



(12) 发明专利

(10) 授权公告号 CN 114064594 B

(45) 授权公告日 2023. 09. 22

(21) 申请号 202111388428.4

CN 111913665 A, 2020.11.10

(22) 申请日 2021.11.22

CN 112799740 A, 2021.05.14

(65) 同一申请的已公布的文献号

CN 112965761 A, 2021.06.15

申请公布号 CN 114064594 A

US 2017124345 A1, 2017.05.04

(43) 申请公布日 2022.02.18

US 2018157752 A1, 2018.06.07

(73) 专利权人 马上消费金融股份有限公司

US 2021034398 A1, 2021.02.04

地址 401120 重庆市渝北区黄山大道中段

US 2021294778 A1, 2021.09.23

52号渝兴广场B2栋4至8楼

WO 2019015288 A1, 2019.01.24

(72) 发明人 张伟 吴海英 权圣 蒋宁

WO 2020168692 A1, 2020.08.27

王洪斌 李云彬 韩卫强

CN 113515346 A, 2021.10.19

(74) 专利代理机构 北京国昊天诚知识产权代理

CN 111552508 A, 2020.08.18

有限公司 11315

CN 111782339 A, 2020.10.16

专利代理师 徐晨影

CN 112379828 A, 2021.02.19

(51) Int. Cl.

CN 112905537 A, 2021.06.04

G06F 16/176 (2019.01)

CN 113296792 A, 2021.08.24

G06F 9/54 (2006.01)

CN 113342280 A, 2021.09.03

US 2013227563 A1, 2013.08.29

(56) 对比文件

钱建梅等. 风云气象卫星数据存档与服务系统. 应用气象学报. 2012, (第03期), 第115-122页.

CN 106209741 A, 2016.12.07

张永夏. 管理Docker容器的数据卷. 网络安全和信息化. 2017, (第10期), 第85-87页.

CN 108205623 A, 2018.06.26

CN 109165206 A, 2019.01.08

CN 109274722 A, 2019.01.25

CN 109739619 A, 2019.05.10

审查员 田志方

权利要求书2页 说明书11页 附图3页

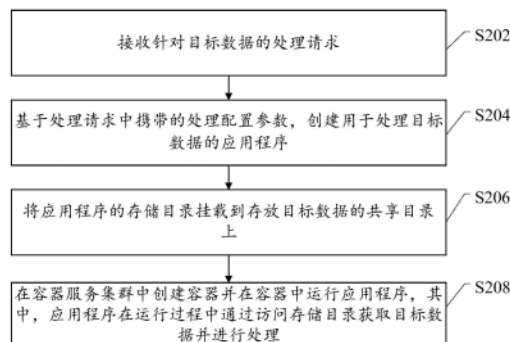
(54) 发明名称

数据处理方法和装置

并在所述容器中运行所述应用程序,其中,所述应用程序在运行过程中通过访问所述存储目录获取所述目标数据并进行处理。

(57) 摘要

本说明书实施例公开了一种数据处理方法和装置,以解决传统的数据处理方案存在的处理效率低、处理成本高的问题。所述方法包括:接收针对目标数据进行处理所需的处理配置参数;基于所述处理配置参数,创建用于处理所述目标数据的应用程序;将所述应用程序的存储目录挂载到存放目标数据的共享目录上,所述存储目录被挂载到所述共享目录后,所述应用程序在运行过程中对所述存储目录的访问操作被映射到所述共享目录上;在容器服务集群中创建容器



CN 114064594 B

1. 一种数据处理方法,其特征在于,包括:

接收针对目标数据的处理请求,所述处理请求中携带对所述目标数据进行处理所需的处理配置参数;

基于所述处理配置参数,创建用于处理所述目标数据的应用程序;

将所述应用程序的存储目录挂载到存放所述目标数据的共享目录上,所述应用程序在运行过程中对所述存储目录的访问操作被映射到所述共享目录上,所述共享目录为在云存储平台中为所述目标数据分配的、被至少两个使用方共享的目录,所述应用程序包括所述至少两个使用方分别对应的应用程序;

在容器服务集群中创建容器并在所述容器中运行所述应用程序,其中,所述应用程序在运行过程中通过访问所述存储目录获取所述目标数据并进行处理。

2. 根据权利要求1所述的方法,其特征在于,所述将所述应用程序的存储目录挂载到存放所述目标数据的共享目录上,包括:

获取与存放所述目标数据的共享目录绑定的存储申明;

将所述应用程序的存储目录挂载到所述存储申明上。

3. 根据权利要求2所述的方法,其特征在于,在接收针对目标数据的处理请求之前,所述方法还包括:

接收针对所述目标数据的共享请求,所述共享请求用于请求在不同使用方之间共享所述目标数据;

在云存储平台中为所述目标数据分配相应的共享目录,并将所述目标数据存储至所述共享目录下;

创建存储申明,并将所创建的存储申明与所述共享目录进行绑定。

4. 根据权利要求3所述的方法,其特征在于,在将所述目标数据存储至所述共享目录下之后,所述方法还包括:

生成所述目标数据与所述共享目录之间的对应关系信息;

向所述共享请求的发起方返回所述对应关系信息。

5. 根据权利要求1所述的方法,其特征在于,在所述在容器服务集群中创建容器并在所述容器中运行所述应用程序之后,所述方法还包括:

从所述应用程序的存储目录中加载对所述目标数据进行处理所得的处理结果;

将所述处理结果存储至所述共享目录下。

6. 根据权利要求5所述的方法,其特征在于,在将所述处理结果存储至所述共享目录下之前,所述方法还包括:

基于所述处理结果,确定所述目标数据是否被修改;

所述将所述处理结果存储至所述共享目录下,包括:

如果所述目标数据被修改,则将修改后的所述目标数据覆盖存储至所述共享目录下。

7. 根据权利要求1至6中任一项所述的方法,其特征在于,所述目标数据包括语料数据和/或机器学习模型,所述处理配置参数为用于利用所述目标数据进行模型训练的训练脚本。

8. 一种数据处理系统,其特征在于,包括:管理控制平台和容器服务集群;

所述管理控制平台,用于接收针对目标数据的处理请求,所述处理请求中携带对所述

目标数据进行处理所需的处理配置参数,基于所述处理配置参数,创建用于处理所述目标数据的应用程序,将所述应用程序的存储目录挂载到存放所述目标数据的共享目录上,所述存储目录被挂载到所述共享目录后,所述应用程序在运行过程中对所述存储目录的访问操作被映射到所述共享目录上,以及向所述容器服务集群发送容器创建请求,所述容器创建请求用于请求创建运行所述应用程序的容器,所述共享目录为在云存储平台中为所述目标数据分配的、被至少两个使用方共享的目录,所述应用程序包括所述至少两个使用方分别对应的应用程序;

所述容器服务集群,用于接收来自所述管理控制平台的容器创建请求,创建容器并在所述容器中运行所述应用程序,其中,所述应用程序在运行过程中通过访问所述存储目录获取所述目标数据并进行处理。

9. 根据权利要求8所述的系统,其特征在于,所述系统还包括云存储平台;

所述管理控制平台,还用于接收针对所述目标数据的共享请求,将所述目标数据发送给所述云存储平台,所述共享请求用于请求在不同使用方之间共享所述目标数据;

所述云存储平台,用于为所述目标数据分配相应的共享目录并将所述目标数据存储至所述共享目录下。

10. 根据权利要求9所述的系统,其特征在于,所述云存储平台包括网络文件系统NFS和/或数据库。

11. 一种数据处理装置,其特征在于,包括:

第一接收模块,用于接收针对目标数据的处理请求,所述处理请求中携带对所述目标数据进行处理所需的处理配置参数;

第一创建模块,用于基于所述处理配置参数,创建用于处理所述目标数据的应用程序;

挂载模块,用于将所述应用程序的存储目录挂载到存放所述目标数据的共享目录上,所述应用程序在运行过程中对所述存储目录的访问操作被映射到所述共享目录上,所述共享目录为在云存储平台中为所述目标数据分配的、被至少两个使用方共享的目录,所述应用程序包括所述至少两个使用方分别对应的应用程序;

第二创建模块,用于在容器服务集群中创建容器并在所述容器中运行所述应用程序,其中,所述应用程序在运行过程中通过访问所述存储目录获取所述目标数据并进行处理。

12. 一种电子设备,其特征在于,包括:

处理器;

用于存储所述处理器可执行指令的存储器;

其中,所述处理器被配置为执行所述指令,以实现如权利要求1至7中任一项所述的方法。

13. 一种计算机可读存储介质,其特征在于,当所述存储介质中的指令由电子设备的处理器执行时,使得电子设备能够执行如权利要求1至7中任一项所述的方法。

数据处理方法和装置

技术领域

[0001] 本申请涉及计算机技术领域,尤其涉及一种数据处理方法和装置。

背景技术

[0002] 在一些业务场景下,比如对于机器学习模型的训练,传统的数据处理方案通常需由使用方将所需的数据从源端下载到本机,然后根据业务需求对所需的数据进行相应的数据处理操作。但是,如果上述数据需要经过不同的使用方进行相应处理,那么,同一份数据就需要被不同的使用方重复下载,这不仅降低了处理效率,还增加了处理成本。

发明内容

[0003] 本说明书实施例的目的是提供一种数据处理方法和装置,以解决传统的数据处理方案存在的处理效率低、处理成本高的问题。

[0004] 为了实现上述目的,本说明书实施例采用下述技术方案:

[0005] 第一方面,提供一种数据处理方法,包括:

[0006] 接收针对目标数据的处理请求,所述处理请求中携带对所述目标数据进行处理所需的处理配置参数;

[0007] 基于所述处理配置参数,创建用于处理所述目标数据的应用程序;

[0008] 将所述应用程序的存储目录挂载到存放所述目标数据的共享目录上,所述应用程序在运行过程中对所述存储目录的访问操作被映射到所述共享目录上;

[0009] 在容器服务集群中创建容器并在所述容器中运行所述应用程序,其中,所述应用程序在运行过程中通过访问所述存储目录获取所述目标数据并进行处理。

[0010] 第二方面,提供一种数据处理系统,包括:管理控制平台和容器服务集群;

[0011] 所述管理控制平台,用于接收针对目标数据的处理请求,所述处理请求中携带对所述目标数据进行处理所需的处理配置参数,基于所述处理配置参数,创建用于处理所述目标数据的应用程序,将所述应用程序的存放目录挂载到存放所述目标数据的共享目录上,所述存储目录被挂载到所述共享目录后,所述应用程序在运行过程中对所述存储目录的访问操作被映射到所述共享目录上,以及向所述容器服务集群发送容器创建请求,所述容器创建请求用于请求创建运行所述应用程序的容器;

[0012] 所述容器服务集群,用于接收来自所述管理控制平台的容器创建请求,创建容器并在所述容器中运行所述应用程序,其中,所述应用程序在运行过程中通过访问所述存储目录获取所述目标数据并进行处理。

[0013] 第三方面,提供一种数据处理装置,包括:

[0014] 第一接收模块,用于接收针对目标数据的处理请求,所述处理请求中携带对所述目标数据进行处理所需的处理配置参数;

[0015] 第一创建模块,用于基于所述处理配置参数,创建用于处理所述目标数据的应用程序;

[0016] 挂载模块,用于将所述应用程序的存储目录挂载到存放所述目标数据的共享目录上,所述应用程序在运行过程中对所述存储目录的访问操作被映射到所述共享目录上;

[0017] 第二创建模块,用于在容器服务集群中创建容器并在所述容器中运行所述应用程序,其中,所述应用程序在运行过程中通过访问所述存储目录获取所述目标数据并进行处理。

[0018] 第四方面,提供一种电子设备,包括:

[0019] 处理器;

[0020] 用于存储所述处理器可执行指令的存储器;

[0021] 其中,所述处理器被配置为执行所述指令,以实现如第一方面所述的方法。

[0022] 第五方面,提供一种计算机可读存储介质,当所述存储介质中的指令由电子设备的处理器执行时,使得电子设备能够执行如第一方面所述的方法。

[0023] 本说明书实施例的方案,通过将待共享的数据存储在能够被多方共享的共享目录下,然后,在使用方需要进行数据处理时,创建一个云环境的用于进行数据处理的应用程序,将应用程序的存储目录挂载到共享目录上,并通过容器服务集群中的容器来运行应用程序,这样就能够将应用程序在运行过程中对存储目录的访问操作映射到存放共享数据的目录上,这样,应用程序在运行过程中通过访问其存储目录便可获取到目标数据并对待共享数据进行处理,使得应用程序能够像在本机一样对共享数据进行处理,而无需将共享数据下载到本地,尤其在共享数据需要经不同使用方进行处理时,能够避免共享数据被重复下载,从而提高数据处理效率,降低数据处理成本。

附图说明

[0024] 此处所说明的附图用来提供对本说明书的进一步理解,构成本说明书的一部分,本说明书的示意性实施例及其说明用于解释本说明书,并不构成对本说明书的不当限定。在附图中:

[0025] 图1为本说明书的一个实施例提供的一种数据处理方法所适用的实施环境的示意图;

[0026] 图2为本说明书的一个实施例提供的一种数据处理方法的流程示意图;

[0027] 图3为本说明书的一个实施例提供的一种数据处理装置的结构示意图;

[0028] 图4为本说明书的一个实施例提供的一种电子设备的结构示意图。

具体实施方式

[0029] 为使本说明书的目的、技术方案和优点更加清楚,下面将结合本说明书具体实施例及相应的附图对本说明书技术方案进行清楚、完整地描述。显然,所描述的实施例仅是本说明书一部分实施例,而不是全部的实施例。基于本说明书中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都应当属于本文件保护的范围。

[0030] 本申请的说明书和权利要求书中的术语“第一”、“第二”等是用于区别类似的对象,而不用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换,以便本申请的实施例能够以除了在这里图示或描述的那些以外的顺序实施。此外,说明

书以及权利要求中“和/或”表示所连接对象的至少其中之一，字符“/”，一般表示前后关联对象是一种“或”的关系。

[0031] 如前所述，在一些业务场景下，比如对于机器学习模型的训练，传统的数据处理方案通常需要由使用方将所需的数据下载到本机，然后根据业务需求对所需的数据进行相应的数据处理操作。但是，如果上述数据需要经过不同的使用方进行相应处理，那么，同一份数据就需要被不同的使用方重复下载，这不仅降低了处理效率，还增加了处理成本。

[0032] 为此，本说明书实施例旨在提供一种数据处理方案，通过将待共享的数据存储在能够被多方共享的共享目录下，然后，在使用方需要进行数据处理时，创建一个云环境的用于进行数据处理的应用程序，将应用程序的存储目录挂载到共享目录上，并通过容器服务集群中的容器来运行应用程序，这样就能够将应用程序在运行过程中对存储目录的操作映射到存放共享数据的目录上，使得应用程序能够像在本地一样对共享数据进行处理，而无需将共享数据下载到本地，尤其在共享数据需要经不同使用方进行处理时，能够避免共享数据被重复下载，从而提高数据处理效率，降低数据处理成本。

[0033] 以下结合附图，详细说明本说明书各实施例提供的技术方案。

[0034] 为便于理解，首先结合图1对本说明书的一个实施例提供的一种数据处理方法所适用的实施环境进行说明。如图1所示，本说明书实施例的一种数据处理方法所适用的实施环境可包括管理控制平台、容器服务集群和云存储平台。

[0035] 在本说明书实施例中，管理控制平台能够基于硬件资源和软件资源的服务，提供计算、网络和存储等功能。具体而言，管理控制平台可以提供、数据上传、数据搜索、应用程序的创建及开启、目录挂载以及地址映射等功能。其中，数据上传功能用于供数据的提供方上传数据并对提供方上传的数据进行存储，数据搜索功能用于对使用方指定的数据进行搜索，应用程序的创建及开启功能用于创建并开启云环境的应用程序以执行相应的数据处理操作，目标挂载功能及地址映射功能用于将应用程序的存储目录挂载到相应的挂载点，从而将对应用程序的存储目录的操作映射到该挂载点上，使得应用程序能够像在本地一样对该挂载点进行操作。

[0036] 容器服务集群能够通过运行多个容器(Container)，由多个容器作为一个整体来对外提供服务，达到提高并发访问能力和避免单点故障的目的。每个容器上可以安装并运行相应的应用程序，以执行相应的处理操作。在本说明书实施例中，容器服务集群可以采用本领域技术人员已知的各种能够提供容器服务的集群，本说明书实施例对此不作具体限定。例如，容器服务集群可以为Kubernetes集群(简称“k8s集群”)，相应地，可以在Kubernetes集群中创建Pod，在Pod中运行应用程序。

[0037] 云存储平台能够提供数据存储功能，其可以存储提供方上传到管理控制平台的数据。可选地，云存储平台可以包括但不限于网络文件系统NFS和/或数据库等。

[0038] 具体而言，管理控制平台可以将提供方指定的待共享数据发送到云存储平台，由云存储平台为该待共享数据分配一相应的共享目录并将该待共享数据存储在该共享目录下，此外，管理控制平台还可以创建存储申明，并将所创建的存储申明与该共享目录绑定，由此使得使用方只需使用该存储申明即可访问该共享目录下的数据。其中，存储申明用于指定所要使用的存储资源。可选地，存储申明可以为持久化卷申明(Persistent Volume Claim,PVC)。

[0039] 当使用方需要使用该待共享数据时,可通过管理控制平台指定该待共享数据并提供用于对该待共享数据进行数据处理所需的处理配置参数等,管理控制平台可根据处理配置参数创建相应的应用程序以用于对待共享数据进行数据处理,并将应用程序的存储目录挂载到上述与共享目录绑定的存储申明上,以将应用程序在运行过程中对存储目录的访问操作映射到存放共享数据的目录上。然后,管理控制平台可通过容器服务集群建立容器并在容器中运行应用程序,这样,应用程序在运行过程中通过访问其存储目录便可获取到目标数据并对待共享数据进行处理,使得应用程序能够像在本地一样对共享目录下的待共享数据进行处理,而无需将共享数据下载到本地。

[0040] 例如,如图1所示,假设使用方A和使用方B均需要使用该待共享数据,且两者各自对待共享数据进行的处理操作不同,由此,管理控制平台可为使用方A创建应用程序A,为使用方B创建应用程序B,并将应用程序A和应用程序B各自的存储目录均挂载到与共享目录绑定的存储申明上。为了提高并发访问能力和避免单点故障,管理控制平台可在容器服务集群中分别建立容器A和容器B,在容器A上运行应用程序A,并在容器B上运行应用程序B,这样,应用程序A和应用程序B各自在运行过程中对其存储目录的访问操作就会映射到共享目录上,使得应用程序A和应用程序B均能够像在本地一样对共享目录下的待共享数据进行处理,而无需将待共享数据下载到本地,由此在共享数据需要经不同使用方进行处理时,能够避免共享数据被重复下载,从而提高数据处理效率,降低数据处理成本。

[0041] 基于上述实施环境,下面对本说明书实施例提供的一种数据处理方法进行说明。

[0042] 请参考图2,为本说明书的一个实施例提供的一种数据处理方法的流程示意图,该方法可应用于上述实施环境中的管理控制平台。如图2所示,该方法包括如下步骤:

[0043] S202,接收针对目标数据的处理请求。

[0044] 其中,处理请求用于请求对目标数据进行处理。处理请求中携带对目标数据进行处理所需的处理配置参数。当然,处理请求中还可携带目标数据的标识信息。

[0045] 本说明书实施例中的目标数据可以例如包括但不限于语料数据、机器学习模型、生物特征信息、脚本等,具体可根据实际业务需求进行选择,本说明书实施例对此不作具体限定。例如,在进行情感识别模型的训练时,目标数据可以为包含用户与智能客服之间的聊天记录记录的语料数据。

[0046] 对目标数据进行处理所需的处理配置参数可根据目标数据及对目标数据的处理操作的不同而不同。例如,如果目标数据为语料数据,对目标数据的处理操作为基于语料数据进行情感识别模型的训练,则对目标数据进行处理所需的处理配置参数可以例如包括但不限于用于进行情感识别模型训练的训练脚本、用于运行训练脚本的环境的相关参数等。

[0047] 又如,如果目标数据为机器学习模型,对目标数据的处理操作为对机器学习模型进行迭代更新,则对目标数据进行处理所需的处理配置参数可以例如包括但不限于对机器学习模型进行迭代更新的脚本、用于运行训练脚本的环境的相关参数等。

[0048] S204,基于处理请求中携带的处理配置参数,创建用于处理目标数据的应用程序。

[0049] 在本说明书实施例中,用于处理目标数据的应用程序可根据目标数据及对目标数据的处理操作的不同而不同。例如,仍以目标数据为语料数据、对目标数据的处理为利用语料数据训练情感识别模型为例,那么,用于处理目标数据的应用程序可以为用于运行训练脚本的编辑器。更为具体地,该编辑器可以为Jupiter Notebook,其为一种以网页的形式打

开、且可以在网页页面中直接编写代码和运行代码的编辑器。当然,在其他一些可选的方案中,编辑器也可以采用本领域技术人员已知的任意适当的编辑器。

[0050] S206,将应用程序的存储目录挂载到存放目标数据的共享目录上。

[0051] 存储目录被挂载到共享目录后,应用程序在运行过程中对存储目录的访问操作被映射到共享目录上。

[0052] 通常情况下,应用程序的存储目录用于存储应用程序运行所需的数据,应用程序在运行过程中通过访问存储目录以获取所需的数据并进行处理。这样,通过将应用程序的存储目录挂载到存放目标数据的共享目录上,应用程序在运行过程中通过访问其存储目录便可获取到目标数据并对待共享数据进行处理,进而使得应用程序能够像在本地一样对共享数据进行处理,而无需将共享数据下载到本地。

[0053] 可选地,上述S206可以包括:获取与存放目标数据的共享目录绑定的存储申明,并将应用程序的存储目录挂载到该存储申明上。

[0054] 可以理解的是,存储申明是一种用于指定所要使用的存储资源的申明,该存储申明与共享目录绑定后,任何使用该存储申明的使用方都可访问该共享目录以从中获取所需的数据。将应用程序的存储目录挂载到该存储申明上,相当于在应用程序的存储目录与共享目录之间建立一座“桥梁”,应用程序对其存储目录的访问操作会映射到共享目录上,因而使得应用程序无需将共享目录中的数据下载到存储目录中,而是可以通过访问其存储目录来达到访问共享目录的效果。

[0055] 本说明书实施例中,存储申明可以具有任意适当的形式,具体可以根据实际需要设置,本说明书实施例对此不作具体限定。在一种较为优选的实现方式中,存储申明可以为PVC。由于PVC是用于数据存储的一种申明,其可以请求特定的存储空间和访问模式,使得使用方在进行数据存储及访问时,无需了解底层的实现细节,只需要直接使用PVC即可,因而通过将存放数据的目录与PVC预先绑定,使得使用方只需使用与该目录绑定的PVC即可访问该目录下的数据。在此基础上,通过获取与存放目标数据的共享目录绑定的PVC,并将应用程序的存储目录挂载到该PVC上,可以确保对应用程序在运行过程中对其存储目录的操作都将会被映射到共享目录上,进而后能确保后续通过应用程序对目标数据的处理过程的可靠性。

[0056] 当然,在其他一些可选的实施方式中,也可以采用本领域技术人员已知的各种技术手段实现将应用程序的存储目录挂载到存放目标数据的共享目录上,本说明书实施例对此不作具体限定。

[0057] S208,在容器服务集群中创建容器并在容器中运行应用程序,其中,应用程序在运行过程中通过访问存储目录获取目标数据并进行处理。

[0058] 考虑到在一些业务场景下,同一份目标数据可能需要经过多个不同的处理方进行处理,而容器服务集群(Cluster)能够通过运行多个容器(Container),由多个容器作为一个整体来对外提供服务,达到提高并发访问能力和避免单点故障的目的,基于此,可以在容器服务集群中创建相应的容器,通过所创建的容器来安装并运行用于对目标数据进行处理的应用程序。

[0059] 在本说明书实施例中,容器服务集群可以采用本领域技术人员已知的各种能够提供容器服务的集群,本说明书实施例对此不作具体限定。例如,容器服务集群可以为

Kubernetes集群,相应地,可以在Kubernetes集群中创建Pod,在Pod中运行用于对目标数据进行处理的应用程序。

[0060] 本说明书实施例提供的数据处理方法,通过将待共享的数据存储在能够被多方共享的共享目录下,然后,在使用方需要进行数据处理时,创建一个云环境的用于进行数据处理的应用程序,将应用程序的存储目录挂载到共享目录上,并通过容器服务集群中的容器来运行应用程序,这样就能够将应用程序在运行过程中对存储目录的访问操作映射到存放共享数据的目录上,这样,应用程序在运行过程中通过访问其存储目录便可获取到目标数据并对待共享数据进行处理,使得应用程序能够像在本地一样对共享数据进行处理,而无需将共享数据下载到本地,尤其在共享数据需要经不同使用方进行处理时,能够避免共享数据被重复下载,从而提高数据处理效率,降低数据处理成本。

[0061] 在上述实施例的基础上,本说明书实施例的数据处理方法还可以包括对目标数据的存储及将存放目标数据的共享目录与存储申明进行绑定。具体而言,在上述S202之前,本说明书实施例的数据处理方法还可以包括:接收针对目标数据的共享请求,该共享请求用于请求在不同使用方之间共享目标数据;在云存储平台中为目标数据分配相应的共享目录,并将目标数据存储至该共享目录下;创建存储申明,并将所创建的存储申明与共享目录进行绑定。

[0062] 需要说明的是,实际应用中,目标数据的提供方可以向管理控制平台发送携带有目标数据的共享请求,以请求管理控制平台在不同使用方之间共享目标数据;或者,目标数据的提供方也可以在管理控制平台的前端界面指定需要共享的目标数据,由管理控制平台对目标数据标记为共享模式。进一步地,管理控制平台即可在云存储平台中为需要共享的目标数据分配相应的共享目录并将目标数存储至该共享目录下。其次,目标数据的提供方可以对目标数据进行修改后再次上传到管理控制平台,在此情况下,管理控制平台可在接收到提供方上传的目标数据后,将该目标数据与共享目录下的目标数据进行比对,以确定目标数据是否被修改,如果是,则可将目标数据覆盖存储在共享目录下,以确保使用方能够获得到正确的目标数据,进而确保使用方对目标数据的处理能够正常进行。另外,云存储平台可以例如包括但不限于网络文件系统(Network File System,NFS)、数据库等,其中,数据库可以例如包括但不限于任意适当类型的数据库,例如MySQL数据库等。

[0063] 可以理解的是,通过在云存储平台中为目标数据分配相应的共享目录,创建存储申明并将该存储申明与将该共享目录进行绑定,使得任意需要使用目标数据的一方通过该存储申明即可访问该共享目录以获取到目标数据。

[0064] 可选地,在上述实施例的基础上,在上述S208之后,本说明书实施例的数据处理方法还可以包括:生成目标数据与共享目录之间的对应关系信息,以及向共享请求的发起方返回该对应关系信息。由此,使得共享请求的发起方能够清楚地获知目标数据的存储位置,以便能够快速地对目标数据进行搜索查询。

[0065] 可选地,考虑到目标数据的处理结果可供其他使用方所用,比如在多方联合进行机器学习模型的训练时,一方对目标数据的处理结果(比如特征提取结果、训练出的模型等)需要共享给其他使用方进行进一步处理,为此,在上述实施例的基础上,在上述S208之后,本说明书实施例的数据处理方法还可以包括:从应用程序的存储目录中加载对目标数据进行处理所得的处理结果,将处理结果存储至共享目录下。可以理解的是,通过将目标数

据的处理结果存储至共享目录下,可以实现该处理结果在不同使用方之间的共享,方便其他使用方进行进一步的数据处理。

[0066] 可选地,考虑到在对目标数据进行处理时可能对目标数据进行修改,例如在机器学习模型的训练过程中会通过不断迭代训练对初始模型进行优化,以供其他使用方使用,为此,在上述实施例的基础上,本说明书实施例的数据处理方法还包括:在将对目标数据进行处理所得的处理结果存储至共享目录下之前,基于该处理结果,确定目标数据是否被修改,如果目标数据被修改,则将修改后的目标数据覆盖存储至共享目录下,使得共享目录下仅保留修改后的目标数据,确保其他使用方能够获取到正确的目标数据,进而确保使用方对目标数据的处理能够正常进行。

[0067] 当然,在其他一些可选的方案中,也可将修改后的目标数据直接存储在共享目录下而不对共享目录下原有的目标数据进行覆盖,使得目标数据及修改后的目标数据均存储在共享目录下。

[0068] 此外,与上述图2所示的数据处理方法相对应地,本说明书实施例还提供一种数据处理装置。图3是本说明书实施例提供的一种数据处理装置300的结构示意图,包括:

[0069] 第一接收模块310,用于接收针对目标数据的处理请求,所述处理请求中携带对所述目标数据进行处理所需的处理配置参数;

[0070] 第一创建模块320,用于基于所述处理配置参数,创建用于处理所述目标数据的应用程序;

[0071] 挂载模块330,用于将所述应用程序的存储目录挂载到存放所述目标数据的共享目录上,所述存储目录被挂载到所述共享目录后,所述应用程序在运行过程中对所述存储目录的访问操作被映射到所述共享目录上;

[0072] 第二创建模块340,用于在容器服务集群中创建容器并在所述容器中运行所述应用程序,其中,所述应用程序在运行过程中通过访问所述存储目录获取所述目标数据并进行处理。

[0073] 本说明书实施例提供的数据处理装置,通过将待共享的数据存储在能够被多方共享的共享目录下,然后,在使用方需要进行数据处理时,创建一个云环境的用于进行数据处理的应用程序,将应用程序的存储目录挂载到共享目录上,并通过容器服务集群中的容器来运行应用程序,这样就能够将应用程序在运行过程中对存储目录的访问操作映射到存放共享数据的目录上,这样,应用程序在运行过程中通过访问其存储目录便可获取到目标数据并对待共享数据进行处理,使得应用程序能够像在本地一样对共享数据进行处理,而无需将共享数据下载到本地,尤其在共享数据需要经不同使用方进行处理时,能够避免共享数据被重复下载,从而提高数据处理效率,降低数据处理成本。

[0074] 可选地,所述挂载模块330包括:

[0075] 申明获取子模块,用于获取与存放所述目标数据的共享目录绑定的存储申明;

[0076] 挂载子模块,用于将所述应用程序的存储目录挂载到所述存储申明上。

[0077] 可选地,所述装置还包括:

[0078] 第二接收模块,用于接收针对所述目标数据的共享请求,所述共享请求用于请求在不同使用方之间共享所述目标数据;

[0079] 第一存储模块,用于在云存储平台中为所述目标数据分配相应的共享目录,并将

所述目标数据存储至所述共享目录下；

[0080] 第三创建模块,用于创建存储申明,并将所创建的存储申明与所述共享目录进行绑定。

[0081] 可选地,所述装置还包括:

[0082] 生成模块,用于生成所述目标数据与所述共享目录之间的对应关系信息;

[0083] 发送模块,用于向所述共享请求的发起方返回所述对应关系信息。

[0084] 可选地,所述装置还包括:

[0085] 加载模块,用于从所述应用程序的存储目录中加载对所述目标数据进行处理所得的处理结果;

[0086] 第二存储模块,用于将所述处理结果存储至所述共享目录下。

[0087] 可选地,所述装置还包括:

[0088] 检测模块,用于基于所述处理结果,确定所述目标数据是否被修改;

[0089] 所述第二存储模块包括:

[0090] 存储子模块,用于在所述目标数据被修改时,将修改后的所述目标数据覆盖存储至所述共享目录下。

[0091] 可选地,所述目标数据包括语料数据和/或机器学习模型,所述处理配置参数为用于利用所述目标数据进行模型训练的训练脚本。

[0092] 显然,本说明书实施例的数据处理装置可以作为上述图1所示的数据处理方法的执行主体,因此能够实现数据处理方法在图1所实现的功能。由于原理相同,在此不再赘述。

[0093] 本说明书实施例还提供一种数据处理系统,该系统包括管理控制平台(比如图1所示实施环境中的管理控制平台1)和容器服务集群(比如图1所示实施环境中的容器服务集群2)。

[0094] 所述管理控制平台,用于接收针对目标数据的处理请求,所述处理请求中携带对所述目标数据进行处理所需的处理配置参数,基于所述处理配置参数,创建用于处理所述目标数据的应用程序,将所述应用程序的存放目录挂载到存放所述目标数据的共享目录上,所述存储目录被挂载到所述共享目录后,所述应用程序在运行过程中对所述存储目录的访问操作被映射到所述共享目录上,以及向所述容器服务集群发送容器创建请求,所述容器创建请求用于请求创建运行所述应用程序的容器;

[0095] 所述容器服务集群,用于接收来自所述管理控制平台的容器创建请求,创建容器并在所述容器中运行所述应用程序,其中,所述应用程序在运行过程中通过访问所述存储目录获取所述目标数据并进行处理。

[0096] 可选地,所述管理控制平台,用于获取与存放所述目标数据的共享目录绑定的存储申明,将所述应用程序的存储目录挂载到所述存储申明上。

[0097] 可选地,所述系统还包括云存储平台;

[0098] 所述管理控制平台,还用于接收针对所述目标数据的共享请求,将所述目标数据发送给所述云存储平台,所述共享请求用于请求在不同使用方之间共享所述目标数据;

[0099] 所述云存储平台,用于为所述目标数据分配相应的共享目录并将所述目标数据存储至所述共享目录下。

[0100] 可选地,所述云存储平台包括网络文件系统NFS和/或数据库。

[0101] 可选地,所述管理控制平台,还用于在所述云存储平台将所述目标数据存储至所述共享目录下之后,生成所述目标数据与所述共享目录之间的对应关系信息,以及向所述共享请求的发起方返回所述对应关系信息。

[0102] 可选地,所述管理控制平台,还用于在所述容器服务集群创建容器,以在所述容器中运行所述应用程序之后,从所述应用程序的存储目录中加载对所述目标数据进行处理所得的处理结果,并将所述处理结果存储至所述共享目录下。

[0103] 可选地,所述管理控制平台,还用于在将所述处理结果存储至所述共享目录下之前,基于所述处理结果,确定所述目标数据是否被修改,如果所述目标数据被修改,则将修改后的所述目标数据覆盖存储至所述共享目录下。

[0104] 可选地,所述目标数据包括语料数据和/或机器学习模型,所述处理配置参数为用于利用所述目标数据进行模型训练的训练脚本。

[0105] 本说明书实施例提供的数据处理系统,通过统一的管理控制平台通过将待共享的数据存储在能够被多方共享的共享目录下,然后,在使用方需要进行数据处理时,创建一个云环境的用于进行数据处理的应用程序,将应用程序的存储目录挂载到共享目录上,并通过容器服务集群中的容器来运行应用程序,这样就能够将应用程序在运行过程中对存储目录的访问操作映射到存放共享数据的目录上,这样,应用程序在运行过程中通过访问其存储目录便可获取到目标数据并对待共享数据进行处理,使得应用程序能够像在本地一样对共享数据进行处理,而无需将共享数据下载到本地,尤其在共享数据需要经不同使用方进行处理时,能够避免共享数据被重复下载,从而提高数据处理效率,降低数据处理成本。

[0106] 图4是本说明书的一个实施例电子设备的结构示意图。请参考图4,在硬件层面,该电子设备包括处理器,可选地还包括内部总线、网络接口、存储器。其中,存储器可能包含内存,例如高速随机存取存储器(Random-Access Memory, RAM),也可能还包括非易失性存储器(non-volatile memory),例如至少1个磁盘存储器等。当然,该电子设备还可能包括其他业务所需要的硬件。

[0107] 处理器、网络接口和存储器可以通过内部总线相互连接,该内部总线可以是ISA (Industry Standard Architecture,工业标准体系结构) 总线、PCI (Peripheral Component Interconnect,外设部件互连标准) 总线或EISA(Extended Industry Standard Architecture,扩展工业标准结构) 总线等。所述总线可以分为地址总线、数据总线、控制总线等。为便于表示,图4中仅用一个双向箭头表示,但并不表示仅有一根总线或一种类型的总线。

[0108] 存储器,用于存放程序。具体地,程序可以包括程序代码,所述程序代码包括计算机操作指令。存储器可以包括内存和非易失性存储器,并向处理器提供指令和数据。

[0109] 处理器从非易失性存储器中读取对应的计算机程序到内存中然后运行,在逻辑层面上形成数据处理装置。处理器,执行存储器所存放的程序,并具体用于执行以下操作:

[0110] 接收针对目标数据的处理请求,所述处理请求中携带对所述目标数据进行处理所需的处理配置参数;

[0111] 基于所述处理配置参数,创建用于处理所述目标数据的应用程序;

[0112] 将所述应用程序的存储目录挂载到存放所述目标数据的共享目录上,所述存储目录被挂载到所述共享目录后,所述应用程序在运行过程中对所述存储目录的访问操作被映

射到所述共享目录上；

[0113] 在容器服务集群中创建容器并在所述容器中运行所述应用程序，其中，所述应用程序在运行过程中通过访问所述存储目录获取所述目标数据并进行处理。

[0114] 上述如本说明书图2所示实施例揭示的数据处理装置执行的方法可以应用于处理器中，或者由处理器实现。处理器可能是一种集成电路芯片，具有信号的处理能力。在实现过程中，上述方法的各步骤可以通过处理器中的硬件的集成逻辑电路或者软件形式的指令完成。上述的处理器可以是通用处理器，包括中央处理器(Central Processing Unit, CPU)、网络处理器(Network Processor, NP)等；还可以是数字信号处理器(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。可以实现或者执行本说明书实施例中的公开的各方法、步骤及逻辑框图。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。结合本说明书实施例所公开的方法的步骤可以直接体现为硬件译码处理器执行完成，或者用译码处理器中的硬件及软件模块组合执行完成。软件模块可以位于随机存储器，闪存、只读存储器，可编程只读存储器或者电可擦写可编程存储器、寄存器等本领域成熟的存储介质中。该存储介质位于存储器，处理器读取存储器中的信息，结合其硬件完成上述方法的步骤。

[0115] 应理解，本说明书实施例的电子设备可以实现数据处理装置在图2所示实施例的功能。由于原理相同，本说明书实施例在此不再赘述。

[0116] 当然，除了软件实现方式之外，本说明书的电子设备并不排除其他实现方式，比如逻辑器件抑或软硬件结合的方式等等，也就是说以下处理流程的执行主体并不限于各个逻辑单元，也可以是硬件或逻辑器件。

[0117] 本说明书实施例还提出了一种计算机可读存储介质，该计算机可读存储介质存储一个或多个程序，该一个或多个程序包括指令，该指令当被包括多个应用程序的便携式电子设备执行时，能够使该便携式电子设备执行图2所示实施例的方法，并具体用于执行以下操作：

[0118] 接收针对目标数据的处理请求，所述处理请求中携带对所述目标数据进行处理所需的处理配置参数；

[0119] 基于所述处理配置参数，创建用于处理所述目标数据的应用程序；

[0120] 将所述应用程序的存储目录挂载到存放所述目标数据的共享目录上，所述存储目录被挂载到所述共享目录后，所述应用程序在运行过程中对所述存储目录的访问操作被映射到所述共享目录上；

[0121] 在容器服务集群中创建容器并在所述容器中运行所述应用程序，其中，所述应用程序在运行过程中通过访问所述存储目录获取所述目标数据并进行处理。

[0122] 上述对本说明书特定实施例进行了描述。其它实施例在所附权利要求书的范围内。在一些情况下，在权利要求书中记载的动作或步骤可以按照不同于实施例中的顺序来执行并且仍然可以实现期望的结果。另外，在附图中描绘的过程不一定要示出的特定顺序或者连续顺序才能实现期望的结果。在某些实施方式中，多任务处理和并行处理也是可以的或者可能是有利的。

[0123] 总之,以上所述仅为本说明书的较佳实施例而已,并非用于限定本说明书的保护范围。凡在本说明书的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本说明书的保护范围之内。

[0124] 上述实施例阐明的系统、装置、模块或单元,具体可以由计算机芯片或实体实现,或者由具有某种功能的产品来实现。一种典型的实现设备为计算机。具体的,计算机例如可以为个人计算机、膝上型计算机、蜂窝电话、相机电话、智能电话、个人数字助理、媒体播放器、导航设备、电子邮件设备、游戏控制台、平板计算机、可穿戴设备或者这些设备中的任何设备的组合。

[0125] 计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括,但不限于相变内存(PRAM)、静态随机存取存储器(SRAM)、动态随机存取存储器(DRAM)、其他类型的随机存取存储器(RAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器(CD-ROM)、数字多功能光盘(DVD)或其他光学存储、磁盒式磁带,磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质,可用于存储可以被计算设备访问的信息。按照本文中的界定,计算机可读介质不包括暂存电脑可读媒体(transitory media),如调制的数据信号和载波。

[0126] 还需要说明的是,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、商品或者设备不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、商品或者设备所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括所述要素的过程、方法、商品或者设备中还存在另外的相同要素。

[0127] 本说明书中的各个实施例均采用递进的方式描述,各个实施例之间相同相似的部分互相参见即可,每个实施例重点说明的都是与其他实施例的不同之处。尤其,对于系统实施例而言,由于其基本相似于方法实施例,所以描述的比较简单,相关之处参见方法实施例的部分说明即可。

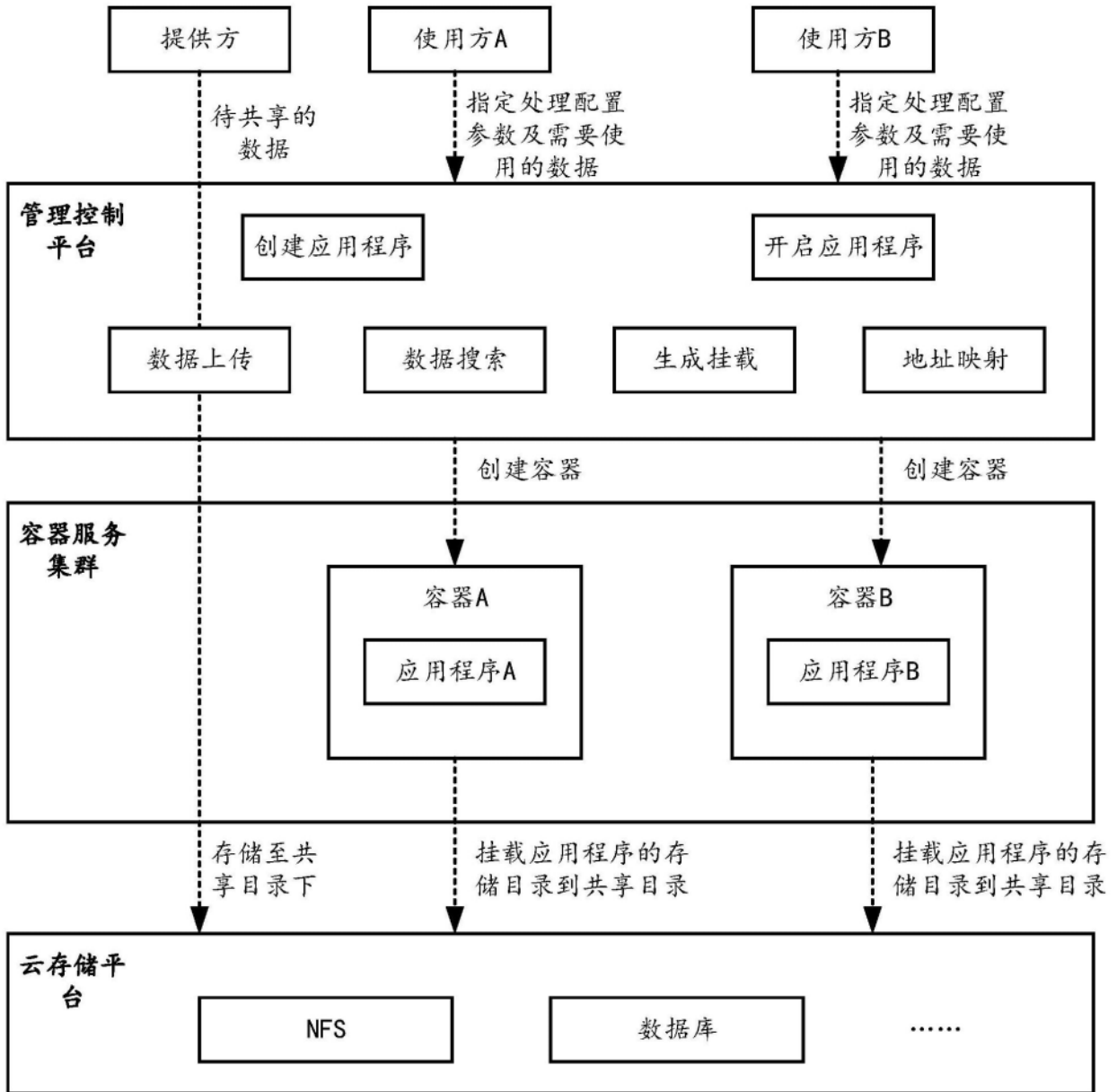


图1

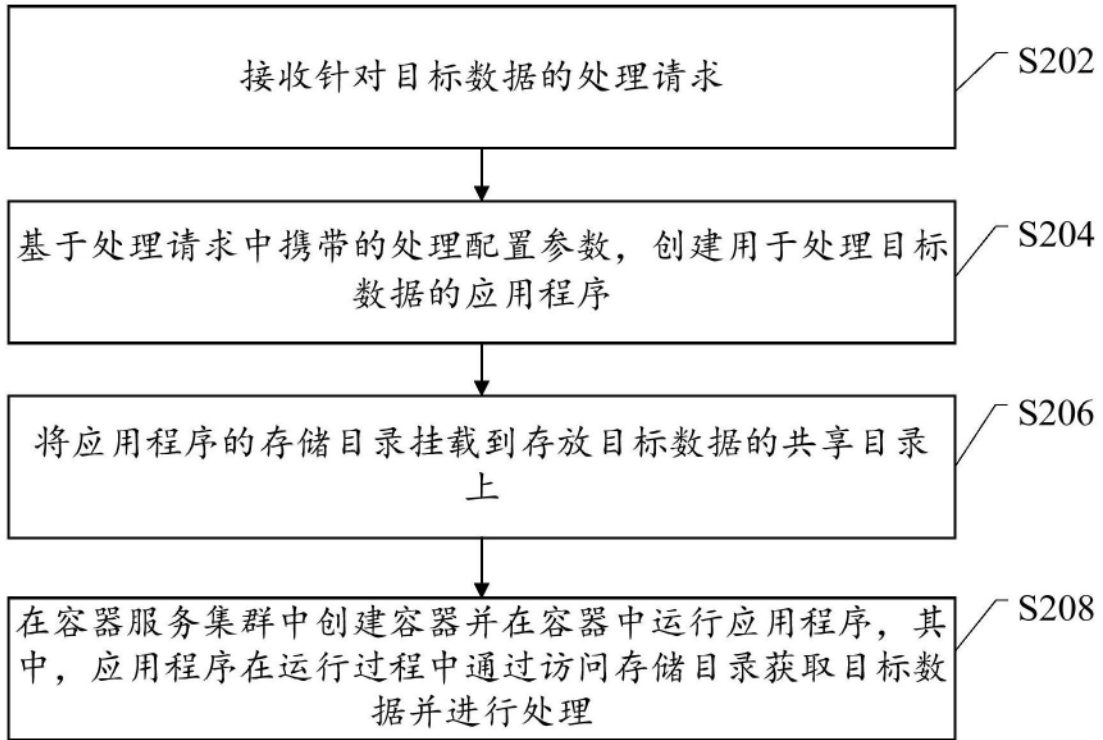


图2

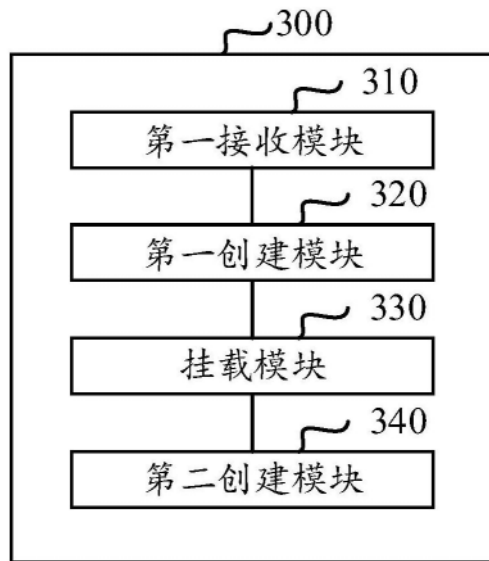


图3

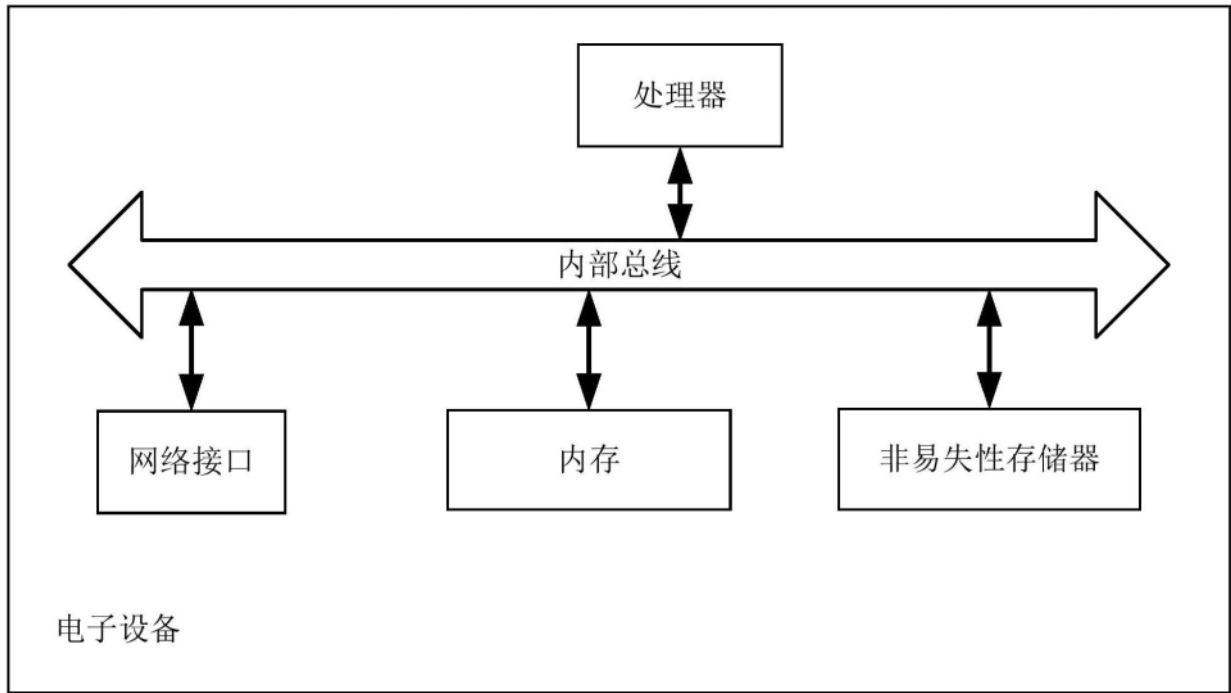


图4