写指令, 执行 RDMA 写操作。RDMA 读操作可以包括: 目的设备的 CPU 将数据存储在内内存区域, 将内存区域的虚拟地址以及目的设备的内存区域的权限信息发送给源设备的 CPU, 源设备的 CPU 接收虚拟地址以及权限信息, 根据源设备的内存区域的虚拟地址、目的设备的内存区域的虚拟地址、目的设备的内存区域的权限信息, 生成 RDMA 读指令, 将 RDMA 读指令加入发送队列中。之后, 源设备会通过门铃机制, 通知 RNIC 执行发送队列中的指令。RNIC 会从发送队列中读取指令, 得到 RDMA 读指令, 根据 RDMA 读指令, 执行 RDMA 读操作。消息操作可以用于控制消息, 消息操作可以包括 RDMA 发送操作和 RDMA 接收操作。

在本申请的图 26 实施例中, AI 装置中的第一网络接口 (可以称为 AI 参数网络接口) 可以是 RNIC, 可以通过上面描述的 RDMA 方式, 来访问其他存储设备中的 AI 内存, 或者通过 RDMA 来将 AI 参数写入其他存储设备中的 AI 内存; 在本申请的图 22 实施例中, 存储装置中的第二网络接口 (可以称为后端交换接口) 可以是 RNIC, 可以通过上面描述的 RDMA 方式, 来访问其他存储设备中的内存, 或者通过 RDMA 来将业务数据写入其他存储设备中的内存。

以下, 对本申请提供的分布式存储系统进行示例性描述。

图 1 是本申请实施例提供的一种分布式存储系统的架构图, 分布式存储系统 1 包括多个存储设备, 该多个存储设备包括第一存储设备 10。

第一存储设备 10 用于存储业务数据以及处理 AI 计算, 第一存储设备 10 包括第一存储装置 100 和第一 AI 装置 101, 第一 AI 装置 101 设置于第一存储设备 10 内部。第一存储装置 100 可以包括第一处理器 1001 和第一硬盘 1002。

第一 AI 装置 101 和第一处理器 1001 之间通过高速互连网络通信。

第一处理器 1001, 用于接收业务数据, 并将业务数据存储至第一硬盘 1002 中。

第一 AI 装置 101, 用于向第一处理器 1001 发送数据获取请求以获取业务数据, 并对业务数据进行 AI 计算。

可选地, 数据获取请求包括第一数据获取请求;

第一处理器 1001, 用于响应于第一数据获取请求, 从第一硬盘 1002 中获取业务数据, 并将业务数据发送给第一 AI 装置 101。

可选地, 数据获取请求包括第二数据获取请求;

第一处理器 1001, 用于响应于第二数据获取请求, 将业务数据的元数据发送给第一 AI 装置 101, 元数据用于指示业务数据的地址;

第一 AI 装置 101, 用于当元数据指示业务数据位于第一存储设备本地时, 向第一硬盘 1002 发送第一数据访问请求, 第一数据访问请求包括元数据;

第一硬盘 1002, 用于根据元数据获取业务数据, 并通过 DMA 方式将业务数据写入第一 AI 装置 101。

可选地, 该多个存储设备还包括第二存储设备 11, 该第一存储设备 10 和第二存储设备 11 为两个不同的存储设备, 第一存储设备 10 与第二存储设备 11 之间通过网络通信。

第二存储设备 11 用于存储业务数据以及处理 AI 计算, 第二存储设备 11 包括第二存储装置 110 和第二 AI 装置 111, 第二 AI 装置 111 设置于第二存储设备 11 内部。

其中, 第二存储设备 11 包括第二处理器 1101 和第二硬盘 1102, 第二 AI 装置和第二处理器 1101 之间通过高速互连网络通信;

第二处理器 1101, 用于接收业务数据, 并将业务数据存储至第二硬盘 1102 中;

第二 AI 装置, 用于向第二处理器 1101 发送数据获取请求以获取业务数据, 并对业务数据进行 AI 计算。

可选地, 数据获取请求包括第三数据获取请求;

第一处理器 1001, 用于响应于第三数据获取请求, 将业务数据的元数据发送给第一 AI 装置 101, 元数据用于指示业务数据的地址;

第一 AI 装置 101, 用于当元数据指示业务数据位于多个存储设备中的第二存储设备 11 时, 向第二存储设备 11 发送第二数据访问请求, 第二数据访问请求包括元数据; 接收第二存储设备 11 响应于第二数据访问请求所发送的业务数据。

可选地, 第二存储设备 11 用于通过第二网络 12 将业务数据传输给第一存储设备 10; 第二存储设备 11, 还用于通过第一网络 13 将 AI 参数传输给第一存储设备 10, AI 参数用于对业务数据进行 AI 计算。

第一网络 13 可以称为 AI 参数网络, 用于在第一存储设备 10 和第二存储设备 11 之间传输 AI 参数。第一网络 13 中可以设置一个或多个网络设备, 例如路由器、交换机等。具体地, 该网络设备 13 可以分别和第一 AI 装置 101 以及第二 AI 装置 111 相连, 第一 AI 装置 101 和第二 AI 装置 111 之间通过网络设备传输 AI 参数。例如, 网络设备可以和第一 AI 装置 101 的第一网络接口以及第二 AI 装置 111 的第一网络接口相连。

第二网络 12 可以称为存储后端网络, 用于在第一存储设备 10 和第二存储设备 11 之间传输业务数据。第二网络 12 中可以设置一个或多个网络设备, 例如路由器、交换机等。具体地, 该网络设备可以和第一存储装置 100 以及第二存储装置 110 相连, 第一存储装置 100 和第二存储装置 110 之间通过网络设备传输业务数据。例如, 网络设备可以和第一存储装置 100 的第二网络接口以及第二存储装置 110 的第二网络接口相连。

通过部署第一网络 13 和第二网络 12, AI 参数可以通过第一网络 13 传输, 而业务数据可以通过第二网络 12 传输, 因此, 对于存储系统 1 而言, 转发 AI 参数所使用的网络资源与转

发其他数据所使用的网络资源能够相互分离,可以避免传递 AI 参数时占用存储设备原有的存储网络资源,从而避免 AI 计算的过程中因占用存储设备的网络带宽而导致存储设备的网络传输性能下降。并且,第一网络 13 能够专用于转发 AI 相关的业务数据,因此基于该网络,可以避免对已有的业务数据中心的组网或存储设备集群产生影响。

可选地,还可以为分布式存储系统 1 部署第三网络 14,第三网络 14 可以称为业务网络。第一存储设备 10 和第二存储设备 11 可以通过第三网络 14 通信。第三网络 14 可以包括一个或多个网络设备,网络设备可以分别和第一存储装置 100 的业务接口以及第二存储装置 110 的业务接口相连。应理解,本实施例仅是以分布式存储系统 1 中的第一存储设备 10 以及第二存储设备 11 为例进行描述,本领域技术人员可以知晓,分布式存储系统 1 包含的存储设备的数量可以更多或更少,例如分布式存储系统 1 可以仅包括第一存储设备 10 以及第二存储设备 11,或者包括几十个或几百个存储设备,或者更多存储设备,此时上述分布式存储系统 1 还包括第一存储设备 10 以及第二存储设备 11 之外的其他存储设备,本实施例对存储设备的数量不做限定。尤其是,随着 AI 计算对算力需求的增长,本实施例所示的存储系统 1 的规模可以相应增加,例如存储系统 1 可以包括百万个存储设备,从而将系统 1 整体的算力向百万级以上提升。

另外,第一存储设备 10 可以泛指分布式存储系统 1 中多个存储设备中的任一个存储设备,第二存储设备 11 可以泛指分布式存储系统 1 中多个存储设备中第一存储设备 10 之外的任一个存储设备,本实施例仅是以第一存储设备 10 和第二存储设备 11 来举例说明。

本实施例提供的分布式存储系统,位于不同存储设备的两个 AI 装置,利用第一网络来交换 AI 参数;位于不同存储设备的两个存储装置,利用第二网络来交换业务数据,从而根据 AI 参数以及业务数据来协同进行 AI 计算,通过融合多个存储设备的存储能力和 AI 算力,能够提高系统整体的存储能力和 AI 算力。其中,第一网络可以是 PCIe 高速互联网络、Fibre Channel (FC)、SCSI、以太网、RDMA、或者 memory fabric 等。第二网络可以是 PCIe 高速互联网络、FC、SCSI、以太网、RDMA、或者 memory fabric 等。

可选地,第二存储设备 110,还用于通过第一网络 13 将其他业务数据传输给第一存储设备 10;第二存储设备 110,还用于通过第二网络 12 传输其他 AI 参数,其他 AI 参数用于对其其他业务数据进行 AI 计算。

例如,当第二网络资源低于设定的存储网络资源阈值时,第一存储装置 100 和第二存储装置 110 之间可以通过第一网络 13 传输其他业务数据。具体来说,第一存储装置 100 包括第一处理器和第一内存,第二存储装置 110 包括第二处理器和第二内存。第一 AI 装置 101 包括第一 AI 处理器和第一 AI 内存。第一 AI 处理器,向第二 AI 处理器发送第一网络的网络资源请求,并且向第二处理器发送内存 RDMA 访问请求。第一 AI 装置 101 从第一内存中读取其他业务数据,通过第一网络 13 将其他业务数据发送给第二 AI 装置 111。第二 AI 装置 111 将其他业务数据写入第二内存。

可选地,当第一网络资源低于设定的 AI 网络资源阈值时,第一 AI 装置 101 和第二 AI 装置 111 之间通过第二网络 12 传输其他 AI 参数。具体来说,第一 AI 处理器,向第二处理器发送第二网络的网络资源请求。第一 AI 处理器,用于从第一 AI 内存中获取其他 AI 参数,将其他 AI 参数以 RDMA 方式通过第二网络 12 传输至第二 AI 内存。

可选地,系统 1 还包括管理装置,管理装置是一个软件模块,从功能上讲它用于管理分

布式存储系统 1 中的所有存储设备，例如进行存储设备的调度以处理 AI 计算。管理装置可以位于一个独立的设备，例如可以是主机、服务器、个人电脑或其他设备。另外，管理装置也可以位于任意一个存储设备中。当管理装置是一个独立的设备时，它可以通过网络与该系统中的任意一个存储设备通信。

下面举一个具体的例子来说明管理装置如何实现各个存储设备之间的调度。当管理装置接收第一作业请求时，根据第一作业请求确定待训练的数据集的分布。当确定该数据集中的业务数据分布在第一存储设备 10 时，管理装置向第一存储设备 10 的第一 AI 装置 101 发送第一计算请求。第一 AI 装置 101，用于基于第一计算请求，从第一存储设备 10 获取业务数据，对业务数据进行 AI 计算，获得第一计算结果。这是一种近数据的 AI 计算方式，即优先选择数据集所在的存储设备的 AI 装置对该数据集进行 AI 计算。可以在一定程度上避免跨网络传输数据，节省了网络资源。应理解，在某些情况下可能数据集所在的存储设备的运行状态不佳，以致于不适合接收 AI 计算任务，因此在本实施例中管理装置向第一存储设备 10 的第一 AI 装置 101 发送第一计算请求之前，还用于确定第一存储设备 10 的运行状态满足设定条件。

那么，当第一存储设备 10 的运行状态不满足设定条件时，可以选择与第一存储设备 10 之间的距离低于设定的距离阈值的存储设备，例如第二存储设备 11，来进行 AI 计算。

在一些实施例中，待训练的数据集可以还包括其他业务数据；

管理装置，还用于当确定其他业务数据分布在多个存储设备中的第二存储设备 11 中时，进一步确定第二存储设备 11 的运行状态，当第二存储设备 11 的运行状态不满足设定条件时，向第一存储设备 10 发送第二计算请求，第一存储设备 10 与第二存储设备 11 之间的距离低于设定的距离阈值；

第一 AI 装置 101，还用于基于第二计算请求，从第二存储设备 11 中获取其他业务数据，对其他业务数据进行 AI 计算，获得第二计算结果。

在一些可能的实施例中，存储系统 1 还可以包括主机（图 1 中未示出），主机用于和第一存储设备 10、第二存储设备 11 协同进行 AI 计算。其中，主机可以是 AI 服务器，也可以是应用服务器。AI 服务器用于执行 AI 计算任务。例如，AI 服务器可以进行模型训练或业务数据处理。在 AI 服务器运行的过程中，AI 服务器可以与第一存储设备 10 和/或第二存储设备 11 交换模型参数。应用服务器可以从终端接收模型训练或业务数据处理的指令，为 AI 计算的过程进行资源调度、存储 AI 算法或更新 AI 算法等。其中，主机可以通过 CPU 等通用的处理器，来提供 AI 算力，也可以通过图形处理器（英文：graphics processing unit，简称：GPU）、神经网络处理器（英文：neural-network processing unit，简称：NPU）、张量处理单元（英文：tensor processing unit，简称：TPU）或现场可编程门阵列（英文：field-programmable gate array，简称：FPGA）等处理器，来提供 AI 算力。主机可以是物理的设备，也可以是虚拟化设备，比如从云平台租用的弹性云服务器。在 AI 计算的过程中，第一存储设备 10 和第二存储设备 11 可以承担主要的计算任务，主机承担次要的计算任务。或者，主机承担主要的计算任务，第一存储设备 10 和第二存储设备 11 承担次要的计算任务。

在一些可能的实施例中，第二网络 12 可以包括内置网络设备以及外置网络设备。

内置网络设备可以设置在第一存储设备 10 和/或第二存储设备 11 的内部，内置网络设备可以通过高速互联网络与第一存储设备 10 和/或第二存储设备 11 的部件连接。内置网络设备可以为总线，具体可以是串行接口总线，比如 PCIE 总线、HCCS 总线、以太网、IB、FC 中

的任意一种。第一存储设备 10 和/或第二存储设备 11 可以通过内置网络设备,对存储的数据进行备份,从而保护业务数据以及 AI 参数。

外置网络设备可以设置在第一存储设备 10 和第二存储设备 11 外部,外置网络设备与第一存储设备 10 和第二存储设备 11 通过网络相连。外置网络设备可以为交换机、路由器等。第一存储设备 10 和第二存储设备 11 可以通过外置网络设备,向应用提供数据存储业务,例如文件存储、对象存储或者块存储业务。外置网络设备可以为交换机、路由器等。

在一些可能的实施例中,第一网络 13 中可以设置内置网络设备以及外置网络设备。第一网络中的内置网络设备以及外置网络设备分别与第二网络 12 中的内置网络设备以及外置网络设备同理,在此不做赘述。

在一些可能的实施例中,第二网络 12 和/或第一网络 13 中的内置网络设备和外置网络设备可以合一部署。

在一些可能的实施例中,系统 1 中存储设备的数量可以根据对 AI 算力的需求,动态地进行伸缩,实现弹性扩容或弹性缩容的功能。具体来讲,存储系统 1 可以视为一个资源池,存储设备可以视为资源池中的资源,该资源能够同时提供 AI 算力以及存储能力。如果存储系统 1 的 AI 算力不足,比如所有 AI 装置的负载已经超过负载阈值,可以向存储系统 1 添加存储设备,通过存储系统 1 已有的存储设备以及新增的存储设备共同提供 AI 算力,从而实现扩容,减轻单个存储设备中 AI 装置的负载。如果存储系统 1 空闲,可以减少存储系统 1 中存储设备的数量,从而实现缩容。

以下,对本申请提供的存储设备进行示例性描述。

图 2 是本申请实施例提供的一种存储设备的结构示意图,存储设备 2 可以是图 1 所示系统 1 中的第一存储设备 10,也可以是图 1 所示系统 1 中的第二存储设备 11,当然也可以是图 1 未示出的其他存储设备。

如图 2 所示,该存储设备 2 包括存储装置 20 和 AI 装置 21。

存储装置 20 包括硬盘 201 和处理器 202。

处理器 202 用于接收业务数据,并将业务数据存储至硬盘 201 中。其中,业务数据可以作为 AI 计算的输入,例如,业务数据可以为用于模型训练的样本集、用于推理预测的数据集等。比如说,在模型训练阶段,业务数据可以包括样本图像、样本视频、样本语音或者样本文本中的一项或多项,样本图像可以用于训练得到图像识别模型,样本视频可以用于训练得到视频识别模型或目标分析模型,样本文本可以用于训练得到语义识别模型等。又比如说,在模型应用阶段,业务数据可以是待识别的图像、视频、语音或者文本等,可以用于进行图像识别,人脸识别,语音识别或文本理解等。在一些实施例中,处理器 202 可以提供计算资源,存储装置 20 可以通过处理器 202 来处理业务数据。

硬盘 201 可以提供存储空间,硬盘 201 用于存储业务数据。其中,硬盘 201 可以为智能硬盘,智能硬盘具有自己的处理器、内存、DMA 控制器等组件,因此可以承担比传统硬盘更多的功能,例如通过 DMA 方式传输数据等。

AI 装置 21,用于向处理器 202 发送数据获取请求以获取业务数据,并对业务数据进行 AI 计算。AI 装置设置于存储设备 2 内部。AI 装置 21 为存储设备 2 提供 AI 计算的能力,通过运行 AI 装置 21,可以执行下述方法实施例中的 AI 计算方法。该 AI 装置 21 的形态可以是

一个芯片或者其他物理组件，例如可以是用于构建神经网络模型的训练芯片，也可以是利用神经网络模型进行推理的推理芯片。其中，数据获取请求用于请求获取硬盘 201 存储的业务数据。

AI 装置 21 和处理器 202 之间通过高速互连网络 22 通信。其中，高速互连网络 22 用于提供 AI 装置 21 和处理器 202 之间的数据通信功能，高速互连网络 22 可以是 PCIe、memory fabric、高速以太网、HCCS、无限带宽（infiniband，IB）、光纤通道（fibre channel，FC）中的任意一种，高速互连网络 22 的形态可以是总线，那么，高速互连网络 22 也可以称为高速互连交换机或高速互连总线。示例性地，存储设备 2 可以包括高速互连总线，AI 装置 21 和处理器 202 可以与该高速互连总线连接，从而接入高速互连网络。在一些可能的实施例中，AI 装置 21 可以包括高速互连网络接口，处理器 202 可以包括高速互连网络接口，AI 装置 21 通过 AI 装置 21 的高速互连网络接口与高速互连总线连接，处理器 202 通过处理器 202 的高速互连网络接口与高速互连总线连接。其中，高速互连网络接口可以为串行总线接口，具体地，高速互连网络接口可以为 PCIE 接口、HCCS 接口、以太网接口、IB 接口、FC 接口中的任意一种。如果高速互连网络接口的类型不同，AI 装置 21 与处理器 202 之间传输业务数据的速度可以也不同。经实验，基于各种不同类型的接口，可以将业务数据的加载速率提升 2-10 倍不等。此外，存储设备 2 可以通过 AI 装置 21 的高速互连网络接口，向 AI 装置 21 提供电能。应理解，高速互连总线仅是高速互连网络 22 的举例说明，高速互连网络 22 也可以不是高速互连总线，而是具有内存直通功能的其他总线，本实施例对高速互连网络 22 的具体类型并不做限定。

在存储设备中，通过高速互连网络 22 通信的部件可以有多种。例如，存储设备中的内存 203 和 AI 内存 210 之间可以通过高速互连网络 22 通信；又如，硬盘 201 和 AI 内存 210 之间可以通过高速互连网络 22 通信；又如，硬盘 201 和内存 203 之间可以通过高速互连网络 22 通信；又如，存储设备中的处理器 202 和 AI 处理器 213 之间可以通过高速互连网络 22 通信。此外，如果内存 203 的数量为多个，不同内存之间可以通过高速互连网络 22 通信；如果 AI 内存 210 的数量为多个，不同 AI 内存之间可以通过高速互连网络 22 通信，当然，上述进行通信的部件仅是举例说明，通过高速互连网络 22 相连的不同部件均可以通过高速互连网络 22 通信，本实施例对通过高速互连网络 22 通信的执行主体不做限定。

通过高速互连网络 22 进行通信的方式可以有多种，以下通过两种方式进行举例说明。

方式一、采用 DMA 的方式进行通信。DMA 的通信方式可以参见下述图 9 实施例以及上述对 DMA 技术原理的介绍，在此不做赘述。

方式二、采用内存结构（memory fabric）进行通信。

参见图 3，存储设备可以包括 memory fabric23，memory fabric23 可以集成高速互连网络 22 的功能和第一网络接口 211 的功能。AI 内存 210 和内存 203 可以通过 memory fabric23 连接，AI 内存 210 和内存 203 可以通过 memory fabric23 通信。此外，memory fabric23 也可以实现跨设备的内存之间的通信。例如，AI 内存 210 和其他存储设备的 AI 内存也可以通过 memory fabric23 连接，AI 内存 210 和其他存储设备的 AI 内存也可以通过 memory fabric23 通信；又如，内存 203 和其他存储设备的内存也可以通过 memory fabric23 连接，内存 203 和其他存储设备的内存也可以通过 memory fabric23 通信；AI 内存 210 和其他存储设备的内存也可以通过 memory fabric23 连接，AI 内存 210 和其他存储设备的内存也可以通过 memory

memory fabric23 通信；又如，内存 203 和其他存储设备的 AI 内存也可以通过 memory fabric23 连接，内存 203 和其他存储设备的 AI 内存也可以通过 memory fabric23 通信。此外，在一些实施例中，memory fabric23 还可以集成第二网络接口 204 的功能，或者集成存储设备中其他用于进行网络通信的部件的功能，此时，存储设备中内存之外的其他部件也可以通过 memory fabric23 来进行通信。

关于通过 memory fabric23 通信的方式，在一些可能的实施例中，可以将通过 memory fabric23 互联的每个内存组成内存资源池，为该内存资源池中的每个内存的内存空间统一分配对应的地址，使得该内存资源池中的每个内存的内存空间属于同一个地址范围。将业务数据或 AI 参数写入该内存资源池中任一个内存中的任一内存空间后，可以根据该内存空间的地址，获取业务数据或 AI 参数的元数据，此后根据该元数据，可以通过 memory fabric23 寻址至对应的内存空间，从内存空间中读取到业务数据或 AI 参数。其中，寻址的对象可以是存储设备本地的内存空间，也可以是远端存储设备的内存空间。

示例性地，参见图 4，第一内存、第二内存、第一 AI 内存、第二 AI 内存可以通过 memory fabric 通信，可以将第一内存、第二内存、第一 AI 内存、第二 AI 内存组成为系统 1 的内存资源池。例如，当第一 AI 处理器或第一处理器要获取 AI 参数时，会根据 AI 参数的元数据，确定 AI 参数在该内存资源池中所处的内存空间的地址，通过 memory fabric 寻址到该内存空间。

通过 memory fabric，可以实现同一存储设备中的内存与 AI 内存、不同存储设备的内存、不同存储设备的 AI 内存统一调度，从而提高存储系统调度内存资源和使用内存资源的效率。

AI 装置 21 与存储装置 20 通过高速互连网络 22 通信的效果至少可以包括：在 AI 计算的过程中，当 AI 装置 21 要加载 AI 计算所需的业务数据时，业务数据能够从存储装置 20 通过高速互连网络 22 发送至 AI 装置 21，载入 AI 装置 21 的 AI 内存 210。当 AI 装置 21 要存储 AI 计算的计算结果时，计算结果能够从 AI 装置 21 通过高速互连网络 22 发送至存储装置 20 的硬盘 201，得以在硬盘 201 中存储。如此，一方面，通过将业务数据的加载路径以及存储路径，从在存储设备 2 的外部，经过一条或多条网络链路以及一个或多个交换机的转发，改进为在存储设备 2 的内部，通过高速互连网络 22 进行传输，可以极大地缩短业务数据的传输路径，从而让 AI 装置 21 可以就近获取到存储装置 20 存储的业务数据，就近将计算结果存入存储装置 20。另一方面，通过将业务数据的传输方式，从通常基于 TCP/IP 协议的远程网络传输方式，改进为基于高速互连网络 22 的业务数据传输方式，由于高速互连网络 22 的传输速度通常远快于远程网络的传输速度，并且可以免去网络通信时所需进行的三次握手等建立网络通信连接的繁琐流程，因此，AI 装置 21 能够从存储装置 20 快速地加载 AI 计算所需的业务数据，将 AI 计算的计算结果快速地存入存储装置 20 的硬盘 201。经实验，每一次加载业务数据可以缩短 30% 以上的时延。

应理解，存储设备 2 包括的 AI 装置 21 的数量可以为一个或多个，AI 装置 21 的数量可以根据存储设备 2 的构造、空间或需求设置，本实施例对存储设备 2 包括多少个 AI 装置 21 不做限定。

如果存储设备 2 中包括多个 AI 装置 21，多个 AI 装置 21 可以通过高速互连网络 22 进行通信，例如通过高速互连网络 22 交换 AI 参数。其中，存储设备 2 的不同的 AI 装置 21 的形状和构造可以相同，或者根据实际需求存在细微的差异。

存储设备 2 可以是分布式控制的存储设备 2, 也可以是集中式控制的存储设备 2。存储设备 2 的形态可以而不限存储服务器、存储阵列、或者其他专用的存储设备 2, 存储设备 2 可以而不限于运行在云环境、边缘环境或终端环境。

本实施例提供的存储设备, 通过设置有存储装置以及 AI 装置, 该存储设备既可以通过 AI 装置, 提供 AI 计算的能力, 又可以通过存储装置, 提供业务数据存储的能力, 从而实现了存储与 AI 算力的融合, AI 参数以及业务数据在存储设备的内部通过高速互联网络进行传输, 而无需经过外部网络的转发, 因此极大地缩短了业务数据以及 AI 参数的传输路径, 能够近距离地载入业务数据, 从而提高了加载速度。此外, AI 装置能够借用存储装置的计算资源来处理业务数据, 从而通过算力的协同, 提高了 AI 装置的算力, 加快 AI 计算的速度。

上述实施例中, AI 装置 21 可以通过向处理器 202 发送数据获取请求, 来获取硬盘 201 中的业务数据, 以便根据业务数据来进行 AI 计算。通过这种方式, AI 装置 21 从本地的存储设备即可得到业务数据, 免去了通过网络向远端存储设备请求业务数据带来的通信开销, 从而节省了获取业务数据的时延。

其中, AI 装置 21 获取业务数据的实现方式具体可以包括多种, 相应地, AI 装置发送的数据获取请求可以包括多种情况。不同情况下的数据获取请求可以存在细微的差异, 以下通过方式一、方式二和方式三进行举例说明。为了区分描述, 将方式一中的数据获取请求称为第一数据获取请求, 将方式二中的数据获取请求称为第二数据获取请求, 将方式三中的数据获取请求称为第三数据获取请求:

方式一、存储设备可以利用处理器 202, 将硬盘 201 中存储的业务数据传输至 AI 装置 21。具体来讲, AI 装置 21, 还用于向处理器 202 发送第一数据获取请求, 第一数据获取请求用于请求获取业务数据。处理器 202, 还用于接收第一数据获取请求, 响应于第一数据获取请求, 从硬盘 201 中获取业务数据, 并将业务数据发送给 AI 装置 21。

方式二、存储设备可以通过 DMA 的方式, 将硬盘 201 中存储的业务数据传输至 AI 装置 21。具体来讲, AI 装置 21, 还用于向处理器 202 发送第二数据获取请求。第二数据获取请求用于请求获取业务数据。处理器 202, 还用于接收第二数据获取请求, 响应于第二数据获取请求, 将业务数据的元数据发送给 AI 装置 21; AI 装置 21 可以根据元数据, 判断业务数据是否位于存储设备本地, 当元数据指示业务数据位于存储设备本地时, AI 装置 21 向硬盘 201 发送第一数据访问请求, 第一数据访问请求包括元数据, 例如, 第一数据访问请求可以是 DMA 请求。硬盘 201, 用于根据元数据获取业务数据, 并通过 DMA 方式将业务数据写入 AI 装置 21。其中, 元数据用于指示业务数据的地址, 在方式二中, 业务数据的地址位于存储设备本地。

其中, 硬盘 201 与 AI 装置 21 可通信, 因此能够实现上述基于第二数据访问请求的交互流程。在一个示例性实施例中, 硬盘 201 可以包括高速互联网络接口, 硬盘 201 可以通过高速互联网络接口与 AI 装置 21 相连, AI 装置 21 可以通过高速互联网络接口对应的控制器或驱动来读写硬盘 201。例如, 硬盘 201 的高速互联网络接口可以是串行连接 SCSI (Serial Attached SCSI, SAS) 接口, AI 装置 21 可以通过 SAS 控制器, 与 SAS 接口进行通信, 从而读写硬盘 201。当然, 硬盘 201 的高速互联网络接口还可以是 SAS 接口之外的其他接口, 例如高技术配置 (advanced technology attachment, ATA) 接口、电子集成驱动器 (integrated drive electronics, IDE) 接口、FC、小型计算机系统接口 (small computer system interface, SCSI)

等。

方式三、存储设备可以通过 RDMA 的方式，将其他存储设备中存储的业务数据传输至 AI 装置 21。具体来讲，AI 装置 21，还用于向处理器 202 发送第三数据获取请求。第三数据获取请求用于请求获取业务数据。处理器 202，还用于接收第三数据获取请求，响应于第三数据获取请求，将业务数据的元数据发送给 AI 装置 21；AI 装置 21 可以根据业务数据的元数据，判断业务数据是否位于其他存储设备。当元数据指示业务数据位于其他存储设备时，AI 装置 21 向其他存储设备发送第二数据访问请求，第二数据访问请求包括元数据，例如，第二数据访问请求可以是 RDMA 请求。其他存储设备可以响应于第二数据访问请求，根据元数据获取业务数据，并通过 RDMA 方式将业务数据写入 AI 装置 21。其中，元数据用于指示业务数据的地址，在方式三中，业务数据的地址位于其他存储设备。

AI 装置 21 可以包括 AI 内存 210，AI 内存 210 用于缓存 AI 参数和/或 AI 算力单元 212 或 AI 处理器 213 刚用过或循环使用的业务数据。如果 AI 算力单元 212 或 AI 处理器 213 需要再次使用业务数据，可从 AI 内存 210 中直接调用，从而避免重复存取，减少了 AI 算力单元 212 或 AI 处理器 213 的等待时间，提高了计算效率。在进行 AI 计算的过程中，AI 内存 210 可以缓存 AI 计算的输入业务数据、中间结果或者最终结果。示例性地，AI 内存 210 可以是高速缓冲存储器。AI 内存 210 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、通用闪存存储器(universal flash storage, UFS)等。

AI 装置 21，用于向硬盘 201 发送第一数据访问请求，第一数据访问请求包括元数据。元数据(metadata)用于指示业务数据的地址。作为示例，元数据可以包括磁道标识和扇区标识，磁道标识用于标识业务数据在硬盘中所处的磁道，例如，磁道标识可以是磁道 ID；扇区标识用于标识业务数据在硬盘中所处的扇区，例如，扇区标识可以是扇区 ID。硬盘根据磁道标识，可以找到对应的磁道，根据扇区标识，可以在该磁道上找到对应的扇区，然后读取该扇区中的数据，得到 AI 装置 21 通过第一数据访问请求来请求的业务数据。应理解，元数据指示地址的方式可以有多种，例如，元数据可以通过指示业务数据的起始地址和长度的方式，来指示业务数据的地址。又如，元数据可以通过指示业务数据的起始地址和结束地址的方式，来指示业务数据的地址，本实施例对此不做限定。

硬盘 201，用于根据元数据获取业务数据，并通过 DMA 方式将业务数据写入 AI 内存 210。

可选地，存储设备 2 与其他存储设备通信。例如，存储设备 2 可以是图 1 所示系统 1 中的第一存储设备 10，其他存储设备可以是图 1 所示系统 1 中的第二存储设备 11，第一存储设备 10 可以与第二存储设备 11 通信。

在一种可能的实现方式中，AI 装置 21 可以包括第一网络接口 211，第一网络接口 211 的形态可以为网卡，第一网络接口 211 可以提供网络通信的能力，AI 装置 21 能够通过第一网络接口 211 与其他存储设备 2 通信，例如通过第一网络接口 211 与其他存储设备 2 交换 AI 参数。可选地，第一网络接口 211 可以支持远程直接内存访问(remote direct memory access, 简称: RDMA)的功能，第一网络接口 211 可以是远程直接数据存取网络接口控制器(RDMA network interface controller, 简称: RNIC)。

在一个示例性应用场景中，在训练或推理的过程中，如果 AI 装置 21 计算出 AI 参数，AI 装置 21 可以通过第一网络接口 211，向其他存储设备 2，发送 AI 参数，以便其他存储设备 2 接收 AI 参数。同理地，如果其他存储设备 2 计算出 AI 参数，其他存储设备 2 可以将 AI

参数发送至第一网络接口 211，第一网络接口 211 可以接收 AI 参数。

AI 装置 21，还用于根据元数据确定业务数据位于其他存储设备 2 时，向其他存储设备 2 发送第二数据访问请求，使得业务数据以 RDMA 方式写入 AI 内存 210，第二数据访问请求包括元数据。

可选地，AI 装置 21 还包括 AI 算力单元 212。AI 算力单元 212 用于提供 AI 算力。在一个示例性实施例中，AI 算力单元 212 上可以运行 AI 算法，从而进行模型训练或推理预测，其中，该 AI 算法可以为神经网络模型，从本质上讲，AI 算法包括矩阵或向量的乘法以及加法，另外还可以包括除法以及指数运算。AI 算力单元 212 可以包括一个或多个图形处理器 (graphics processing unit, GPU)、神经网络处理器 (neural-network processing unit, NPU)、张量处理单元 (tensor processing unit, TPU)、现场可编程门阵列 (field-programmable gate array, FPGA)、专用集成电路 (application specific integrated circuit, ASIC)、类脑芯片、可重构通用 AI 芯片、CPU、可编程逻辑器件 (programmable logic device, PLD)、控制器、状态机、门逻辑、分立硬件部件、或者其他能够提供 AI 算力的电路的任意组合。AI 算力单元 212 可以包括一个或多个处理核心。

AI 算力单元 212 具体用于从 AI 内存 210 中获取业务数据，并进行 AI 计算。

可选地，AI 装置 21 包括 AI 处理器 213，AI 处理器 213 和 AI 内存 210 之间可以通过高速互连网络 22 相连。AI 处理器 213 可以为 CPU，AI 处理器 213 可以用于进行管理和资源调度。AI 处理器 213 可以运行操作系统或者应用程序。AI 处理器 213 可以包括一个或多个处理核心。

可选地，存储装置 20 包括处理器 202 和内存 203，处理器 202 和内存 203 之间可以通过高速互连网络 22 相连。处理器 202 可以是存储装置 20 的中央处理器 (central processing unit, CPU)。处理器 202 可以采用数字信号处理 (digital signal processing, DSP)、FPGA、可编程逻辑阵列 (programmable logic array, PLA) 中的至少一种硬件形式来实现。

在一些可能的实施例中，存储设备可以是集中式控制的存储设备。具体地，存储设备可以包括一个或多个控制器，可以通过控制器来控制硬盘。此外，存储设备还可以包括一个或多个级联板，级联板可以用于级联存储设备中的不同硬盘，级联板可以和硬盘以及控制器连接。例如，参见图 5，存储设备可以是双控的架构，双控架构是指存储设备中控制器的数量是 2 个，例如图 5 中，系统 1 可以包括控制器 1 和控制器 2 这 2 个控制器。又如，参见图 6，存储设备可以是四控架构，四控架构是指存储设备中控制器的数量是 4 个，例如图 6 中，系统 1 可以包括控制器 1、控制器 2、控制器 3 和控制器 4 这 4 个控制器。

其中，如果存储设备是集中式控制的存储设备，AI 装置可以设置在控制器内部，那么，控制器可以在实现原有的控制硬盘的功能的基础上，还提供 AI 算力。如果存储设备包括多个控制器，可以在每个控制器中均设置 AI 装置，也可以在部分控制器中设置 AI 装置，在另一部分控制器中不设置 AI 装置，本实施例对此不做限定。

内存 203：用于缓存业务数据，可选地，在支持 AI 内存 210 与内存 203 直通的功能下，内存 203 也可以用于缓存 AI 参数。通过内存 203 来缓存业务数据和/或 AI 参数，处理器 202、AI 算力单元 212 或 AI 处理器 213 需要使用业务数据和/或 AI 参数时，可从内存 203 中直接调用，从而避免重复存取，减少了处理器 202、AI 算力单元 212 或 AI 处理器 213 的等待时间，提高了计算效率。示例性地，内存 203 可以是高速缓冲存储器。内存 203 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、通用闪存存储器(universal flash storage, UFS)等。

AI 装置 21 包括 AI 处理器 213, AI 处理器 213 和 AI 内存 210 之间可以通过高速互联网络 22 相连。AI 处理器 213 可以为 CPU, AI 处理器 213 可以用于进行管理和资源调度。在一个示例性实施例中, AI 处理器 213 可以运行操作系统或者应用程序

通过在 AI 装置 21 中设置 AI 处理器 213, 可以将 AI 计算任务和非 AI 计算任务分配给不同的处理器执行, 通过 AI 算力单元 212 执行 AI 计算任务, 通过 AI 处理器 213 执行非 AI 计算任务, 从而实现计算任务的负载分担, 免去 AI 算力单元 213 为执行非 AI 计算任务而导致算力被占用的情况。在一些实施例中, 存储设备还可以使用 AI 处理器 201 可以分担处理器 202 的计算任务, 例如, 处理器 202 可以将从硬盘中获取业务数据的任务发送给 AI 处理器 201, AI 处理器 201 可以执行从硬盘中获取业务数据的任务, 将得到的业务数据发送给 AI 处理器 201。外, 能够利用不同处理器各自的性能优势, 比如利用 CPU 对逻辑控制的优势, 将 CPU 作为 AI 处理器 213 以进行资源调度和管理, 利用 GPU 或 NPU 对浮点运算以及并行计算的优势, 将 GPU 或 NPU 作为 AI 算力单元 212 来进行模型训练或者其他 AI 计算, 从而提高 AI 计算的整体效率, 帮助 AI 加速。此外, 能够避免 AI 计算的过程与资源调度管理的过程相互干扰, 提高 AI 计算的整体效率, 帮助 AI 加速。

可选地, 处理器 202, 还用于从内存 203 中划分一段内存空间预留给 AI 装置 21 使用。

可选地, AI 处理器 213, 用于当 AI 内存 210 的可用容量达到预设阈值时, 向处理器 202 发送内存申请请求。处理器 202, 用于基于内存申请请求, 从内存 203 中划分一段内存空间预留给 AI 装置 21 使用。

可选地, AI 内存 210 的可用容量通过设定的 batchsize 确定。

可选地, AI 处理器 213, 用于将计算任务划分为至少两个子任务, 将至少两个子任务中的第一子任务发送给处理器 202; 处理器 202, 还用于执行第一子任务, 并将计算结果发送给 AI 处理器 213。这一可选方式可以应用在多种情况下, 例如, AI 处理器可以当确定 AI 处理器 213 的算力不足时, 执行该可选方式。

可选地, 存储装置 20 可以包括第二网络接口 204。第二网络接口 204 用于进行网络通信, 第二网络接口 204 可以是网卡, 第二网络接口 204 可以通过高速互联网络 22 与 AI 装置 21、处理器 202、内存 203 连接。

可选地, 存储装置 20 可以包括业务接口 205。

在一些可能的实施例中, 存储装置 20 可以用于将操作系统加载至 AI 装置 21, 通过操作系统启动 AI 装置 21。具体来讲, 存储装置 20 可以存储操作系统的镜像文件, 如果存储装置 20 要启动 AI 装置 21 时, 存储装置 20 可以通过高速互联网络 22, 向 AI 装置 21 发送操作系统的镜像文件, 将操作系统的镜像文件加载至 AI 装置 21 的 AI 内存 210, 基于操作系统的镜像文件运行操作系统, 从而启动 AI 装置 21。其中, 操作系统可以为 LINUX™、UNIX™、或 WINDOWS™等。

通过由存储装置 20 将操作系统加载至 AI 装置 21, AI 装置 21 可以无需安装操作系统, 从而避免操作系统占用 AI 装置 21 的存储空间, 因此可以保证 AI 装置 21 的存储容量。尤其是, 如果存储装置 20 包括多个 AI 装置 21, 通常来讲, 多个 AI 装置 21 的操作系统均相同, 因此存储装置 20 可以将同一操作系统, 批量化的分别加载至每个 AI 装置 21, 从而批量化的

启动各个 AI 装置 21。那么，既能够避免每个 AI 装置 21 分别安装相同的操作系统，导致相同的操作系统在每个 AI 装置 21 上均占用存储空间的情况，从而节省每个 AI 装置 21 的存储空间，提高存储效率，并且，可以通过批量化的启动方式，提高多个 AI 装置 21 的启动效率。

可选地，AI 装置 21 还可以包括非易失性存储介质，下述方法实施例的全部或部分步骤可以以计算机软件产品的形式体现出来，该计算机软件产品可以存储在 AI 装置 21 的非易失性存储介质中或者上面描述的 AI 内存 210 中，计算机软件产品包括一个或多个指令，用以使得 AI 装置 21 执行全部或部分步骤。

可选地，AI 装置 21 和存储装置 20 可以集成在一起，作为一套产品共同销售或使用，AI 装置 21 与存储装置 20 不相互分离。在另一些可能的实施例中，AI 装置 21 也可以作为一个独立的产品销售，并配合存储装置 20 使用，例如 AI 装置 21 可以插入存储设备 2 的高速互联总线，或者从存储设备 2 的高速互联总线拔出，从而实现和存储装置 20 的接触和分离。

应理解，图 2 所示的存储设备 2 的每个部件的数量可以更多或更少，比如部件的数量可以为一个，或者可以为几十个或几百个，或者更多数量。例如，AI 装置 21 可以包括多个 AI 算力单元 212，本实施例对存储设备 2 的每个部件的数量不做限定。

应理解，本实施例示意的结构并不构成对存储设备 2 的具体限定。在本申请另一些实施例中，存储设备 2 可以包括比图 2 所示的更多或更少的部件，或者将图 2 中的两个或两个以上的部件进行组合，或者将图 2 中某一个部件拆分为两个或两个以上的节点，或者将图 2 中的两个或两个以上的部件布置在不同的位置。

以上介绍了本申请实施例提供的分布式存储系统以及存储设备，以下示例性介绍应用于该分布式存储系统以及存储设备的方法流程。

图 7 是本申请实施例提供的一种数据处理方法的流程图，该方法可以应用于存储设备，该存储设备可以是图 1 或图 2 所示的任意一个存储设备，该方法包括：

S71，存储设备的处理器将业务数据存储至硬盘中。

例如，客户端可以将业务数据发送给存储设备，存储设备可以接收来自客户端的业务数据，将业务数据写入至硬盘，从而通过硬盘来存储业务数据。

S72，存储设备的 AI 装置向存储设备的处理器发送数据获取请求以获取业务数据，并对业务数据进行 AI 计算。

可选地，存储设备的处理器的计算资源以及 AI 装置的计算资源可以组成资源池，存储设备可以调度资源池中的资源，来处理 AI 计算。通过协同处理器的算力以及 AI 装置的算力，能够提高存储设备的整体的 AI 算力，从而加速 AI 计算的过程。

其中，AI 计算的处理过程可以而不限于下述情况（1）至（2）中的一项或多项。

（1）模型训练。

具体地，存储设备可以通过存储设备自身的处理器的计算资源以及 AI 装置的计算资源对样本集进行模型训练，得到 AI 模型。示例性地，AI 模型可以是神经网络模型。比如，可以通过样本图像训练得到图像识别模型，该图像识别模型可以而不限于是卷积神经网络（convolutional neural networks, CNN），当然也可以是建立在 CNN 基础上的其他神经网络，比如区域 CNN（region-CNN, R-CNN）、快的 R-CNN（fast R-CNN）、更快的 R-CNN（faster R-CNN）、单阶段多框检测（single shot multibox detector, SSD）等。又如，AI 装置可以通

过样本文本，训练得到语义识别模型，该语义识别模型可以而不限于是循环神经网络(recurrent neural networks, RNN)。又如，AI 装置可以通过样本语音，训练得到语音识别模型。

(2) 推理预测。

具体地，存储设备可以将待识别的业务数据输入至已训练的 AI 模型，通过处理器的计算资源以及 AI 装置的计算资源，使用 AI 模型进行业务数据推理，得到业务数据的识别结果。比如，可以将待识别的图像输入至图像识别模型，输出图像识别结果。该图像识别结果可以而不限于是图像的类别、图像的特征或图像中物体的位置中的一项或多项。该图像的类别可以指示图像中包含了哪一种物体，比如，如果是人脸识别的场景，图像识别模型为人脸识别模型，图像的类别可以为人脸的类别；如果是场景识别的场景，图像识别模型为场景识别模型，图像的类别可以为场景的类别，比如是天花板，还是草坪，还是地面；如果是字符识别的场景，图像识别模型为字符识别模型。图像的特征可以而不限于是是一维的特征值、二维的特征图、三维的特征立方体或者更高维的张量。图像中物体的位置可以通过物体所处的边界框(bounding box)的坐标来表示。又如，可以将待识别的文本输入至语义识别模型，输出语义识别结果。又如，可以将待识别的语音输入至语音识别模型，输出语音识别结果。

本实施例提供的存储设备，通过设置有处理器以及 AI 装置，该存储设备既可以通过 AI 装置，提供 AI 计算的能力，又可以通过处理器，提供业务数据存储的能力，从而实现了存储与 AI 算力的融合，AI 参数以及业务数据在存储设备的内部通过高速互连网络进行传输，而无需经过外部网络的转发，因此极大地缩短了业务数据以及 AI 参数的传输路径，能够近距离地加载业务数据，从而提高了加载速度。此外，AI 装置能够借用处理器的计算资源来处理业务数据，从而提高了 AI 装置的算力，加快 AI 计算的速度。

上述图 7 实施例中，AI 装置获取业务数据的实现方式具体可以包括多种，以下通过图 8 实施例、图 9 实施例和图 10 实施例进行举例说明。

在一些实施例中，AI 装置能够从存储设备本地就近获取业务数据。以下通过图 8 实施例和图 9 实施例，对就近获取业务数据的实现方式进行举例说明。

示例性地，参见图 8，图 8 是本申请实施例提供的一种数据处理方法的流程图，AI 装置就近获取业务数据的流程可以包括以下步骤：

S81，AI 处理器向处理器发送第一数据获取请求。

AI 处理器可以生成第一数据获取请求，通过高速互连网络，将第一数据获取请求发送给处理器。其中，第一数据获取请求用于请求硬盘中存储的业务数据，第一数据获取请求可以携带业务数据的标识，业务数据的标识可以是业务数据的 ID。

S82，处理器响应于第一数据获取请求，从硬盘中获取业务数据，并将业务数据发送给 AI 装置。

处理器可以通过高速互连网络，接收第一数据获取请求。处理器可以解析第一数据获取请求，得到第一数据获取请求携带的业务数据的标识。处理器可以根据业务数据的标识，确定业务数据在硬盘中的地址。处理器可以根据业务数据的地址，访问硬盘，得到硬盘存储的业务数据。处理器可以通过高速互连网络，将业务数据返回给 AI 装置。

S83，AI 装置接收处理器发送的业务数据，对业务数据进行 AI 计算。

本实施例提供了一种 AI 装置就近获取业务数据的方法，由于存储设备内部包含了 AI 装置、处理器和硬盘，当 AI 装置要获取业务数据时，通过向处理器发送数据获取请求，由存储设备的处理器从本地的硬盘中获取业务数据，发送给 AI 处理器，使得 AI 处理器能够从本地得到业务数据，免去了通过网络向远端存储设备请求业务数据带来的通信开销，从而节省了获取业务数据的时延。

上述图 8 实施例提供了一种就近获取业务数据的流程，在本申请的另一些实施例中，就近获取业务数据也可以通过其他实现方式实现。以下通过图 9 实施例示例性说明。

参见图 9，图 9 是本申请实施例提供的一种数据处理方法的示意图，AI 装置就近获取业务数据的流程可以包括以下步骤：

S91，AI 装置向处理器发送第二数据获取请求。

当 AI 装置要从处理器获取业务数据时，AI 处理器可以生成第二数据获取请求，通过高速互联网络，将第二数据获取请求发送给处理器。其中，第二数据获取请求用于请求获取业务数据，第二数据获取请求可以携带业务数据的标识，该业务数据的标识可以是业务数据的 ID。

S92，处理器响应于第二数据获取请求，将业务数据的元数据发送给 AI 装置，元数据用于指示业务数据的地址。

处理器可以解析第二数据获取请求，得到第二数据获取请求携带的业务数据的标识，获取该业务数据的元数据，将元数据通过高速互联网络发送给 AI 处理器。其中，元数据用于指示业务数据的地址，即业务数据的存储位置。例如，元数据可以指示业务数据的起始地址以及业务数据的长度，或者，元数据可以指示业务数据的起始地址和结束地址。

参见图 10，其示出了一种 AI 处理器和存储设备的处理器进行交互的示意图。

可选地，处理器可以根据业务数据的地址，判断业务数据是否能够通过 DMA 的方式发送给 AI 装置，如果业务数据能够通过 DMA 的方式发送给 AI 装置，则将业务数据的元数据发送给 AI 装置，以触发图 9 实施例提供的方法流程。此外，处理器可以判断业务数据是否能够通过 RDMA 的方式发送给 AI 装置，如果业务数据能够通过 RDMA 的方式发送给 AI 装置，则将业务数据的元数据发送给 AI 装置，以触发图 12 实施例提供的方法流程。此外，如果业务数据不能通过 DMA 的方式或 RDMA 的方式发送给 AI 装置，则将业务数据从硬盘载入内存，再将业务数据从内存传输至 AI 内存。

S93，当元数据指示业务数据位于存储设备本地时，AI 装置向硬盘发送第一数据访问请求。

第一数据访问请求包括业务数据的元数据，第一数据访问请求可以是 DMA 请求。在一个示例性实施例中，AI 装置的 AI 处理器可以根据业务数据的元数据，确定业务数据的地址，判断业务数据的地址在本地的硬盘还是在其他存储设备，如果确定业务数据的地址位于本地的硬盘，则根据元数据，生成第一数据访问请求，通过高速互联网络，将第一数据访问请求发送给硬盘。

S94，硬盘根据元数据获取业务数据，并通过 DMA 方式将业务数据写入 AI 装置。

硬盘可以解析第一数据访问请求，得到第一数据访问请求携带的元数据，根据元数据确定业务数据的地址，访问该地址，得到业务数据。硬盘和 AI 内存之间可以建立 DMA 通道，

硬盘可以将业务数据通过该 DMA 通道发送到 AI 内存中。其中，硬盘可以是智能盘，硬盘可以包括 CPU，硬盘可以通过自身的 CPU 来执行写入业务数据的操作。

示例性地，参见图 11，图 11 是本申请实施例提供的一种业务数据传输的示意图，从图 11 可以看出，硬盘存储的业务数据能够通过高速互连网络，以 DMA 的方式传输至 AI 内存。

本实施例提供的方法，可以实现 AI 装置和硬盘之间的 DMA 直通，通过在 AI 装置和硬盘之间建立 DMA 通路，AI 装置和硬盘能够利用 DMA 通路进行快速业务数据交换，从而提高 AI 装置载入业务数据的速度，进而提高 AI 装置可同时处理的业务数据量，从而降低 AI 装置之间的传输 AI 参数的通信开销，加快 AI 训练的速度。

在另一些实施例中，AI 装置能够从其他存储设备获取业务数据。以下通过图 12 实施例，对从其他存储设备获取业务数据进行举例说明。具体地，参见图 12，本申请实施例提供了一种数据处理方法，该方法可以应用于存储设备，该存储设备可以是图 1 或图 2 所示的任意一个存储设备，包括以下步骤。

S1201，AI 装置向处理器发送第三数据获取请求。

S1202，处理器响应于第三数据获取请求，将业务数据的元数据发送给 AI 装置。

S1203，当元数据指示业务数据位于其他存储设备时，AI 装置向其他存储设备发送第二数据访问请求。

AI 处理器可以根据业务数据的元数据，确定业务数据的地址，根据业务数据的地址，判断业务数据存储在本地产还是在存储在其他存储设备，如果确定业务数据存储在其他存储设备，则 AI 处理器生成第二数据访问请求，将第二数据访问请求发送给第一网络接口，第一网络接口可以将第二数据访问请求发送给其他存储设备。其中，第二数据访问请求可以包括业务数据的元数据。第二数据访问请求可以是 RDMA 请求，第二数据访问请求可以包括存储设备的 AI 内存中的目的地址，以指示其他存储设备将业务数据写入至该目的地址。

S1204，其他存储设备响应于第二数据访问请求，将业务数据发送至 AI 装置。

例如，其他存储设备可以采用 RDMA 方式，将业务数据写入 AI 内存。具体地，其他存储设备可以解析第二数据访问请求，得到第二数据访问请求携带的业务数据的元数据，根据业务数据的元数据，确定业务数据的地址，访问存储介质中的该地址，得到业务数据。此外，其他存储设备可以根据第二数据访问请求，建立与 AI 装置的 AI 内存之间的 RDMA 通路，将业务数据通过 RDMA 通路写入到 AI 装置的 AI 内存。具体地，其他存储设备可以根据 AI 内存的目的地址以及业务数据，生成 RDMA 消息，向 AI 装置发送 RDMA 消息，该 RDMA 消息包括业务数据以及 AI 内存的目的地址，AI 装置的第一网络接口可以接收 RDMA 消息，解析 RDMA 消息，得到 RDMA 消息携带的业务数据以及 AI 内存的目的地址，通过高速互连网络访问 AI 内存，将业务数据写入到该 AI 内存的目的地址中。

示例性地，参见图 13，图 13 是本申请实施例提供的一种业务数据传输的示意图，从图 13 可以看出，其他存储设备存储的业务数据能够通过高速互连网络，以 RDMA 的方式传输至 AI 内存。

通过这种可选方式，AI 装置的 AI 内存与其他存储设备实现了 RDMA 直通，通过 AI 内存和其他存储设备进行快速业务数据交换，加快 AI 训练的速度。

可选地，AI 装置可以通过 AI 算力单元来提供 AI 算力。例如，参见图 14，本申请实施例提供了一种 AI 计算的处理方法，该方法可以应用于存储设备，该存储设备可以是图 1 或图 2 所示的任意一个存储设备，该方法的执行主体为存储设备的 AI 算力单元，包括以下步骤。

S1401，AI 算力单元从 AI 内存中获取业务数据。

AI 算力单元和 AI 内存之间可以通过高速互联网络通信，AI 算力单元可以通过高速互联网络，访问 AI 内存，得到 AI 内存缓存的业务数据。

S1402，AI 算力单元对业务数据进行 AI 计算。

进行 AI 计算的细节可以参见上述图 7 实施例的步骤 S72，在此不做赘述。

通过这种可选方式，AI 装置能够通过 AI 算力单元来提供 AI 算力，从而避免 AI 计算占用存储装置本身的算力，也就可以避免 AI 计算对存储设备本身的性能造成严重影响。

可选地，存储设备的内存和 AI 内存可以实现内存资源的协同。实现内存资源协同的方式可以有多种，以下通过图 15 实施例和图 16 实施例举例说明。

在一些实施例中，AI 装置能够借用存储装置的内存来进行 AI 计算。参见图 15，本申请实施例提供了一种数据处理方法，该方法可以应用于存储设备，该存储设备可以是图 1 所示的系统 1 中的第一存储设备 10 或第二存储设备 11，也可以是图 2 所示的存储设备 2，该方法的交互主体包括存储设备的 AI 处理器以及该存储设备的处理器，包括以下步骤。

S1501，AI 处理器确定 AI 内存的可用容量达到预设阈值。

S1502，AI 处理器向处理器发送内存申请请求。

其中，内存申请请求用于借用处理器的内存，内存申请请求可以携带待借用的内存空间的大小。可以理解的是，S1501 中所规定的触发条件仅是一种示例，AI 处理器也可以在其他情况下，或者没有任何触发条件的情况下向处理器发送内存申请请求。

S1503，处理器基于内存申请请求，从内存中划分一段内存空间预留给 AI 处理器使用。

S1504，AI 处理器使用存储设备的内存中的内存空间进行 AI 计算。

AI 处理器可以通过 PCIe 总线或 memory fabric，访问存储设备的处理器预留的内存空间，使用内存的内存空间以及 AI 内存的内存空间来进行 AI 计算。在进行 AI 计算的过程中，内存可以和 AI 内存之间通过 DMA、PCIe 或 memory fabric 进行快速交换，那么由于内存空间更大，可以提高单个 AI 处理器一批可处理的数据量，降低不同 AI 处理器之间的参数通信开销，以加速 AI 训练的速度。此外，当 AI 处理器停止进行 AI 计算，不再使用存储设备的内存的内存空间时，AI 处理器可以向处理器发送内存释放请求，使得处理器释放预留的内存空间，将内存空间还给处理器使用。

相关技术中，GPU 的显存是固定的，这就导致进行 AI 计算时经常面临内存不足的问题。而通过这种可选方式，AI 装置能够借用存储装置的内存来进行 AI 计算，从而扩大了 AI 装置的可用内存空间，让 AI 装置能够在更大的内存上进行 AI 计算，从而提高 AI 计算的效率。

需要说明的一点是，存储设备的处理器在 AI 处理器发送了内存申请请求的触发条件下，划分内存空间给 AI 处理器的方式仅是可选方式。在另一些实施例中，存储设备的处理器也可以在其他情况下，从内存中划分一段内存空间预留给 AI 装置使用。例如，存储设备的处理器可以主动划分内存空间，预留给 AI 装置，例如，存储设备的处理器可以当确定内存的可用容量高于预设阈值时，则划分内存空间预留给 AI 装置使用，又如，存储设备的处理器可以监控

AI 内存的可用容量，存储设备的处理器确定 AI 内存的可用容量达到预设阈值时，则划分内存空间预留给 AI 装置使用。应理解，本实施例并不限定存储设备的处理器划分内存空间预留给 AI 处理器这一特征的应用场景，在一些实施例中，存储设备的处理器可以在任意预先设定的条件下，划分内存空间预留给 AI 处理器。

可选地，AI 内存的可用容量通过设定的批尺寸 (batchsize) 确定。其中，批尺寸为一次训练使用的数据数量。具体地，AI 处理器可以根据设定的批尺寸，确定 AI 内存的可用容量，即要基于设定的批尺寸进行 AI 计算，需要多大容量的 AI 内存。AI 处理器可以对 AI 内存的可用容量与预设阈值进行比较，如果 AI 内存的可用容量达到预设阈值，表明 AI 内存不足，则触发执行借用存储设备的内存的流程。

通过这种可选方式，AI 处理器可以利用存储设备的内存的内存空间来进行训练，那么由于可用的内存空间更大，能够提高 AI 训练的批尺寸，从而提高 AI 装置一批可处理的业务数据量，降低不同 AI 装置之间交换 AI 参数的通信开销，提高 AI 训练的速度。经实验，如果仅通过 AI 内存进行 AI 训练，批尺寸最高设定为 256，而通过这种方式，可以将批尺寸设定为 32000，由此可见，批尺寸得到显著提升。

在一些实施例中，存储设备在读写 AI 参数的过程中，可以优先使用 AI 内存，其次使用存储设备的内存，以便利用 AI 内存通常比存储设备的内存性能更好的优势（例如 AI 内存的访问速度可能高于内存），来加快 AI 参数的读写速度。具体地，可以将 AI 内存为第一层级，以内存为第二层级，对 AI 参数进行分层缓存，其中，第一层级的优先级高于第二层级的优先级。通过这一方式，实现了存储设备的内存与 AI 内存之间的分层，会优先将 AI 参数缓存在 AI 内存中。作为一种可能的实现，介质控制器可以建立 AI 内存和存储设备的内存之间的映射关系，如果 AI 内存溢出，则将 AI 内存缓存的 AI 参数存入存储设备的内存。

在一些实施例中，存储装置能够借用 AI 装置的内存来处理业务数据的读写，从而加速访问业务数据的速度或存储业务数据的速度。参见图 16，本申请实施例提供了一种数据处理方法，该方法可以应用于存储设备，该存储设备可以是图 1 或图 2 所示的任意一个存储设备，包括以下步骤。

S1601，存储设备的处理器确定存储设备的内存的可用容量达到预设阈值。

S1602，存储设备的处理器向 AI 处理器发送内存申请请求。

其中，内存申请请求用于借用 AI 处理器的 AI 内存，内存申请请求可以携带待借用的内存空间的大小。可以理解的是，S1111 中所规定的触发条件仅是一种示例，存储设备的处理器也可以在其他情况下，或者没有任何触发条件的情况下向 AI 处理器发送内存申请请求。

S1603，AI 处理器基于内存申请请求，从 AI 内存中划分一段内存空间预留给存储设备的处理器使用。

S1604，存储设备的处理器使用 AI 内存中的内存空间进行 AI 计算。

存储设备的处理器可以通过高速互联网络或 memory fabric，访问 AI 处理器预留的内存空间，使用 AI 内存的内存空间以及内存的内存空间来读写业务数据。在读写业务数据的过程中，AI 内存可以和内存之间通过 DMA、高速互联网络或 memory fabric 进行快速交换，那么

由于内存空间更大，可以提高单个处理器一批可处理的数据量，降低不同处理器之间的参数通信开销，以加速业务数据的读写。此外，当存储设备的处理器停止读取业务数据或写入业务数据，不再使用 AI 内存的内存空间时，处理器可以向 AI 处理器发送内存释放请求，使得 AI 处理器释放预留的内存空间，将内存空间还给 AI 处理器使用。

在一个示例性应用场景中，用户对一个人脸图片集进行了标注，要将该人脸图片集存储到存储设备中，以便后续利用该人脸图片集来进行 AI 训练。此时，用户可以通过对客户端触发操作，使得客户端将人脸图片集发送至存储设备。存储设备接收到人脸图片集后，存储设备的处理器通常会先将人脸图片集缓存至内存中，再将内存中的人脸图片集写入至硬盘，从而实现人脸图片集的持久化存储。由于人脸图片集的数据量庞大，存储设备的内存的容量可能不满足需求，导致存储设备的内存的可用容量达到预设阈值。此时，存储设备的处理器可以通过执行上述步骤 S1601 至步骤 S1604，借用 AI 内存的内存空间，通过更多的内存空间来缓存人脸图片集，进而加速人脸图片集的存储速度。

相关技术中，存储设备的内存的容量是固定的，这就导致存储业务数据时经常面临内存不足的问题。而通过这种可选方式，存储装置能够借用 AI 装置的 AI 内存来进行业务数据的读写，从而扩大了存储装置的可用内存空间，让存储装置能够在更大的内存上存储业务数据，从而节省业务数据的读写时间，提高业务数据的读写效率。

可选地，存储设备的处理器和 AI 处理器可以实现算力的协同。实现算力协同的方式可以有多种，以下通过图 17 实施例和图 18 实施例举例说明。

在一些实施例中，AI 处理器能够借用存储设备的处理器的算力来处理 AI 计算，从而支持存储设备的处理器和 AI 处理器算力协同的功能。参见图 17，本申请实施例提供了一种数据处理方法，该方法可以应用于存储设备，该存储设备可以是图 1 或图 2 所示的任意一个存储设备，包括以下步骤。

S1701，当确定 AI 处理器的算力不足时，AI 处理器将计算任务划分为至少两个子任务。

在 AI 处理器进行 AI 训练以及资源调度的过程中，AI 处理器由于执行计算操作，导致 CPU 被占用，由此出现算力不足的情况，此时，AI 处理器可以通过对计算任务进行划分，利用处理器分担计算任务。

S1702，AI 处理器将至少两个子任务中的第一子任务发送给存储设备的处理器。

第一子任务为该至少两个子任务中待由存储设备的处理器执行的任务。AI 处理器可以通过高速互联网络来发送第一子任务。

S1703，存储设备的处理器执行第一子任务。

S1704，存储设备的处理器将计算结果发送给 AI 处理器。

存储设备的处理器可以在完成第一子任务后，将第一子任务的计算结果反馈给 AI 处理器，AI 处理器可以接收计算结果，根据计算结果执行下一步计算任务，或根据计算结果来调度 AI 算力单元。

通过这种可选方式，AI 处理器能够在自身的算力不足时，借用存储设备的处理器的算力来处理 AI 计算，从而实现 AI 处理器和存储设备的处理器的算力协同，提高了 AI 处理器的算力，从而加快了 AI 训练的速度。

在一些实施例中，存储设备的处理器能够借用 AI 处理器的算力来处理业务数据的读写，从而加速访问业务数据的速度或存储业务数据的速度。参见图 18，本申请实施例提供了一种数据处理方法，该方法可以应用于存储设备，该存储设备可以是图 1 或图 2 所示的任意一个存储设备，包括以下步骤。

S1811，当确定存储设备的处理器的算力不足时，存储设备的处理器将计算任务划分为至少两个子任务。

在存储设备的处理器处理业务数据的过程中，处理器由于执行读操作或写操作，导致 CPU 不足，由此可以出现算力不足的情况。此时，存储设备的处理器可以通过对计算任务进行划分，利用 AI 处理器分担计算任务。

S1812，存储设备的处理器将至少两个子任务中的第二子任务发送给 AI 处理器。

第二子任务为该至少两个子任务中待由 AI 处理器执行的任务。例如，第二子任务可以从硬盘中获取业务数据。处理器可以通过高速互联网络来发送第二子任务。

S1813，AI 处理器执行第二子任务。

S1814，AI 处理器将计算结果发送给存储设备的处理器。

AI 处理器可以在完成第二子任务后，将第二子任务的计算结果反馈给存储设备的处理器，存储设备的处理器可以接收计算结果，根据计算结果执行下一步计算任务。

例如，如果第二子任务是从硬盘中获取业务数据，AI 处理器从硬盘中获取业务数据后，可以将得到的业务数据发送给存储设备的处理器。

通过这种可选方式，存储设备的处理器能够在自身的算力不足时，借用 AI 处理器的算力来处理业务数据的读写，从而实现 AI 处理器和存储设备的处理器的算力协同，提高了存储设备的处理器的算力，从而加快了业务数据的读写速度。

在一个可选的实施例中，存储设备处理 AI 计算之前，存储设备的处理器可以将操作系统加载到 AI 装置，存储设备的处理器通过操作系统启动 AI 装置。在另一个可选的实施例中，AI 装置也可以预存有操作系统，存储设备的处理器可以向 AI 装置发送启动指令，AI 装置可以接收到启动指令，响应于启动指令，加载操作系统，运行操作系统，从而得以启动。其中，处理器可以从终端或上层应用接收到 AI 计算指令，AI 计算指令用于指示处理器进行 AI 计算，处理器可以在 AI 计算指令的触发下，执行将操作系统加载到 AI 装置的步骤。当然，处理器也可以在其他时机下执行将操作系统加载到 AI 装置的步骤，本实施例对此不做限定。

上述实施例描述了单个存储设备处理 AI 计算的方法流程，在本申请的一些实施例中，可以通过分布式存储系统中的多个存储设备协同进行 AI 计算，从而提高整体的 AI 算力。以下进行具体描述。

图 19 是本申请实施例提供的一种数据处理方法的流程图，该方法可以应用于图 1 所示的存储系统，该方法可以应用于图 1 所示的系统 1，该方法的交互主体包括存储系统中的第一存储设备以及第二存储设备，包括以下步骤。

S1901，第一存储设备的第一 AI 装置和第二存储设备的第二 AI 装置之间通过第一网络传输 AI 参数。

在第一存储设备和第二存储设备处理 AI 计算的过程中，第一存储设备可以和第二存储设

备交换 AI 参数。具体地，AI 计算通常基于神经网络模型进行，而神经网络模型的计算过程主要分为前向传播算法以及后向传播（back propagation neural networks, BP）算法两大部分。前向传播算法，用于计算神经网络模型的输出结果。在运行前向传播算法的过程中，数据会从神经网络模型的输入层开始，按照输入层、一个或多个隐藏层直至输出层的方向，经过每个层的计算，逐层传递，直至从输出层输出。后向传播算法，用于减小模型的输出结果与实际结果之前的误差。在运行后向传播算法的过程中，会通过调整 AI 参数，例如各个神经元的权重，来优化神经网络模型。总结来说，神经网络模型的运行包括前向传播算法与后向传播算法的循环迭代，在此过程中，需要通过交换 AI 参数。来使模型通过 AI 参数的调整不断优化。

当第一 AI 装置计算出 AI 参数，第一 AI 装置可以通过第一网络接口输出 AI 参数，第一网络可以从第一 AI 装置的第一网络接口接收 AI 参数，第一网络可以将接收的 AI 参数发送至第二 AI 装置的第一网络接口，该第二 AI 装置可以从第一网络接口接收输入的 AI 参数。

可选地，如果存储系统还包括用于处理 AI 计算的主机，第一网络可以用于在第一存储设备、第二存储设备和主机之间传输 AI 参数。

S1902, 第一存储设备的第一存储装置和第二存储设备的第二存储装置之间通过第二网络传输业务数据。

示例性地，参见图 20，图 20 是本申请实施例提供的一种数据传输的示意图，从图 20 可以看出，AI 参数可以通过第一网络在不同存储设备的 AI 装置之间传输，业务数据可以通过第二网络在不同存储设备的存储装置之间传输。

本实施例提供的方法，两个存储设备通过各自的 AI 装置，利用第一网络来交换 AI 参数，通过各自的存储装置，利用第二网络来交换业务数据，从而根据 AI 参数以及业务数据来协同进行 AI 计算，通过融合多个存储设备的存储能力和 AI 算力，能够提高系统整体的存储能力和 AI 算力。

可选地，存储装置可以借用第一网络来交换业务数据。具体地，参见图 21，本申请实施例提供了一种业务数据传输方法，该方法可以应用于图 1 所示的系统 1，该方法的交互主体包括第一存储装置和第二存储装置，包括以下步骤。

S2111，第一存储装置确定第二网络的网络资源低于设定的存储网络资源阈值。

第一存储装置或管理装置可以检测第一存储装置的第二网络的网络资源，判断第一存储装置的第二网络的网络资源是否低于设定的存储网络资源阈值，如果确定第二网络的网络资源低于设定的存储网络资源阈值时，可以触发执行步骤 S2112。在一些实施例中，可以还检测第一 AI 装置的第一网络的网络资源以及第二 AI 装置的第一网络的网络资源，判断第一 AI 装置的第一网络的网络资源以及第二 AI 装置的第一网络的网络资源是否高于设定的 AI 网络资源阈值，如果确定第二网络的网络资源低于设定的存储网络资源阈值，且第一 AI 装置的第一网络的网络资源高于设定的 AI 网络资源阈值，且第二 AI 装置的第一网络的网络资源高于设定的 AI 网络资源阈值时，表明当前的第二网络的网络资源不足而 AI 网络资源充足，可以在这种情况的触发下，执行步骤 S2112。另外，可以理解的是，即使没有 S2111 的限定条件，存储装置也可以借用第一网络来交换业务数据。

S2112，第一存储装置和第二存储装置之间通过第一网络传输其他业务数据。

以第一存储装置为源设备，第二存储装置为目的设备为例，其他业务数据的传输流程可以包括：第一 AI 装置的第一网络接口可以通过高速互连网络，访问第一存储装置的硬盘，从第一存储装置的硬盘读取其他业务数据，通过第一网络，将其他业务数据发送给第二存储设备，第二 AI 装置的第一网络接口可以通过第一网络，接收其他业务数据，通过高速互连网络，将其他业务数据写入至第二存储装置的内存。

在一个示例性场景中，当分布式存储系统启动作业，各个存储设备首次加载业务数据时，由于存储设备通过第二网络来加载业务数据，导致第二网络出现拥塞，而此时由于 AI 装置尚未启动作业或者刚开始启动作业，第一网络通常是空闲的，因此能够利用空闲的第一网络来传输业务数据，从而加快业务数据的传输，使得业务数据能够通过第一网络和第二网络的协同快速加载到各个存储设备。

相关技术中，业务数据通常只会通过第二网络传输，相当于仅有第二网络这一条转发路径。通过这种可选方式，达到的效果至少可以包括：为存储装置之间交换业务数据提供了一条新的通路，通过在第二网络的网络资源不足时，使用第一网络来交换业务数据，第一网络可以作为业务数据的一条新增的转发路径，那么业务数据既可以通过第二网络传输，也可以通过第一网络传输，因此提高了传输业务数据的网络带宽，从而缩短了交换业务数据的时延，提高了交换业务数据的速度，进而提高了进行 AI 计算的速度。

可选地，不同存储设备之间可以通过 RDMA 实现内存直通。具体地，参见图 22，本申请实施例提供了一种数据处理方法，该方法可以应用于图 1 所示的系统 1，该方法的交互主体包括第一存储设备中的第一 AI 装置和第二存储设备中的第二 AI 装置，该方法包括以下步骤。

S2211，第一 AI 装置向第二 AI 处理器发送第一网络的网络资源请求，并且向第二处理器发送内存 RDMA 访问请求。

其中，第一网络的网络资源请求用于请求占用第二 AI 装置的部分第一网络的网络资源来传输业务数据。内存 RDMA 访问请求用于以 RDMA 的方式访问第二内存，内存 RDMA 访问请求可以包括第二内存的目的地址，以指明要将业务数据写入到该目的地址。

在一个可选实施例中，第一 AI 装置的第一 AI 处理器可以生成第一网络的网络资源请求和内存 RDMA 访问请求，将第一网络的网络资源请求和内存 RDMA 访问请求发送给第一网络接口，第一网络接口可以将第一网络的网络资源请求发送至第二 AI 处理器，将内存 RDMA 访问请求发送至第二处理器。第二 AI 处理器接收到第一网络的网络资源请求后，可以划分部分第一网络的网络资源，预留给第一 AI 装置使用。第二 AI 处理器接收到内存 RDMA 访问请求后，第二 AI 处理器可以生成内存 RDMA 访问响应，向第一 AI 装置返回内存 RDMA 访问响应，该内存 RDMA 访问响应携带第二内存的目标地址，以便第一 AI 装置以 RDMA 的方式，将数据写到第二内存的目标地址。通过上述流程，可以建立第二 AI 装置与第一 AI 装置之间的 RDMA 通路。

S2212，第一 AI 装置从第一内存中读取其他业务数据。

第一 AI 装置可以包括第一网络接口，该第一网络接口支持 RDMA 功能，第一 AI 装置的第一网络接口可以通过高速互连网络，访问第一内存，根据其他业务数据在第一内存的地址，从第一内存读取其他业务数据。

S2213, 第一 AI 装置通过第一网络将其他业务数据发送给第二 AI 装置。

第一 AI 装置的第一网络接口可以将读取的其他业务数据发送出去, 其他业务数据可以通过第一网络传输至第二 AI 装置。

S2214, 第二 AI 装置将其他业务数据写入第二内存。

具体地, 第二 AI 装置可以包括第一网络接口, 该第一网络接口支持 RDMA 功能, 第二 AI 装置的第一网络接口可以接收到来自第一 AI 装置的其他业务数据, 第二 AI 装置的第一网络接口可以通过高速互联网络, 向第二内存的目的地址写入其他业务数据。

此外, 第二 AI 装置对其他业务数据完成写操作后, 第一 AI 装置与第二 AI 装置可以释放占用的第一网络的网络资源, 结束第一 AI 装置与第二 AI 装置之间的数据交换任务。

参见图 23, 图 23 示出了通过第一网络实现不同内存之间数据直通的示意图, 从图 23 可以看出, 第一网络能够作为一条新的业务数据的转发路径, 第一内存缓存的业务数据能够通过第一网络直接传输至第二内存。

通过这种可选方式, 第一内存与第二内存之间能够通过 RDMA 的方式进行业务数据交换, 如此, 能够免去第一处理器与第二处理器的处理开销, 让业务数据从第一内存直接到达第二内存, 从而加快业务数据交换的速度, 提高业务数据交换的效率。

图 23 实施例示出了两个存储设备的内存之间通过第一网络实现 RDMA 数据直通的流程, 本申请的一些实施例中, 还可以采用同理的方式, 在源存储设备的硬盘与目的存储设备的内存之间实现 RDMA 数据直通, 以下通过图 24 实施例具体描述:

参见图 24, 本申请实施例提供了一种数据处理方法, 该方法可以应用于图 1 所示的系统 1, 该方法的交互主体包括第一存储设备中的第一 AI 装置和第二存储设备中的第二 AI 装置, 该方法包括以下步骤。

S2411, 第一 AI 装置向第二 AI 处理器发送第一网络的网络资源请求, 并且向第二处理器发送内存 RDMA 访问请求。

S2412, 第一 AI 装置从第一存储装置的硬盘中读取其他业务数据。

第一 AI 装置的第一网络接口可以通过高速互联网络, 访问第一存储装置的硬盘, 根据其业务数据在第一存储装置的硬盘的地址, 从第一存储装置的硬盘读取其他业务数据。

S2413, 第一 AI 装置通过第一网络将其他业务数据发送给第二 AI 装置。

S2414, 第二 AI 装置将其他业务数据写入第二内存。

参见图 25, 图 25 示出了通过第一网络实现硬盘与存储设备的内存之间数据直通的示意图, 从图 25 可以看出, 第一网络能够作为一条新的业务数据的转发路径, 第一存储装置的硬盘存储的业务数据能够通过第一网络, 以 RDMA 的方式直接传输至第二内存, 从而实现了源存储节点的硬盘和目标存储节点的内存之间的数据直通。

通过这种可选方式, 第一存储装置的硬盘与第二内存之间能够通过 RDMA 的方式进行业务数据交换, 如此, 能够免去第一处理器与第二处理器的处理开销, 让业务数据从第一存储装置的硬盘直接到达第二内存, 从而加快业务数据交换的速度, 提高业务数据交换的效率。

可选地, AI 装置可以借用第二网络的网络资源来交换 AI 参数。具体地, 第一 AI 装置和第二 AI 装置之间通过第二网络传输其他 AI 参数。

这一功能可以在多种情况下触发,例如,当第一网络的网络资源低于设定的 AI 网络资源阈值时,第一 AI 装置和第二 AI 装置之间通过第二网络传输其他 AI 参数。另外,可以理解的是,即使没有网络资源低于设定的 AI 网络资源阈值的限定条件,第一 AI 装置和第二 AI 装置之间也可以通过第二网络传输其他 AI 参数。

在一个可选实施例中,第一 AI 装置、第二 AI 装置或管理装置可以检测第一网络的网络资源的大小,判断第一网络的网络资源是否低于设定的 AI 网络资源阈值,如果第一网络的网络资源低于设定的 AI 网络资源阈值,则通过第二网络传输其他 AI 参数。其中,第一 AI 装置、第二 AI 装置或管理装置还可以检测第二网络的大小,判断第二网络是否高于设定的存储网络资源阈值,如果第一网络的网络资源低于设定的 AI 网络资源阈值且第二网络高于设定的存储网络资源阈值,则通过第二网络传输其他 AI 参数。

在一个示例性应用场景中,如果第一 AI 装置要加载的业务数据全部或大部分位于本地的第一存储装置时,第一 AI 装置能够就近在本地加载业务数据,从而免去了大量通过第二网络来访问远端存储设备的操作,从而使得本地的第二网络的网络资源充足。而由于频繁的与远端 AI 装置进行 AI 参数交换,导致第一网络的网络资源不足,在这种场景中,可以利用相对充足的第二网络的网络资源作为 AI 参数的新通路,加速 AI 参数的交换速度。

通过这种可选方式,为 AI 装置之间交换 AI 参数提供了一条新的通路,通过在第一网络的网络资源不足时,使用第二网络来交换 AI 参数,提高了传输 AI 参数的网络带宽,从而缩短了交换 AI 参数的时延,提高了交换 AI 参数的速度,进而提高了进行 AI 计算的速度。

可选地,不同存储设备之间可以通过 RDMA 实现 AI 内存直通。具体地,参见图 26,本申请实施例提供了一种数据处理方法,该方法可以应用于图 1 所示的系统 1,该方法的交互主体包括第一存储设备以及第二存储设备,包括以下步骤。

S2601,第一 AI 处理器向第二处理器发送第二网络的网络资源请求。

第一 AI 处理器可以生成第二网络的网络资源请求,发送给第一网络接口,第一网络接口可以将第二网络的网络资源请求发送给第二处理器。其中,第二网络的网络资源请求用于请求占用第二 AI 装置的部分第二网络的网络资源来传输其他 AI 参数,第二网络请求可以携带其他 AI 参数的目的地址。

其中,第二网络的网络资源请求可以是内存 RDMA 访问请求,第二网络的网络资源请求用于请求以 RDMA 的方式访问第二 AI 内存,第二网络请求携带的目的地址可以是第二 AI 内存中的地址。第二处理器接收到第二网络请求后,可以建立第一 AI 内存与第二 AI 内存之间的 RDMA 通路。

S2602,第一 AI 处理器从第一 AI 内存中获取其他 AI 参数。

S2603,第一 AI 处理器将其他 AI 参数以 RDMA 方式通过第二网络传输至第二 AI 内存。

第一 AI 处理器可以访问本地的第一 AI 内存,得到其他 AI 参数,将其他 AI 参数发送给第一 AI 装置的第一网络接口,第一 AI 装置的第一网络接口可以将其他 AI 参数发送给第二 AI 装置的第一网络接口,第二 AI 装置的第一网络接口可以接收其他 AI 参数,将其他 AI 参数写入到第二 AI 内存。在此过程中,可以通过 RDMA 的方式,免去第二 AI 装置的第二 AI 处理器的参与,从而避免占用第二 AI 处理器的处理资源,提高转发性能。

参见图 27,图 27 示出了通过第二网络实现两个存储设备的 AI 内存之间数据直通的示意

图, 从图 27 可以看出, 第二网络能够作为一条新的 AI 参数的转发路径, 第一存储设备的第一 AI 内存的 AI 参数能够通过第一网络, 以 RDMA 的方式直接传输至第二 AI 内存, 从而实现了源存储设备和目标存储设备的 AI 内存之间的数据直通。

通过这种可选方式, 第一 AI 内存与第二 AI 内存之间能够通过 RDMA 的方式进行 AI 参数交换, 如此, 能够免去第一 AI 处理器与第二 AI 处理器的处理开销, 让 AI 参数从第一 AI 内存直接到达第二 AI 内存, 从而加快 AI 参数交换的速度, 提高 AI 参数交换的效率。

可选地, 如果 AI 装置借用存储设备的内存来进行训练, 存储设备的内存可以缓存有 AI 参数, 不同存储设备之间可以通过 RDMA 实现内存直通, 利用第二网络来交换不同存储设备的内存缓存的 AI 参数。具体地, 参见图 28, 本申请实施例提供了一种业务数据传输方法, 该方法的交互主体包括第一存储设备以及第二存储设备, 包括以下步骤。

S2801, 第一 AI 处理器向第二处理器发送第二网络的网络资源请求。

S2802, 第一 AI 处理器从第一内存中获取其他 AI 参数。

第一 AI 处理器可以通过高速互连网络, 访问第一内存, 得到第一内存缓存的其他 AI 参数。

S2803, 第一 AI 处理器将其他 AI 参数以 RDMA 方式通过第二网络传输至第二内存。

参见图 29, 图 29 示出了通过第二网络实现两个存储设备的内存之间 AI 参数直通的示意图, 从图 29 可以看出, 第二网络能够作为一条新的 AI 参数的转发路径, 第一存储设备的第一内存的 AI 参数能够通过第一网络, 以 RDMA 的方式直接传输至第二内存, 从而实现了源存储节点和目标存储节点的内存之间的 AI 参数直通。

通过这种可选方式, 第一内存与第二内存之间能够通过 RDMA 的方式进行 AI 参数交换, 如此, 能够免去第一 AI 处理器与第二 AI 处理器的处理开销, 让 AI 参数从第一内存直接到达第二内存, 从而加快 AI 参数交换的速度, 提高 AI 参数交换的效率。

可选地, 可以通过管理装置来调度存储系统中数据集所在的存储设备进行 AI 计算, 具体参见下述方法实施例。

图 30 是本申请实施例提供的一种数据处理方法的流程图, 该方法可以应用于图 1 所示的系统 1, 该方法的交互主体包括管理装置以及第一存储设备, 该方法包括以下步骤:

S3001, 管理装置接收第一作业请求。

第一作业请求用于请求基于数据集进行训练, 数据集包括业务数据。第一作业请求可以包括数据集的标识, 该数据集的标识用于指示数据集, 例如可以是数据集的 ID、名称等。在一些实施例中, 客户端可以生成第一作业请求, 向管理装置发送第一作业请求, 管理装置可以接收来自客户端的第一作业请求。其中, 第一作业请求可以由用户的输入操作触发, 例如, 客户端可以显示图形用户界面, 用户可以在图形用户界面上输入数据集的标识, 客户端可以接收输入的数据集的标识, 生成第一作业请求。

S3002, 管理装置根据第一作业请求确定待训练的数据集的分布。

管理装置可以解析第一作业请求, 得到第一作业请求携带的数据集的标识, 根据数据集的标识, 查询数据集的标识与数据集的元数据之间的映射关系, 得到数据集的元数据, 该数据集的元数据用于指示数据集的地址, 管理装置可以根据数据集的元数据, 确定数据集的分

布。例如，管理装置可以查询数据集中每个业务数据所在的存储设备，生成存储设备列表，该存储设备列表包括设备标识，该设备标识用于标识目标存储设备，目标存储设备存储有数据集中的部分业务数据或全部业务数据。可选地，存储设备列表中的设备标识的数量可以有多个，多个设备标识的排列顺序用于表示数据集在对应目标存储设备上存储的数据量的大小。例如，存储设备列表中第一个设备标识对应的目标设备存储的数据集的业务数据最多，存储设备列表中最后一个设备标识对应的目标设备存储的数据集的业务数据最少。

S3003，当确定业务数据分布在第一存储设备时，管理装置向第一存储设备发送第一计算请求。

第一计算请求用于请求第一 AI 装置对业务数据进行 AI 计算，第一计算请求可以包括数据集的标识，通过携带数据集的标识来指明要获取数据集中的业务数据。示例性地，如果分布式存储系统中包括 N 个存储设备，分别是存储设备 1、存储设备 2 至存储设备 N，其中数据集分布在存储设备 1、存储设备 2 和存储设备 3，管理装置可以选择在存储设备 1、存储设备 2 和存储设备 3 上进行 AI 计算，则向存储设备 1、存储设备 2 和存储设备 3 的 AI 装置发送第一计算请求。其中 N 为正整数。

S3004，第一 AI 装置基于第一计算请求，从第一存储设备的第一存储装置获取业务数据。

第一 AI 装置可以解析第一计算请求，得到第一计算请求携带的数据集的标识，根据数据集的标识，确定数据集在第一存储装置的地址，通过高速互联网络访问第一存储装置，得到第一存储装置存储的数据集的全部业务数据或部分业务数据。

S3005，第一 AI 装置对业务数据进行 AI 计算，获得第一计算结果。

通过这种可选方式，由于管理装置选择了业务数据所在的第一存储设备来提供 AI 计算，而第一存储设备通过内部的第一存储装置，即可得到业务数据来进行 AI 计算，从而避免业务数据的跨存储设备移动，通过免去了访问其他存储设备以得到业务数据带来的时延，从而减少了获取业务数据的时延，也就提高了进行 AI 计算的速度。

可选地，管理装置可以确定第一存储设备的运行状态是否满足设定条件，如果管理装置确定第一存储设备的运行状态满足设定条件，则管理装置向第一存储设备发送第一计算请求。

运行状态可以包括存储设备的内存的空闲空间、AI 内存的空闲空间、AI 算力单元的占用状态、AI 处理器的占用率、存储设备的处理器的占用率、网络资源的使用率中的一项或多项。预设条件包括以下条件 (1) 至 (5) 中的任意一种及其组合：

- (1) 存储设备的内存的空闲空间是否高于设定的空间阈值。
- (2) AI 内存的空闲空间是否高于设定的空间阈值。
- (3) AI 处理器的占用率是否低于设定的占用率阈值。
- (4) 存储设备的处理器的占用率是否低于设定的占用率阈值。
- (5) 网络资源的使用率是否低于设定的占用率阈值。

通过这种可选方式，能够保证选中的第一存储设备当下未被占用，能够提供 AI 算力，避免选中已被占用的存储设备进行 AI 计算导致设备开销过大、AI 计算任务无法及时完成的问题。

可选地，如果数据集所在的存储设备已被占用，可以通过管理装置来调度存储系统中与

数据集距离近的存储设备进行 AI 计算。具体地，参见图 31，本申请实施例提供了一种数据处理方法的流程图，该方法可以应用于图 1 所示的系统 1，该方法的交互主体包括管理装置、第一存储设备以及第二存储设备，该方法包括以下步骤。

S3101，管理装置接收第二作业请求。

步骤 S3101 与步骤 S3001 同理，在此不做赘述。

S3102，管理装置根据第二作业请求确定待训练的数据集的分布。

步骤 S3101 与步骤 S3002 同理，在此不做赘述。

S3103，当其他业务数据分布在多个存储设备中的第二存储设备时，管理装置进一步确定第二存储设备的运行状态是否满足设定条件。

S3104，当第二存储设备的运行状态不满足设定条件时，管理装置向第一存储设备发送第二计算请求。其中，第一存储设备与第二存储设备之间的距离低于设定的距离阈值。

在一些实施例中，对于分布式存储系统中每个存储设备，管理装置可以确定该存储设备的运行状态以及该存储设备与第二存储设备之间的距离，管理装置可以根据每个存储设备的运行状态以及与第二存储设备之间的距离，从分布式存储系统中确定运行状态满足设定条件，且与第二存储设备之间的距离低于设定的距离阈值的存储设备，得到第一存储设备。

作为一种可能的实现，管理装置可以采用基于代价的优化方式(Cost-Based Optimization, 简称: CBO)确定出第一存储设备。示例性地，对于分布式存储系统中每个存储设备，管理装置可以根据存储设备的运行状态、存储设备与第二存储设备之间的距离、存储设备存储的第二数据集的数据量，计算存储设备的代价值，该代价值表示启用该存储设备进行 AI 计算的开销大小。管理装置可以每个存储设备的代价值，选择代价值满足设定条件的存储设备作为第一存储设备，通过这种算法，由于离数据集最近的存储设备的代价值小，可以选中离数据集最近的存储设备来启用 AI 计算。作为示例，可以为本地存储和远端存储分别分配权重 $x1$ 和权重 $y1$ ， $x1$ 小于 $y1$ ，对于任一个存储设备而言，如果第二数据集在该存储设备本地存储的数据量为 x ，第二数据集不在本存储设备的数据量为 y ，则根据权重 $x1$ 和权重 $y1$ ，对 x 和 y 进行加权求和，将得到的加权和值作为该存储设备的代价值，对每个存储设备按照代价值从小到大的顺序排序，如果 AI 计算需要 n 个存储设备，则从排序结果中选择排在前 n 位的存储设备，作为第一存储设备。示意性地，可以设定从本地加载数据的权重是 1，从远端加载数据的权重是 10，如果某个存储设备中存储有 30% 的数据集，数据集剩余的 70% 不在该存储设备，则该存储设备的代价值为 $30\%*1+70\%*10=7.3$ 。

可选地，管理装置选中第一存储设备后，可以向客户端输出推荐列表，该推荐列表表示管理装置推荐启动进行 AI 计算的存储设备，该推荐列表可以包括第一存储设备的标识。

S3105，第一 AI 装置基于第二计算请求，从第二存储设备获取其他业务数据。

S3106，第一 AI 装置对其他业务数据进行 AI 计算，获得第二计算结果。

通过这种可选方式，如果数据集所在的存储设备已被占用，管理装置能够选择与数据集距离近的存储装置来提供 AI 计算，从而缩短数据集的传输距离，减少业务数据的跨节点移动。

以下，对本申请实施例的应用场景进行示例性介绍。

随着深度学习的应用，经常需要使用包含大量数据的数据集来进行模型训练，从而拟合出神经网络。例如，参见图 32，图 32 示出了模型训练的逻辑流程，模型训练可以包括模型

加载阶段、数据加载阶段、参数初始化阶段、前向传播阶段、损失计算阶段、反向传播阶段、参数更新阶段、权重保存阶段。其中，由于需要大量的迭代训练才能得到模型参数，而迭代次数可以达到从几百到上万不等，因此要加载大量的数据、进行参数更新和交换，导致模型训练非常耗时。由此可见，充分利用软件资源和硬件资源，提高调度性能，优化数据传输路径对于模型训练十分重要。

应用本申请实施例提供的方法，可以对模型训练的各个阶段进行优化。

具体来讲，在模型加载阶段，可以应用图 30 和图 31 实施例提供的方法，来提供近数据的 AI 训练。具体地，可以根据训练模型所需的数据的地址，选择近数据的 AI 装置，作为训练节点来进行启动。

在数据加载阶段，可以应用图 7 实施例、图 8 实施例、图 9 实施例、图 12 实施例、图 14 实施例提供的方法，缩短业务数据的传输路径，基于近存储的方式来执行数据读写操作。此外，还可以利用内存与 AI 内存之间的高速互联，实现内存与 AI 内存之间的 RDMA 直通。

在前向传播阶段，可以应用图 18 实施例提供的方法，基于高速互连网络，来实现处理器与 AI 处理器的算力协同，从而加速 AI 算子的计算速度。

在参数更新阶段，可以应用图 26 实施例提供的方法、图 28 实施例提供的方法，实现不同存储设备的 AI 内存之间的 RDMA 通路，以及不同存储设备的内存之间的 RDMA 通路，通过 RDMA 通路传输 AI 参数，从而加速参数交换。

参见图 33，图 33 是本申请实施例提供的一种 AI 训练的逻辑架构图，用户可以在客户端上执行操作，输入 AI 训练作业，客户端会生成指示执行 AI 训练作业的请求，调用作业提交应用程序编程接口(Application Programming Interface, API)，发给 AI 作业管理服务；AI 作业管理服务接收到请求，会对请求携带的作业进行解析，将资源请求发送给管理装置。管理装置接收到资源请求后，管理装置中的调度器从所管理的底层物理资源池中，选择合适的存储设备，在选择的存储设备的 AI 装置上启动，对应的 AI 训练作业，选择的存储设备的 AI 装置的算力被 AI 训练作业占用，直至 AI 训练作业结束后，会释放该计算资源。

上述所有可选技术方案，可以采用任意结合形成本申请的可选实施例，在此不再一一赘述。

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。计算机程序产品包括一个或多个计算机程序指令。在计算机上加载和执行计算机程序指令时，全部或部分地产生按照本申请实施例的流程或功能。计算机可以是通用计算机、专用计算机、计算机网络、或者其他可编程装置。计算机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，计算机程序指令可以从一个网站站点、计算机、服务器或业务数据中心通过有线或无线方式向另一个网站站点、计算机、服务器或业务数据中心进行传输。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包含一个或多个可用介质集成的服务器、业务数据中心等业务数据存储设备。可用介质可以是磁性介质(例如软盘、硬盘、磁带)、光介质(例如，数字视频光盘(digital video disc, DVD)、或者半导体介质(例如固态硬盘)等。

本领域普通技术人员可以意识到，结合本文中所公开的实施例中描述的各方法步骤和单元，能够以电子硬件、计算机软件或者二者的结合来实现，为了清楚地说明硬件和软件的可互换性，在上述说明中已经按照功能一般性地描述了各实施例的步骤及组成。这些功能究竟以硬件还是软件方式来执行，取决于技术方案的特定应用和设计约束条件。本领域普通技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出本申请的范围。

所属领域的技术人员可以清楚地了解到，为了描述的方便和简洁，上述描述的系统、装置和单元的具体工作过程，可以参见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，该单元的划分，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。另外，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，也可以是电的，机械的或其它的形式连接。

该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申请实施例方案的目的。

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，也可以是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。

该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，可以存储在一个计算机可读存储介质中。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中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（read-only memory，ROM）、随机存取存储器（random access memory，RAM）、磁碟或者光盘等各种可以存储程序代码的介质。

以上，仅为本申请的具体实施方式，但本申请的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本申请揭露的技术范围内，可轻易想到各种等效的修改或替换，这些修改或替换都应涵盖在本申请的保护范围之内。因此，本申请的保护范围应以权利要求的保护范围为准。

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。计算机程序产品包括一个或多个计算机程序指令。在计算机上加载和执行计算机程序指令时，全部或部分地产生按照本申请实施例的流程或功能。计算机可以是通用计算机、专用计算机、计算机网络、或者其他可编程装置。计算机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，计算机程序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。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包含一个或

多个可用介质集成的服务器、数据中心等数据存储设备。可用介质可以是磁性介质 (例如软盘、硬盘、磁带)、光介质(例如, 数字视频光盘 (digital video disc, DVD)、或者半导体介质 (例如固态硬盘) 等。

本领域普通技术人员可以理解实现上述实施例的全部或部分步骤可以通过硬件来完成, 也可以通过程序来指令相关的硬件完成, 的程序可以存储于一种计算机可读存储介质中, 上述提到的存储介质可以是只读存储器, 磁盘或光盘等。

以上所述仅为本申请的可选实施例, 并不用以限制本申请, 凡在本申请的精神和原则之内, 所作的任何修改、等同替换、改进等, 均应包含在本申请的保护范围之内。