(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 118819864 A
(43)申请公布日 2024.10.22

(21)申请号 202411295538.X

(22)申请日 2024.09.18

(71)申请人 北京景行锐创软件有限公司
地址 100010 北京市朝阳区红军营南路15号院1号楼-2至12层101五层507C室

(72)发明人 郑奕

(74)专利代理机构 北京高沃律师事务所 11569
专利代理师 赵昕

(51)Int.Cl.
G06F 9/50(2006.01)
G06F 9/48(2006.01)
H04L 67/1008(2022.01)

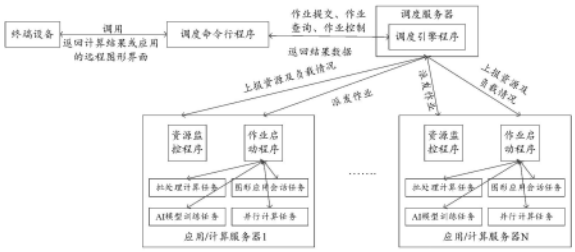
权利要求书4页 说明书16页 附图4页

(54)发明名称

多类型负载的资源统一调度方法及系统

(57)摘要

本申请公开了一种多类型负载的资源统一调度方法和系统,涉及数据处理领域,该方法包括:接收不同类型的目标作业负载类型的应用任务,以及各个目标作业负载类型的应用任务对应的资源需求信息,其中,不同类型的目标作业负载类型的应用任务至少包括图形应用会话任务;根据当前目标作业负载类型的应用任务对应的资源需求信息,在集群资源信息表中查找满足当前目标作业负载类型的应用任务对应的资源需求信息的候选服务器;并将当前目标作业负载类型的应用任务和当前目标作业负载类型的应用任务对应的资源需求信息发送给候选服务器中选择负载最低的服务器。本申请可以提升终端设备连接服务器的效率以及集群中服务器资源的整体利用率。



CN 118819864 A

1.一种多类型负载的资源统一调度方法,其特征在于,所述多类型负载的资源统一调度方法包括:

接收不同类型的目标作业负载类型的应用任务,以及各个所述目标作业负载类型的应用任务对应的资源需求信息,其中,所述不同类型的目标作业负载类型的应用任务至少包括图形应用会话任务;

根据当前目标作业负载类型的应用任务对应的资源需求信息,在集群资源信息表中查找满足所述当前目标作业负载类型的应用任务对应的资源需求信息的候选服务器;所述集群资源信息表中包括:集群中,各个服务器的属性信息、已用资源信息和空闲资源信息;

在所述候选服务器中选择负载最低的服务器作为第一目标服务器;

将所述当前目标作业负载类型的应用任务和所述当前目标作业负载类型的应用任务对应的资源需求信息发送给所述第一目标服务器,以使所述第一目标服务器运行所述当前目标作业负载类型的应用任务。

2.根据权利要求1所述的多类型负载的资源统一调度方法,其特征在于,

所述第一目标服务器为正在运行应用任务但有空闲资源的服务器,所述空闲资源可以满足所述当前目标作业负载类型的应用任务对应的资源需求信息,且所述第一目标服务器正在运行的应用任务与所述当前目标作业负载类型的应用任务不相同,在运行了所述当前目标作业负载类型的应用任务后,所述第一目标服务器中运行了不同类型的应用任务。

3.根据权利要求1所述的多类型负载的资源统一调度方法,其特征在于,所述根据当前目标作业负载类型的应用任务对应的资源需求信息,在集群资源信息表中查找满足所述当前目标作业负载类型的应用任务对应的资源需求信息的候选服务器,包括:

获取所述当前目标作业负载类型的应用任务对应的运行时间,将其放入对应的目标时间窗口中,其中,不同的时间窗口对应不同的运行时间段,且不同类型的目标作业负载类型的应用任务对应不同的运行时间段;

若当前时间满足所述目标时间窗口对应的时间段要求,则根据所述当前目标作业负载类型的应用任务对应的资源需求信息,在所述集群资源信息表中查找满足所述当前目标作业负载类型的应用任务对应的资源需求信息的所述候选服务器;

若当前时间不满足所述目标时间窗口对应的时间段要求,向所述当前目标作业负载类型的应用任务对应的终端设备发送拒绝处理指令,所述拒绝处理指令指示当前时间不满足所述当前目标作业负载类型的应用任务对应的运行时间;或者,将所述当前目标作业负载类型的应用任务和所述当前目标作业负载类型的应用任务对应的资源需求信息放入对应的等待队列中排队,并在检测到满足所述目标时间窗口对应的时间段要求时,根据所述当前目标作业负载类型的应用任务对应的资源需求信息,在所述集群资源信息表中重新查找满足所述当前目标作业负载类型的应用任务对应的资源需求信息的所述候选服务器。

4.根据权利要求1所述的多类型负载的资源统一调度方法,其特征在于,所述多类型负载的资源统一调度方法,还包括:

若在所述集群资源信息表中未查找到满足所述当前目标作业负载类型的应用任务对应的资源需求信息的候选服务器,获取所述当前目标作业负载类型的应用任务的优先级;

检测所述集群的各个所述服务器中是否有第二目标服务器,其中,所述第二目标服务器中正在运行的第一应用任务的负载类型的优先级低于所述当前目标作业负载类型的应

用任务的优先级,且所述第二目标服务器在释放了为正在运行的第一应用任务分配的资源后,所述第二目标服务器中的空闲资源满足所述当前目标作业负载类型的应用任务对应的资源需求信息;

控制所述第二目标服务器将正在运行的第一应用任务挂起或终止,以使所述第二目标服务器释放为所述第一应用任务分配的资源;

将所述当前目标作业负载类型的应用任务和所述当前目标作业负载类型的应用任务对应的资源需求信息发送给所述第二目标服务器,以让所述第二目标服务器运行所述当前目标作业负载类型的应用任务;

在检测到所述当前目标作业负载类型的应用任务运行完成后,控制所述第二目标服务器继续运行所述第一应用任务。

5.根据权利要求1所述的多类型负载的资源统一调度方法,其特征在于,在所述目标作业负载类型的应用任务为图形应用会话任务时,所述多类型负载的资源统一调度方法,还包括:

若在集群资源信息表中未查找到满足所述当前目标作业负载类型的应用任务对应的资源需求信息的候选服务器,检测所述集群中是否有第三目标服务器,其中,所述第三目标服务器中正在运行与所述当前目标作业负载类型的应用任务属于同一终端设备的其他应用任务,所述其他应用任务的负载类型与所述当前目标作业负载类型相同;

若有,将所述当前目标作业负载类型的应用任务和所述当前目标作业负载类型的应用任务对应的资源需求信息发送给所述第三目标服务器。

6.根据权利要求1-5任一项所述的多类型负载的资源统一调度方法,其特征在于,所述多类型负载的资源统一调度方法,还包括:

接收集群中的各个所述服务器发送的集群资源更新信息,所述集群资源更新信息中包括:当前时间点对应的各个所述服务器的已用资源信息和空闲资源信息;

根据所述集群资源更新信息更新所述集群资源信息表。

7.一种多类型负载的资源统一调度系统,其特征在于,所述多类型负载的资源统一调度系统包括:

终端设备、调度服务器和包括多个服务器的集群,其中,多个所述服务器包括第一目标服务器;

所述终端设备,用于向所述调度服务器发送不同类型的目标作业负载类型的应用任务,以及各个所述目标作业负载类型的应用任务对应的资源需求信息,其中,所述不同类型的目标作业负载类型的应用任务至少包括图形应用会话任务;

所述调度服务器,用于根据当前目标作业负载类型的应用任务对应的资源需求信息,在集群资源信息表中查找满足所述当前目标作业负载类型的应用任务对应的资源需求信息的候选服务器;所述集群资源信息表中包括:集群中,各个服务器的属性信息、已用资源信息和空闲资源信息;

所述调度服务器,用于在所述候选服务器中选择负载最低的服务器作为第一目标服务器,并将所述当前目标作业负载类型的应用任务和所述当前目标作业负载类型的应用任务对应的资源需求信息发送给所述第一目标服务器;

所述第一目标服务器,用于在接收到所述当前目标作业负载类型的应用任务和所述当

前目标作业负载类型的应用任务对应的资源需求信息后,根据所述当前目标作业负载类型的应用任务对应的资源需求信息运行所述当前目标作业负载类型的应用任务。

8.根据权利要求7所述的多类型负载的资源统一调度系统,其特征在于,

所述第一目标服务器为正在运行应用任务但有空闲资源的服务器,所述空闲资源可以满足所述当前目标作业负载类型的应用任务对应的资源需求信息,且所述第一目标服务器正在运行的应用任务与所述当前目标作业负载类型的应用任务不相同,在运行了所述当前目标作业负载类型的应用任务后,所述第一目标服务器中运行了不同类型的应用任务。

9.根据权利要求7所述的多类型负载的资源统一调度系统,其特征在于,所述调度服务器,具体用于:

获取所述当前目标作业负载类型的应用任务对应的运行时间,将其放入对应的目标时间窗口中,其中,不同的时间窗口对应不同的运行时间段,且不同类型的目标作业负载类型的应用任务对应不同的运行时间段;

若当前时间满足所述目标时间窗口对应的时间段要求,则根据所述当前目标作业负载类型的应用任务对应的资源需求信息,在所述集群资源信息表中查找满足所述当前目标作业负载类型的应用任务对应的资源需求信息的候选服务器;

若当前时间不满足所述目标时间窗口对应的时间段要求,向所述当前目标作业负载类型的应用任务对应的终端设备发送拒绝处理指令,所述拒绝处理指令指示当前时间不满足所述当前目标作业负载类型的应用任务对应的运行时间;或者,将所述当前目标作业负载类型的应用任务和所述当前目标作业负载类型的应用任务对应的资源需求信息存入缓存中,并在检测到满足所述目标时间窗口对应的时间段要求时,根据所述当前目标作业负载类型的应用任务对应的资源需求信息,在所述集群资源信息表中重新查找满足所述当前目标作业负载类型的应用任务对应的资源需求信息的所述候选服务器。

10.根据权利要求7所述的多类型负载的资源统一调度系统,其特征在于,

所述调度服务器,还用于:若在集群资源信息表中未查找到满足所述当前目标作业负载类型的应用任务对应的资源需求信息的候选服务器,获取所述当前目标作业负载类型的应用任务的优先级;检测所述集群的各个所述服务器中是否有第二目标服务器,其中,所述第二目标服务器中正在运行的第一应用任务的负载类型的优先级低于所述当前目标作业负载类型的应用任务的优先级,且所述第二目标服务器在释放了为正在运行的第一应用任务分配的资源后,所述第二目标服务器中的空闲资源满足所述当前目标作业负载类型的应用任务对应的资源需求信息;控制所述第二目标服务器将正在运行的第一应用任务挂起或终止,以使所述第二目标服务器释放为所述第一应用任务分配的资源;将所述当前目标作业负载类型的应用任务和所述当前目标作业负载类型的应用任务对应的资源需求信息发送给所述第二目标服务器,以让所述第二目标服务器运行所述当前目标作业负载类型的应用任务;在检测到所述当前目标作业负载类型的应用任务运行完成后,控制所述第二目标服务器继续运行所述第一应用任务;

或者,

所述调度服务器,还用于:在所述目标作业负载类型的应用任务为图形应用会话任务时,若在集群资源信息表中未查找到满足所述当前目标作业负载类型的应用任务对应的资源需求信息的候选服务器,检测所述集群中是否有第三目标服务器,其中,所述第三目标服

务器中正在运行与所述当前目标作业负载类型的应用任务属于同一终端设备的其他应用任务,所述其他应用任务的负载类型与所述当前目标作业负载类型相同；若有,将所述当前目标作业负载类型的应用任务和所述当前目标作业负载类型的应用任务对应的资源需求信息发送给所述第三目标服务器。

多类型负载的资源统一调度方法及系统

技术领域

[0001] 本申请涉及数据处理领域,特别是涉及一种多类型负载的资源统一调度方法及系统。

背景技术

[0002] 用户通过终端设备使用应用时,往往是需要启用服务器来进行相关操作的。

[0003] 目前,用户会通过终端设备在服务器集群中自找一台服务器进行尝试连接以启动图形应用,如果尝试连接的服务器的所有资源均已被其他终端设备占用,那么,表明终端设备尝试失败,然后终端设备会再去尝试着连接其他的服务器,直至与某一个服务器连接成功,以让该服务器启动终端设备的图形应用。

[0004] 但该种方式,由于终端设备可能需要多次尝试着去连接服务器,从而使得终端设备的连接时间过长,启动服务器效率较低,并且还会出现不同的终端设备使用同一台服务器来运行应用任务,从而导致有些服务器很忙,有些服务器闲置,集群中服务器资源负载不均衡。

发明内容

[0005] 本申请的目的是提供一种多类型负载的资源统一调度方法及系统,可提升终端设备连接服务器的效率以及集群中服务器资源的整体利用率。

[0006] 为实现上述目的,本申请提供了如下方案:

第一方面,本申请提供了一种多类型负载的资源统一调度方法,所述多类型负载的资源统一调度方法 包括:

接收不同类型的目标作业负载类型的应用任务,以及各个所述目标作业负载类型的应用任务对应的资源需求信息,其中,所述不同类型的目标作业负载类型的应用任务至少包括图形应用会话任务;

根据当前目标作业负载类型的应用任务对应的资源需求信息,在集群资源信息表中查找满足所述当前目标作业负载类型的应用任务对应的资源需求信息的候选服务器;所述集群资源信息表中包括:集群中,各个服务器的属性信息、已用资源信息和空闲资源信息;

在所述候选服务器中选择负载最低的服务器作为第一目标服务器;

将所述当前目标作业负载类型的应用任务和所述当前目标作业负载类型的应用任务对应的资源需求信息发送给所述第一目标服务器,以使所述第一目标服务器运行所述当前目标作业负载类型的应用任务。

[0007] 第二方面,本申请提供了一种多类型负载的资源统一调度系统,所述多类型负载的资源统一调度系统包括:

终端设备、调度服务器和包括多个服务器的集群,其中,多个所述服务器包括第一目标服务器;

所述终端设备,用于向所述调度服务器发送不同类型的目标作业负载类型的应用任务,以及各个所述目标作业负载类型的应用任务对应的资源需求信息,其中,所述不同类型的目标作业负载类型的应用任务至少包括图形应用会话任务;

所述调度服务器,用于根据当前目标作业负载类型的应用任务对应的资源需求信息,在集群资源信息表中查找满足所述当前目标作业负载类型的应用任务对应的资源需求信息的候选服务器;所述集群资源信息表中包括:集群中,各个服务器的属性信息、已用资源信息和空闲资源信息;

所述调度服务器,用于在所述候选服务器中选择负载最低的服务器作为第一目标服务器,并将所述当前目标作业负载类型的应用任务和所述当前目标作业负载类型的应用任务对应的资源需求信息发送给所述第一目标服务器;

所述第一目标服务器,用于在接收到所述当前目标作业负载类型的应用任务和所述当前目标作业负载类型的应用任务对应的资源需求信息后,根据所述当前目标作业负载类型的应用任务对应的资源需求信息运行所述当前目标作业负载类型的应用任务。

[0008] 根据本申请提供的具体实施例,本申请公开了以下技术效果:

本申请提供了一种多类型负载的资源统一调度方法及系统,所述方法包括:接收不同类型的目标作业负载类型的应用任务,以及各个所述目标作业负载类型的应用任务对应的资源需求信息,其中,所述不同类型的目标作业负载类型的应用任务至少包括图形应用会话任务;根据当前目标作业负载类型的应用任务对应的资源需求信息,在集群资源信息表中查找满足所述当前目标作业负载类型的应用任务对应的资源需求信息的候选服务器;所述集群资源信息表中包括:集群中,各个服务器的属性信息、已用资源信息和空闲资源信息;在所述候选服务器中选择负载最低的服务器作为第一目标服务器;将所述当前目标作业负载类型的应用任务和所述当前目标作业负载类型的应用任务对应的资源需求信息发送给所述第一目标服务器,以使所述第一目标服务器运行所述当前目标作业负载类型的应用任务。这样,在为终端设备分配服务器时,调度服务器可以直接从集群资源信息表查找可以满足目标作业的资源需求信息服务器,这样终端设备就可以直接与该目标服务器连接,以让该目标服务器为终端设备处理目标作业,而无需像相关技术中,需要通过终端设备逐一去尝试与集群中的每一个服务器进行连接,直至成功建立连接,从而提升了终端设备连接服务器的效率,再者,调度服务器选择的服务器是集群中负载最低的服务器,从而可以提升集群中服务器资源的整体利用率。

附图说明

[0009] 为了更清楚地说明本申请实施例或现有技术中的技术方案,下面将对实施例中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本申请的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0010] 图1为本申请一实施例示出的多类型负载的资源统一调度方法的流程图;

图2为本申请一实施例示出的相关技术中启动图形应用会话任务的流程图;

图3为本申请一实施例示出的本公开中启动图形应用会话任务的流程图;

图4为本申请一实施例提供的多类型负载的资源统一调度系统功能模块架构图;

图5为本申请一实施例提供的一种多类型负载的资源统一调度装置的功能模块示意图;

图6为本申请一实施例提供的一种计算机设备的结构示意图。

具体实施方式

[0011] 下面将结合本申请实施例中的附图,对本申请实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本申请一部分实施例,而不是全部的实施例。基于本申请中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本申请保护的范围。

[0012] 使本申请的上述目的、特征和优点能够更加明显易懂,下面结合附图和具体实施方式对本申请作进一步详细的说明。

[0013] 图1为本申请一实施例示出的多类型负载的资源统一调度方法的流程图,该多类型负载的资源统一调度方法包括以下步骤S101-S102:

在步骤S101中,接收不同类型的目标作业负载类型的应用任务,以及各个目标作业负载类型的应用任务对应的资源需求信息,其中,不同类型的目标作业负载类型的应用任务至少包括图形应用会话任务。

[0014] 其中,图形应用会话任务包括数字孪生类应用及图形渲染类应用。

[0015] 目标作业负载类型的应用任务还可以包括:HPC并行计算任务、批处理计算任务、AI模型训练任务和AI模型推理任务。

[0016] 资源需求信息包括:CPU数量、内存大小、操作系统版本、GPU的型号、GPU数量、GPU模式和图形渲染库类型。

[0017] 也即,目标作业所需资源信息除了常见的CPU数量、内存大小和操作系统版本之外,还包括GPU的型号、GPU数量(分为流处理器数量和显存数量)、GPU模式(渲染/计算/平衡模式),以及图形渲染库类型,如DirectX、OpenGL、Vulkan。

[0018] 本公开给出几个应用任务以及应用任务对应的资源需求信息的例子,Siemens UG NX,Unreal Engine,PyTorch模型训练,matlab并行计算。

[0019] 1)应用任务为图形应用会话任务(图形渲染类),以工业设计应用Siemens UG NX的图形会话为例,其对应的资源需求信息如下:

CPU型号:“x86, amd64”;

CPU核数:4;

内存数量:32GB;

操作系统:“Windows Server 2016”;

GPU型号:“Quadro M4000”;

GPU数量:“gmem=8GB, cuda_core=1000”;

GPU模式:“渲染”;

渲染库类型:“DirectX 12”。

[0020] 2)应用任务为图形应用会话任务(数字孪生类),以游戏引擎Unreal Engine的图形会话为例,其对应的资源需求信息如下:

CPU型号:“x86, amd64”;

CPU核数:8;

内存数量:128GB;

操作系统:“Ubuntu 22.04”;

GPU型号:“RTX A40”;

GPU数量:“gmem=48GB, cuda_core=10000”;

GPU模式:“渲染”;

渲染库类型:“Vulkan 1.3”。

[0021] 3)应用任务为并行计算应用,以气候预测的科学计算应用WRF为例,其对应的资源需求信息如下:

CPU型号:“aarch64”;

CPU核数:256;

内存数量:512GB;

操作系统:“Kylin Linux Server V10 SP1”。

[0022] 4)应用任务为批处理计算应用,以芯片设计的EDA应用Cadence Virtuoso为例,其对应的资源需求信息如下:

CPU型号:“x86, amd64”;

CPU核数:64;

内存数量:256GB;

操作系统:“CentOS 7”;

GPU型号:“RTX A4000”;

GPU数量:“gmem=4GB, cuda_core=500”;

GPU模式:“平衡”;

渲染库类型:“OpenGL 2.1”。

[0023] 5)应用任务为AI模型训练,以大模型训练框架Deepspeed为例,其对应的资源需求信息如下:

CPU型号:“aarch64”;

CPU核数:8;

内存数量:1024GB;

操作系统:“Ubuntu 22.04”;

GPU型号:“NVIDIA A800”;

GPU数量:“gmem=640GB, cuda_core=100000”;

GPU模式:“计算”。

[0024] 6)应用任务为AI模型推理,以深度学习框架Pytorch为例,其对应的资源需求信息如下:

CPU型号:“aarch64”;

CPU核数:4;

内存数量:128GB;

操作系统:“Ubuntu 20.04”;

GPU型号:“Ascend 310b”;

GPU数量:“gmem=10GB, cuda_core=2000”;

GPU模式:“计算”。

[0025] 现有的HPC集群资源调度软件,比如IBM Spectrum LSF和Slurm调度软件,只能调度并行计算任务和批处理计算任务,不能调度数字孪生类应用任务、图形渲染类应用任务、AI模型训练任务或者AI模型推理任务。容器集群最常用的K8S开源管理软件只能按pod来编排和调度管理经过容器封装的计算任务和AI模型训练任务或者AI模型推理任务,不能调度HPC并行计算作业任务,更不能调度管理数字孪生应用任务和图形渲染类应用任务。云桌面系统软件,比如世界最领先的Citrix和VMware中,只有云桌面会话和远程图形应用会话的管理功能,没有按GPU资源来调度云桌面或图形应用会话任务,更不支持其他类型的计算任务或AI模型训练推理任务的管理和资源调度。本公开可以在公有云和私有云的服务器集群中统一调度Windows和Linux数字孪生类应用、图形渲染类应用和AI模型训练任务及AI模型推理任务,还有传统的并行计算任务和批处理计算任务,进一步的,本公开可以将云端集群中运行的数字孪生类应用任务或图形渲染类应用任务作为一种GPU资源的负载任务抽象出来,作为远程图形会话作业,跟其他负载类型一起进行调度和管理。

[0026] 在本公开中,将数字孪生类应用任务及图形渲染类应用任务、AI模型训练任务、AI模型推理任务、HPC并行计算任务、批处理计算任务都抽象为“作业”对象,终端设备使用调度服务器中调度系统的命令行或者通过图形界面调用调度系统的API,来向调度服务器提交一个作业,通过命令行参数或者作业描述文件说明目标作业的负载类型和目标作业所需资源信息,可以包括:操作系统版本,该图形应用需要的资源,主要是GPU型号、GPU数量、GPU模式和图形渲染库类型和版本。

[0027] 例如:用户可以通过调度系统的命令行jsub来提交作业,将作业对象描述文件作为命令行参数提交给调度系统:"jsub-jobfile myjob.json";

在myjob.json中说明该作业的资源需求和应用启动路径及参数,如:

App_start_cmd="/opt/unreal-engine/Engine/Binaries/Linux/UnrealEditor -graphic"。

[0028] 在步骤S102中,根据当前目标作业负载类型的应用任务对应的资源需求信息,在集群资源信息表中查找满足当前目标作业负载类型的应用任务对应的资源需求信息的候选服务器;集群资源信息表中包括:集群中,各个服务器的属性信息、已用资源信息和空闲资源信息。

[0029] 其中,在集群资源信息表中查找满足当前目标作业负载类型的应用任务对应的资源需求信息的候选服务器可以理解为:为每一个应用任务找到适合的候选服务器,这里的候选服务器可以同时服务于多个应用任务,在服务于多个应用任务时,候选服务器的空闲资源满足该些应用任务的资源需求信息的要求。

[0030] 服务器的属性信息包括:操作系统版本、GPU的型号、GPU模式(渲染/计算/平衡模式)和图形渲染库类型,如DirectX、OpenGL、Vulkan;各个服务器的已用资源信息和各个服务器的空闲资源信息可以包括:已使用的CPU数量、已使用的内存大小、已使用的GPU数量,空闲的CPU数量、空闲的内存大小和空闲的GPU数量。

[0031] 在步骤S103中,在候选服务器中选择负载最低的服务器作为第一目标服务器。

[0032] 为了均衡集群中的各个服务器的利用率,不至于使得某些服务器过载而某些服务

器空闲,本公开中会在候选服务器中选择负载最低的服务器作为第一目标服务器。

[0033] 在步骤S104中,将当前目标作业负载类型的应用任务和当前目标作业负载类型的应用任务对应的资源需求信息发送给第一目标服务器,以使第一目标服务器运行当前目标作业负载类型的应用任务。

[0034] 在为当前的应用任务选择了第一目标服务器后,便可以将当前目标作业负载类型的应用任务和当前目标作业负载类型的应用任务对应的资源需求信息发送给第一目标服务器,这样,第一目标服务器就可以基于接收到的信息去运行当前目标作业负载类型的应用任务。

[0035] 以当前目标作业负载类型的应用任务为图形应用会话任务为例,第一目标服务器为图形服务器为例,调度服务器根据图形应用的资源需求信息和集群中各图形服务器的资源负载情况(上述的集群资源信息表),如果当前有符合图形应用会话任务的资源需求信息要求的候选服务器,便从该些候选服务器中选择负载最低的图形服务器作为第一目标服务器,调度服务器就根据作业描述文件中的应用启动命令行及参数,通过其在该图形服务器上的作业启动程序立即在服务器上启动这个图形应用会话任务,并将图形服务器IP地址和Linux图形DISPLAY号或Windows远程会话号返回给当前目标作业负载类型的应用任务对应的终端设备,供终端设备连接使用;如果当前没有空闲的图形服务器,那么该图形应用会话任务将在调度队列中排队,等待资源满足后再启动。也即,在本公开中,如果调度服务器中的作业较多,此时,终端设备提交的应用任务可以先放置在调度服务器的队列中,然后由调度服务器来调度和派发。

[0036] 其中,目标作业在服务器资源不足时可以在调度服务器中排队等待,不会因为争用相同的服务器资源而启动失败或运行缓慢。

[0037] 其中,在图形服务器上启动图形应用会话任务时,根据图形应用会话任务所需的图形渲染库类型和版本,预先安装或通过系统环境变量及Windows注册表将图形渲染库切换到应用需要的版本,并启动图形应用会话任务对应的图形应用;在启动图形应用前,根据调度分配的具体GPU卡号和CPU核数,会通过系统API和底层资源隔离机制将图形应用会话任务的进程树绑定到分配的GPU卡和CPU核上面去,比如linux系统的cgroup,Windows系统的job object。

[0038] 相关技术中,如图2所示,用户在终端设备中启动一个数字孪生类应用或者图形渲染类应用时,都是登录到图形服务器以获取图形界面或远程图形会话,然后在图形桌面启动图形应用,如果该服务器被其他终端设备占用,那么就需要重新去登录其他的终端设备,比如:有10台服务器,终端设备去登录其中的第4台服务器,如果第4台服务器被其他终端设备所占用,那么就会导致登录失败,终端设备需要去尝试着连接其他服务器。

[0039] 而本公开中,将数字孪生类应用及图形渲染类应用、AI模型训练及推理任务、HPC并行计算任务、批处理计算任务都抽象为“作业”对象,以终端设备使用调度服务器中调度系统的命令行或者通过图形界面调用调度系统的API,来向调度服务器提交一个作业,通过命令行参数或者作业描述文件说明目标作业的负载类型和目标作业所需资源信息,可以包括:操作系统版本,该图形应用会话任务需要的资源,主要是GPU型号、GPU数量、GPU模式和图形渲染库类型和版本。

[0040] 例如:用户可以通过调度系统的命令行jsub来提交作业,将作业对象描述文件作

为命令行参数提交给调度系统:"jsub-jobfile myjob.json";

在myjob.json中说明该作业的资源需求和应用启动路径及参数,如:

App_start_cmd="/opt/unreal-engine/Engine/Binaries/Linux/UnrealEditor -graphic"。

[0041] 如图3所示,以目标作业的负载类型为图形应用会话任务为例,终端设备通过命令行或者API向调度服务器提交图形应用会话任务,调度服务器调度分配图形服务器上的资源,然后调度服务器启动图形应用会话任务,并返回图形界面或远程图形会话。

[0042] 本申请提供了一种多类型负载的资源统一调度方法,包括:接收不同类型的目标作业负载类型的应用任务,以及各个目标作业负载类型的应用任务对应的资源需求信息,其中,不同类型的目标作业负载类型的应用任务至少包括图形应用会话任务;根据当前目标作业负载类型的应用任务对应的资源需求信息,在集群资源信息表中查找满足当前目标作业负载类型的应用任务对应的资源需求信息的候选服务器;集群资源信息表中包括:集群中,各个服务器的属性信息、已用资源信息和空闲资源信息;在候选服务器中选择负载最低的服务器作为第一目标服务器;将当前目标作业负载类型的应用任务和当前目标作业负载类型的应用任务对应的资源需求信息发送给第一目标服务器,以使第一目标服务器运行当前目标作业负载类型的应用任务。这样,在为终端设备分配服务器时,调度服务器可以直接从集群资源信息表查找可以满足目标作业的资源需求信息服务器,这样终端设备就可以直接与该目标服务器连接,以让该目标服务器为终端设备处理目标作业,而无需像相关技术中,需要通过终端设备逐一去尝试与集群中的每一个服务器进行连接,直至成功建立连接,从而提升了终端设备连接服务器的效率,再者,调度服务器选择的服务器是集群中负载最低的服务器,从而可以提升集群中服务器资源的整体利用率。

[0043] 针对跨类型应用,尤其是图形渲染类应用任务而设计了新的调度策略。为了最大化利用集群中的服务器资源,并保证新方式下图形渲染类应用任务的用户使用体验,本公开在调度系统中增加了互补调度策略、时间窗口调度策略、资源抢占式调度策略、以及会话重用调度策略,以下详细介绍。

[0044] 一、互补调度策略

当前目标作业负载类型的应用任务为接收的不同类型的目标作业负载类型的应用任务中当前进行处理的应用任务,可以为一个应用任务,也可以为多个应用任务。

[0045] 以当前目标作业负载类型的应用任务为多个应用任务为例进行说明。

[0046] 在一种可实现方式中,在当前目标作业负载类型的应用任务分配的第一目标服务器为正在运行应用任务但有空闲资源的服务器,此时服务器的空闲资源可以满足当前目标作业负载类型的应用任务对应的资源需求信息,且第一目标服务器正在运行的应用任务与当前目标作业负载类型的应用任务不相同,在运行了当前目标作业负载类型的应用任务后,第一目标服务器中运行了不同类型的应用任务。

[0047] 在获取到第一目标服务器有剩余资源时,可以从接收到的各个目标作业负载类型的应用任务对应的资源需求信息中获取多个目标资源需求信息,其中,多个目标资源需求信息之和小于或等于第一目标服务器有剩余资源,且第一目标服务器中正在运行的应用任务的类型要与各个目标资源需求信息对应的应用任务的类型不相同,这样就可以将多个目标资源需求信息以及各个目标资源需求信息对应的应用任务发送给第一目标服务器,以让

第一目标服务器同时运行该些应用任务。

[0048] 在另一种可实现方式中,在检测到集群中有空闲服务器时,可以从接收到的各个目标作业负载类型的应用任务对应的资源需求信息中获取多个目标资源需求信息,其中,多个目标资源需求信息之和小于或等于第一目标服务器的资源,且各个目标资源需求信息对应的应用任务的类型不相同,这样就可以将多个目标资源需求信息以及各个目标资源需求信息对应的应用任务发送给第一目标服务器,以让第一目标服务器同时运行该些应用任务。

[0049] 具体的,目前存在的作业调度系统一般都是按作业排队顺序将作业逐个派发到服务器上去执行,作业批量派发时也是将同一应用类型的一批作业派发出去,没有考虑过将不同应用类型的作业按资源互补而组合起来批量派发。本发明提出的互补调度策略,当调度队列中有不同应用类型的负载作业排队,而且资源互补时,调度程序可以根据服务器资源配置,将2个或多个完全不同应用类型的作业一次性派发到同一个服务器上面去执行。例如,集群中空闲出了一台服务器A,该服务器配置了32核CPU和2块GPU卡;在排队队列中有一个并行计算作业需要24核CPU,还有一个图形类作业需要2核CPU核+1块GPU卡,恰好还有一个AI模型推理作业需要6核CPU+1块GPU卡;那么调度系统将为这3个作业“拼车”,一次性派发这3个作业到服务器A上启动,保证服务器A的资源最大化利用,实现集群服务器的“一机多用”。

[0050] 由于某些图形应用会话任务重度使用GPU进行图形渲染,而只用很少量的CPU资源,按传统方法分配的物理服务器或者虚拟机,CPU计算资源往往被浪费。本公开可以达到“一机多用”的目的,同一台服务器上可以同时混合运行并行计算任务、批处理计算任务、图形应用会话任务,以及AI模型训练及推理任务,充分利用服务器的CPU和GPU资源。

[0051] 目前存在的作业调度系统一般都是按作业排队顺序将作业逐个派发到服务器上去执行,作业批量派发时也是将同一应用类型的一批作业派发出去,没有考虑过将不同应用类型的作业按资源互补而组合起来批量派发,而本公开中提出了互补调度策略,在互补调度策略中,可以几个终端设备同时使用集群中的同一个服务器,比如:第一目标服务器已经为其他终端设备进行作业,但第一目标服务器还有空闲资源,而该空闲资源可以满足目标作业的资源需求信息的要求,且第一目标服务器已经进行的作业的负载类型与目标作业的负载类型各不相同,那么它们就可以共用这一台第一目标服务器,这样就可以最大化利用第一目标服务器的资源。

[0052] 二、时间窗口调度策略

在一个实施例中,上述步骤S102包括以下子步骤A1-A3:

A1、获取当前目标作业负载类型的应用任务对应的运行时间,将其放入对应的目标时间窗口中,其中,不同的时间窗口对应不同的运行时间段,且不同类型的目标作业负载类型的应用任务对应不同的运行时间段。

[0053] A2、若当前时间满足目标时间窗口对应的时间段要求,则根据当前目标作业负载类型的应用任务对应的资源需求信息,在集群资源信息表中查找满足当前目标作业负载类型的应用任务对应的资源需求信息的候选服务器。

[0054] A3、若当前时间不满足目标时间窗口对应的时间段要求,向当前目标作业负载类型的应用任务对应的终端设备发送拒绝处理指令,拒绝处理指令指示当前时间不满足当前

目标作业负载类型的应用任务对应的运行时间;或者,将当前目标作业负载类型的应用任务和当前目标作业负载类型的应用任务对应的资源需求信息放入对应的等待队列中排队,并在检测到满足目标时间窗口对应的时间段要求时,根据当前目标作业负载类型的应用任务对应的资源需求信息,在集群资源信息表中重新查找满足当前目标作业负载类型的应用任务对应的资源需求信息的候选服务器。

[0055] 在调度服务器中设置若干时间窗口,比如白天工作时段对应的时间窗口、夜晚时段对应的时间窗口及周末休息时段对应的时间窗口。在调度服务器中为每一类应用任务指定运行时间窗口,例如,将需要用户交互操作的图形应用会话任务作业分配到白天工作时段对应的时间窗口,而把批处理计算作业或者AI模型训练作业分配到夜晚时段对应的时间窗口或周末休息时段对应的时间窗口。这种策略可以称之为时间窗口调度策略。当白天工作时段结束时,调度服务器将把在服务器上运行的图形应用环境保存检查点并清理释放资源,自动准备好批处理计算和AI模型训练所需要的运行环境,供计算作业夜间使用资源;当夜晚时段对应的时间窗口或周末休息时段对应的时间窗口结束时,又把计算作业环境保存并清理释放资源,将服务器自动恢复成图形应用环境,供上班人员交互使用。通过时间窗口调度策略可以实现图形应用会话任务和批处理计算任务在相同服务器的分时使用,充分利用服务器上的GPU资源和CPU资源,达到“一机两用”的效果。

[0056] 数字孪生类应用及图形渲染类应用一般是在用户交互操作时才大量使用GPU资源,当晚上用户下班后GPU资源往往就闲置,而无法让AI模型训练及推理任务使用。本公开中的“一机两用”,白天将GPU卡设置为渲染模式,供用户交互使用数字孪生类应用及图形渲染类应用,而在晚上自动将GPU卡切换成计算模式,供AI模型训练及推理任务使用。

[0057] 具体的,在为目标作业分配服务器时,首先可以获取当前目标作业负载类型的应用任务对应的运行时间,并将其放入对应的目标时间窗口中;然后检查当前时间是否满足目标时间窗口对应的时间段要求,若当前时间满足时间窗口对应的时间段要求,则根据当前目标作业负载类型的应用任务对应的资源需求信息,在集群资源信息表中查找满足当前目标作业负载类型的应用任务对应的资源需求信息的候选服务器;若当前时间不满足目标时间窗口对应的时间段要求,向当前目标作业负载类型的应用任务对应的终端设备发送拒绝处理指令,拒绝处理指令指示当前时间不满足当前目标作业负载类型的应用任务对应的运行时间;或者,将当前目标作业负载类型的应用任务和当前目标作业负载类型的应用任务对应的资源需求信息放入对应的等待队列中排队,并在检测到满足目标时间窗口对应的时间段要求时,根据当前目标作业负载类型的应用任务对应的资源需求信息,在集群资源信息表中重新查找满足当前目标作业负载类型的应用任务对应的资源需求信息的候选服务器,这样就可以实现“一机两用”的效果。

[0058] 三、资源抢占式调度策略

在一个实施例中,本公开中的多类型负载的资源统一调度方法,还包括以下子步骤B1- B5:

B1、若在集群资源信息表中未查找到满足当前目标作业负载类型的应用任务对应的资源需求信息的候选服务器,获取当前目标作业负载类型的应用任务的优先级。

[0059] B2、检测集群的各个服务器中是否有第二目标服务器,其中,第二目标服务器中正在运行的第一应用任务的负载类型的优先级低于当前目标作业负载类型的应用任务的优

先级,且第二目标服务器在释放了为正在运行的第一应用任务分配的资源后,第二目标服务器中的空闲资源满足当前目标作业负载类型的应用任务对应的资源需求信息。

[0060] B3、控制第二目标服务器将正在运行的第一应用任务挂起或终止,以使第二目标服务器释放为第一应用任务分配的资源。

[0061] B4、将当前目标作业负载类型的应用任务和当前目标作业负载类型的应用任务对应的资源需求信息发送给第二目标服务器,以让第二目标服务器运行当前目标作业负载类型的应用任务。

[0062] B5、在检测到当前目标作业负载类型的应用任务运行完成后,控制第二目标服务器继续运行第一应用任务。

[0063] 在实际使用场景中,并行计算和AI模型训练往往需要长时间持续使用CPU和GPU服务器资源,当用户急需资源使用某款图形应用会话任务或者使用AI模型进行推理时,需要立即获得所需资源并启动相关应用。调度服务器可以通过抢占式作业调度策略来应对上述场景。在调度服务器中可以为不同的负载类型配置不同的优先级,优先级越高,应用运行的紧急程度越高,当遇到优先级高的紧急任务时,调度服务器将在集群中找到可被抢占运行低优先级作业的服务器,并将该低优先级作业挂起或终止该低优先级作业以释放资源,供该紧急的图形应用会话任务使用。当该图形应用会话任务使用完毕后,调度系统将服务器环境自动还原,恢复低优先级作业从检查点继续运行。

[0064] 示例的,调度服务器可以配置多个排队队列,不同排队队列可以设置不同的优先级,当遇到紧急任务时,比如:图形应用会话任务,调度服务器可以将图形应用会话任务提交到高优先级的可抢占队列中,调度服务器将在集群中可被抢占的低优先级队列中找到符合要求的低优先级作业和服务器,并将该低优先级作业挂起或终止该低优先级作业以释放资源,供该紧急的图形应用会话任务使用。当该图形应用会话任务使用完毕后,调度系统将服务器环境自动还原,恢复低优先级作业从检查点继续运行。

[0065] 四、会话重用调度策略

在一个实施例中,在目标作业负载类型的应用任务为图形应用会话任务时,本公开中的多类型负载的资源统一调度方法,还包括以下子步骤C1- C2:

C1、若在集群资源信息表中未查找到满足当前目标作业负载类型的应用任务对应的资源需求信息的候选服务器,检测集群中是否有第三目标服务器,其中,第三目标服务器中正在运行与当前目标作业负载类型的应用任务属于同一终端设备的其他应用任务,其他应用任务的负载类型与当前目标作业负载类型相同;

C2、若有,将当前目标作业负载类型的应用任务和当前目标作业负载类型的应用任务对应的资源需求信息发送给第三目标服务器。

[0066] 当集群中各个服务器的GPU资源紧张时,为每个图形会话类任务创建一个单独的会话或分配一个单独的GPU显卡会加重集群的资源负担。为此,调度服务器可以提供一种会话重用的调度策略,即对同一个终端设备启动的多个同类型图形应用会话任务,调度分配到同一个会话作业中,使用同一张GPU显卡。因为同一个终端设备在交互操作过程中,同一时间一般只会操作其中一个应用,这样该终端设备的其他图形应用会话任务占用的GPU资源就比较低,不会影响当前操作的图形应用会话任务的交互性能,通过终端设备级别的会话重用调度策略,不会引起各个终端设备之间的资源冲突和协调工作,既保证了各个终端

设备之间的安全隔离,又充分利用了GPU资源。

[0067] 由于集群中各个服务器的资源是动态变化的,因此,还需要去更新集群资源信息表,此时,多类型负载的资源统一调度方法,还包括以下子步骤D1- D2:

D1、接收集群中的各个服务器发送的集群资源更新信息,集群资源更新信息中包括:当前时间点对应的各个服务器的已用资源信息和空闲资源信息。

[0068] D2、根据集群资源更新信息更新集群资源信息表。

[0069] 在集群中的各类服务器上识别当前时间点已用资源信息和空闲资源信息,上报给调度服务器,供调度服务器更新集群资源信息表所用,从而在调度服务器为终端设备的目标作业分配服务器时,可以更加准确。

[0070] 本公开还提供一种多类型负载的资源统一调度系统,多类型负载的资源统一调度系统包括:

终端设备、调度服务器和包括多个服务器的集群,其中,多个服务器包括第一目标服务器;

终端设备,用于向调度服务器发送不同类型的目标作业负载类型的应用任务,以及各个目标作业负载类型的应用任务对应的资源需求信息,其中,不同类型的目标作业负载类型的应用任务至少包括图形应用会话任务;

调度服务器,用于根据当前目标作业负载类型的应用任务对应的资源需求信息,在集群资源信息表中查找满足当前目标作业负载类型的应用任务对应的资源需求信息的候选服务器;集群资源信息表中包括:集群中,各个服务器的属性信息、已用资源信息和空闲资源信息;

调度服务器,用于在候选服务器中选择负载最低的服务器作为第一目标服务器,并将当前目标作业负载类型的应用任务和当前目标作业负载类型的应用任务对应的资源需求信息发送给第一目标服务器;

第一目标服务器,用于在接收到当前目标作业负载类型的应用任务和当前目标作业负载类型的应用任务对应的资源需求信息后,根据当前目标作业负载类型的应用任务对应的资源需求信息运行当前目标作业负载类型的应用任务。

[0071] 在一个实施例中,第一目标服务器为正在运行应用任务但有空闲资源的服务器,空闲资源可以满足当前目标作业负载类型的应用任务对应的资源需求信息,且第一目标服务器正在运行的应用任务与当前目标作业负载类型的应用任务不相同,在运行了当前目标作业负载类型的应用任务后,第一目标服务器中运行了不同类型的应用任务。

[0072] 在一个实施例中,调度服务器,具体用于:

获取当前目标作业负载类型的应用任务对应的运行时间,将其放入对应的目标时间窗口中,其中,不同的时间窗口对应不同的运行时间段,且不同类型的目标作业负载类型的应用任务对应不同的运行时间段;

若当前时间满足目标时间窗口对应的时间段要求,则根据当前目标作业负载类型的应用任务对应的资源需求信息,在集群资源信息表中查找满足当前目标作业负载类型的应用任务对应的资源需求信息的候选服务器;

若当前时间不满足目标时间窗口对应的时间段要求,向当前目标作业负载类型的应用任务对应的终端设备发送拒绝处理指令,拒绝处理指令指示当前时间不满足当前目标

作业负载类型的应用任务对应的运行时间;或者,将当前目标作业负载类型的应用任务和当前目标作业负载类型的应用任务对应的资源需求信息存入缓存中,并在检测到满足目标时间窗口对应的时间段要求时,根据当前目标作业负载类型的应用任务对应的资源需求信息,在集群资源信息表中重新查找满足当前目标作业负载类型的应用任务对应的资源需求信息的候选服务器。

[0073] 在一个实施例中,调度服务器,还用于:若在集群资源信息表中未查找到满足当前目标作业负载类型的应用任务对应的资源需求信息的候选服务器,获取当前目标作业负载类型的应用任务的优先级;检测集群的各个服务器中是否有第二目标服务器,其中,第二目标服务器中正在运行的第一应用任务的负载类型的优先级低于当前目标作业负载类型的应用任务的优先级,且第二目标服务器在释放了为正在运行的第一应用任务分配的资源后,第二目标服务器中的空闲资源满足当前目标作业负载类型的应用任务对应的资源需求信息;控制第二目标服务器将正在运行的第一应用任务挂起或终止,以使第二目标服务器释放为第一应用任务分配的资源;将当前目标作业负载类型的应用任务和当前目标作业负载类型的应用任务对应的资源需求信息发送给第二目标服务器,以让第二目标服务器运行当前目标作业负载类型的应用任务;在检测到当前目标作业负载类型的应用任务运行完成后,控制第二目标服务器继续运行第一应用任务;

或者,

调度服务器,还用于:在目标作业负载类型的应用任务为图形应用会话任务时,若在集群资源信息表中未查找到满足当前目标作业负载类型的应用任务对应的资源需求信息的候选服务器,检测集群中是否有第三目标服务器,其中,第三目标服务器中正在运行与当前目标作业负载类型的应用任务属于同一终端设备的其他应用任务,其他应用任务的负载类型与当前目标作业负载类型相同; 若有,将当前目标作业负载类型的应用任务和当前目标作业负载类型的应用任务对应的资源需求信息发送给第三目标服务器。

[0074] 在现有的公有云和私有云中,用户使用数字孪生及图形渲染类应用时,所需的图形服务器资源往往按GPU服务器整机分配,或者使用虚拟化技术将服务器隔离成多个虚拟机,利用GPU穿透或者vGPU技术将Windows和linux虚拟机固定分配给不同的用户来交互使用。这样就容易造成服务器的GPU资源利用率低下的问题,而且图形类应用的交互性能也受到虚拟化的影响而大幅损耗。

[0075] 某些数字孪生及图形类应用重度使用GPU进行图形渲染,而只用很少量的CPU资源,按传统方法分配的物理服务器或者虚拟机,CPU计算资源往往被浪费。

[0076] 有些公有云和私有云通过使用容器化技术和Kubernetes(K8S)来解决Linux应用和资源的动态匹配问题,但数字孪生及图形渲染类应用多数使用Windows操作系统,而Windows容器中不支持图形界面的应用,这样Windows图形类应用就只能使用集群中的Windows物理服务器或者云桌面虚拟机。

[0077] 数字孪生及图形渲染类应用一般是在用户交互操作时才大量使用GPU资源,当晚上用户下班后GPU资源往往就闲置,而无法让AI模型训练及推理任务使用。

[0078] 现有的HPC集群资源调度软件,比如IBM Spectrum LSF和Slurm调度软件,只能调度并行计算和批处理计算任务,不能调度数字孪生应用、图形渲染应用或者AI模型训练推理任务。容器集群最常用的K8S开源管理软件只能按pod来编排和调度管理经过容器封装的

计算任务和AI模型训练推理任务,不能调度HPC并行计算作业,更不能调度管理数字孪生和图形应用。云桌面系统软件,比如世界最领先的Citrix和VMware中,只有云桌面会话和远程图形应用会话的管理功能,没有按GPU资源来调度云桌面或图形应用会话,更不支持其他类型的计算任务或AI模型训练推理任务的管理和资源调度。

[0079] 本公开的效果和优点:

1.本公开可以在公有云和私有云的服务器集群中统一调度Windows和Linux数字孪生应用、图形渲染应用和AI模型训练及推理任务,还有传统的并行计算任务和批处理计算任务。

[0080] 2.将云端集群中运行的数字孪生应用或图形渲染应用作为一种GPU资源的负载任务抽象出来,作为远程图形会话作业,跟其他类型的计算作业一起进行调度和管理。

[0081] 3.图形类应用的作业在服务器GPU资源不足时可以在资源调度系统中排队等待,不会因为争用相同的GPU资源而启动失败或运行缓慢。

[0082] 4.不依赖虚拟化技术,数字孪生和图形渲染应用可以和AI模型训练及推理任务共享使用同一块GPU卡,性能更好。

[0083] 5.“一机多用”,同一台服务器上可以同时混合运行并行计算任务、批处理计算任务、数字孪生和图形渲染应用,以及AI模型训练及推理任务,充分利用服务器的CPU和GPU资源。

[0084] 6.“一机两用”,白天将GPU卡设置为渲染模式,供用户交互使用数字孪生及图形渲染类应用,而在晚上自动将GPU卡切换成计算模式,供AI模型训练及推理任务使用。

[0085] 图4为本申请一实施例提供的多类型负载的资源统一调度系统的功能模块架构图,如图4所示,包括:终端设备、调度服务器和包括多个服务器的集群,其中,集群中的服务器可以为应用/计算服务器,多个服务器包括第一目标服务器;

其中,调度命令行程序为一组命令行程序,包括作业提交、作业查询、作业控制等命令行程序。终端设备可以在操作系统中手工调用这些命令行进行作业提交、作业查询和作业控制操作,命令行程序给终端设备返回作业计算结果或者图形应用的远程界面连接信息。终端设备提交作业前,将该应用程序需要的资源类别、资源数量、渲染库类型版本、应用程序启动路径和参数写在一个作业描述文件中,在提交作业时指定作业描述文件。调度命令行程序一般运行在终端设备上,也可以运行在调度服务器上。

[0086] 调度服务器中的调度引擎程序,用于接受调度命令行提交的作业提交、作业查询、作业控制等请求,将作业数据或图形界面连接信息返回给调度命令行程序。调度引擎程序内部有配置文件,管理员可以预先设置调度策略相关的参数,例如:互补调度策略开关参数、每一类负载类型的时间窗口设置、抢占式调度策略的排队队列,或者会话重用调度测试开关等。调度引擎程序定时(比如每隔15秒)接收每个应用/计算服务器上的资源监控程序上报的当前时间点对应的已用资源信息和空闲资源信息,根据目标作业的资源需求信息、当前集群中各个应用/计算服务器的资源状况、以及管理员配置的调度策略,将目标作业调度到合适的应用/计算服务器上。调度引擎程序将根据该应用/计算服务器的资源设备情况为目标作业指定服务器上可用的CPU、GPU设备号和资源数量,这些信息将随同作业对象通过网络通讯发送给应用/计算服务器上的作业启动程序,实现目标作业的派发。调度引擎程序运行在调度服务器上。

[0087] 资源监控程序运行在每一个应用/计算服务器上,通过操作系统命令、系统API和IPMI等各种方法对所在的应用/计算服务器系统进行扫描,自动识别应用/计算服务器上的CPU、内存和GPU等设备和资源使用情况,以及应用/计算服务器上运行的各类作业实际使用的资源数量(也即上述实施例中的各个应用/计算服务器的属性信息、已用资源信息和空闲资源信息),通过网络通讯定时上报给调度服务器上的调度引擎程序。

[0088] 作业启动程序运行在每一个应用/计算服务器上,当收到调度引擎程序通过网络通讯派发来的作业对象信息(上述实施例中的目标作业的资源需求信息)后,根据作业对象信息中的渲染库类型及版本要求,通过设置操作系统的环境变量和注册表的方式,将当前程序运行环境的渲染库切换成作业要求的类型和版本,如果切换过程中系统因为缺少相关依赖库而报错,作业启动程序会自动从网上下载安装。运行环境准备完毕后,作业启动程序将执行作业对象中的应用启动路径和启动参数来启动目标作业,并调用操作系统的API来设置应用程序的进程属性,从而按作业对象中的资源配额来限制该目标作业能够使用的CPU、内存、GPU等设备和数量。目标作业启动后,作业启动程序将对每个作业的进程树进行跟踪扫描,后续可以根据调度引擎程序发来的作业控制指令来挂起、恢复或终止作业进程。

[0089] 相关技术中,有些公有云和私有云通过使用容器化技术和Kubernetes(K8S)来解决Linux应用和资源的动态匹配问题,但数字孪生类应用及图形渲染类应用多数使用Windows操作系统,而Windows容器中不支持图形界面的应用,这样Windows图形类应用就只能使用集群中的Windows物理服务器或者云桌面虚拟机。而本公开不依赖虚拟化技术,数字孪生类应用和图形渲染类应用可以和AI模型训练及推理任务共享使用同一块GPU卡,性能更好。

[0090] 基于同样的发明构思,本申请实施例还提供了一种用于实现上述所涉及的多类型负载的资源统一调度方法的多类型负载的资源统一调度装置。该装置所提供的解决问题的实现方案与上述方法中所记载的实现方案相似,故下面所提供的一个或多个多类型负载的资源统一调度装置实施例中的具体限定可以参见上文中对于多类型负载的资源统一调度方法的限定,在此不再赘述。

[0091] 在一个示例性的实施例中,如图5所示,提供了一种多类型负载的资源统一调度装置,所述多类型负载的资源统一调度装置包括:

接收模块11,用于接收不同类型的目标作业负载类型的应用任务,以及各个所述目标作业负载类型的应用任务对应的资源需求信息,其中,所述不同类型的目标作业负载类型的应用任务至少包括图形应用会话任务;

查找模块12,用于根据当前目标作业负载类型的应用任务对应的资源需求信息,在集群资源信息表中查找满足所述当前目标作业负载类型的应用任务对应的资源需求信息的候选服务器;所述集群资源信息表中包括:集群中,各个服务器的属性信息、已用资源信息和空闲资源信息;

选择模块13,用于在所述候选服务器中选择负载最低的服务器作为第一目标服务器;

发送模块14,用于将所述当前目标作业负载类型的应用任务和所述当前目标作业负载类型的应用任务对应的资源需求信息发送给所述第一目标服务器,以使所述第一目标服务器运行所述当前目标作业负载类型的应用任务。

[0092] 在一示例性的实施例中,提供了一种计算机设备,该计算机设备可以是服务器或者终端,其内部结构图可以如图6所示。该计算机设备包括处理器、存储器、输入/输出接口(Input/Output,简称I/O)和通信接口。其中,处理器、存储器和输入/输出接口通过系统总线连接,通信接口通过输入/输出接口连接到系统总线。其中,该计算机设备的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质和内存储器。该非易失性存储介质存储有操作系统、计算机程序和数据库。该内存储器为非易失性存储介质中的操作系统和计算机程序的运行提供环境。该计算机设备的数据库用于存储多类型负载的资源统一调度所需的数据。该计算机设备的输入/输出接口用于处理器与外部设备之间交换信息。该计算机设备的通信接口用于与外部的终端通过网络连接通信。该计算机程序被处理器执行时以实现一种多类型负载的资源统一调度方法。

[0093] 本领域技术人员可以理解,图6中示出的结构,仅仅是与本申请方案相关的部分结构的框图,并不构成对本申请方案所应用于其上的计算机设备的限定,具体的计算机设备可以包括比图中所示更多或更少的部件,或者组合某些部件,或者具有不同的部件布置。

[0094] 在一个示例性的实施例中,还提供了一种计算机设备,包括存储器和处理器,存储器中存储有计算机程序,该处理器执行计算机程序时实现上述各方法实施例中的步骤。

[0095] 在一个示例性的实施例中,提供了一种计算机可读存储介质,存储有计算机程序,该计算机程序被处理器执行时实现上述各方法实施例中的步骤。

[0096] 在一个示例性的实施例中,提供了一种计算机程序产品,包括计算机程序,该计算机程序被处理器执行时实现上述各方法实施例中的步骤。

[0097] 需要说明的是,本申请所涉及的用户信息(包括但不限于用户设备信息、用户个人信息等)和数据(包括但不限于用于分析的数据、存储的数据、展示的数据等),均为经用户授权或者经过各方充分授权的信息和数据,且相关数据的收集、使用和处理需要符合相关规定。

[0098] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程,是可以通过计算机程序来指令相关的硬件来完成,所述的计算机程序可存储于一非易失性计算机可读取存储介质中,该计算机程序在执行时,可包括如上述各方法的实施例的流程。其中,本申请所提供的各实施例中所使用的对存储器、数据库或其它介质的任何引用,均可包括非易失性和易失性存储器中的至少一种。非易失性存储器可包括只读存储器(Read-OnlyMemory,ROM)、磁带、软盘、闪存、光存储器、高密度嵌入式非易失性存储器、阻变存储器(ReRAM)、磁变存储器(Magnetoresistive Random Access Memory,MRAM)、铁电存储器(Ferroelectric Random Access Memory,FRAM)、相变存储器(Phase Change Memory,PCM)、石墨烯存储器等。易失性存储器可包括随机存取存储器(Random Access Memory,RAM)或外部高速缓冲存储器等。作为说明而非局限,RAM可以是多种形式,比如静态随机存取存储器(Static Random Access Memory,SRAM)或动态随机存取存储器(Dynamic RandomAccess Memory,DRAM)等。

[0099] 本申请所提供的各实施例中所涉及的数据库可包括关系型数据库和非关系型数据库中至少一种。非关系型数据库可包括基于区块链的分布式数据库等,不限于此。本申请所提供的各实施例中所涉及的处理器可为通用处理器、中央处理器、图形处理器、数字信号处理器、可编程逻辑器、基于量子计算的数据处理逻辑器等,不限于此。

[0100] 以上实施例的各技术特征可以进行任意的组合,为使描述简洁,未对上述实施例中的各个技术特征所有可能的组合都进行描述,然而,只要这些技术特征的组合不存在矛盾,都应当认为是本说明书记载的范围。

[0101] 本文中应用了具体个例对本申请的原理及实施方式进行了阐述,以上实施例的说明只是用于帮助理解本申请的方法及其核心思想;同时,对于本领域的一般技术人员,依据本申请的思想,在具体实施方式及应用范围上均会有改变之处。综上所述,本说明书内容不应理解为对本申请的限制。

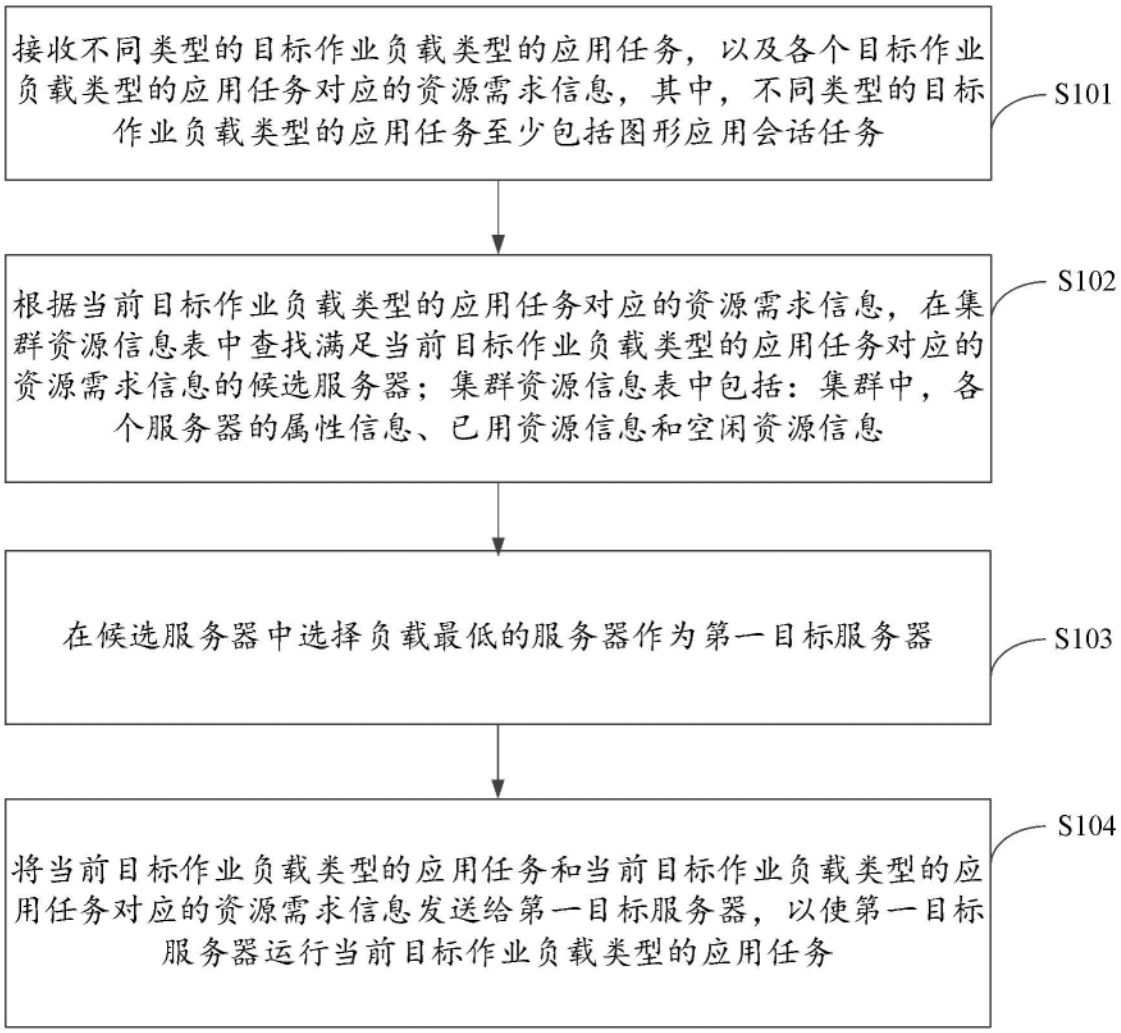

接收不同类型的目标作业负载类型的应用任务，以及各个目标作业负载类型的应用任务对应的资源需求信息，其中，不同类型的目标作业负载类型的应用任务至少包括图形应用会话任务
S101
根据当前目标作业负载类型的应用任务对应的资源需求信息，在集群资源信息表中查找满足当前目标作业负载类型的应用任务对应的资源需求信息的候选服务器；集群资源信息表中包括：集群中，各个服务器的属性信息、已用资源信息和空闲资源信息
S102
在候选服务器中选择负载最低的服务器作为第一目标服务器
S103
将当前目标作业负载类型的应用任务和当前目标作业负载类型的应用任务对应的资源需求信息发送给第一目标服务器，以使第一目标服务器运行当前目标作业负载类型的应用任务
S104

图1

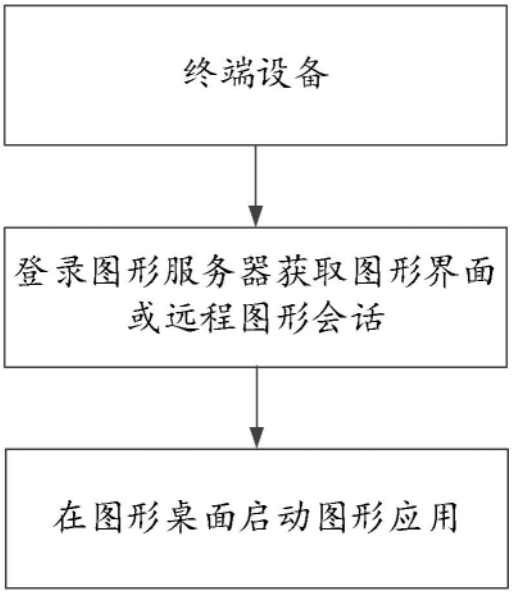

终端设备
登录图形服务器获取图形界面或远程图形会话
在图形桌面启动图形应用

图2

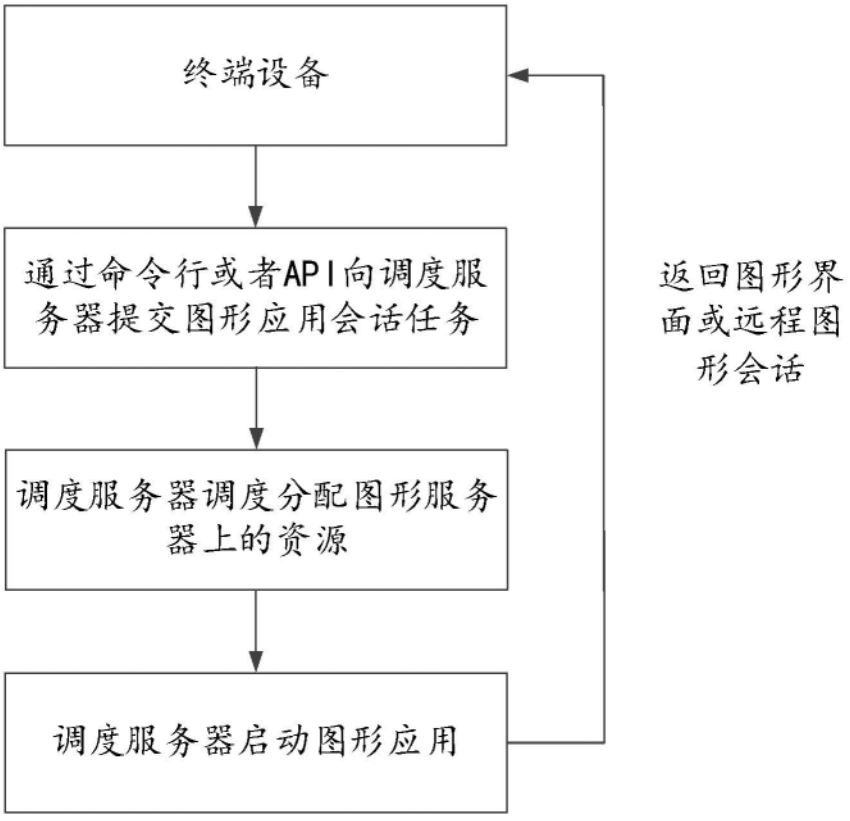

终端设备
通过命令行或者API向调度服务器提交图形应用会话任务
调度服务器调度分配图形服务器上的资源
调度服务器启动图形应用
返回图形界面或远程图形会话

图3

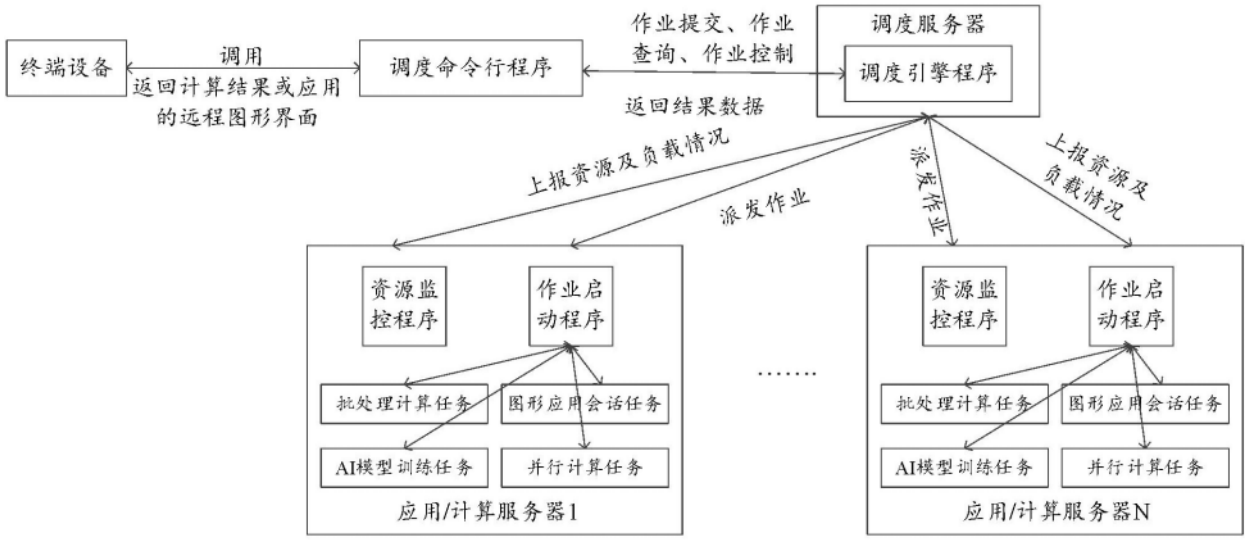

终端设备
调用
返回计算结果或应用的远程图形界面
调度命令行程序
作业提交、作业查询、作业控制
调度服务器
调度引擎程序
返回结果数据
上报资源及负载情况
派发作业
派发作业
上报资源及负载情况
资源监控程序
作业启动程序
批处理计算任务
图形应用会话任务
AI模型训练任务
并行计算任务
应用/计算服务器1
.....
资源监控程序
作业启动程序
批处理计算任务
图形应用会话任务
AI模型训练任务
并行计算任务
应用/计算服务器N

图4

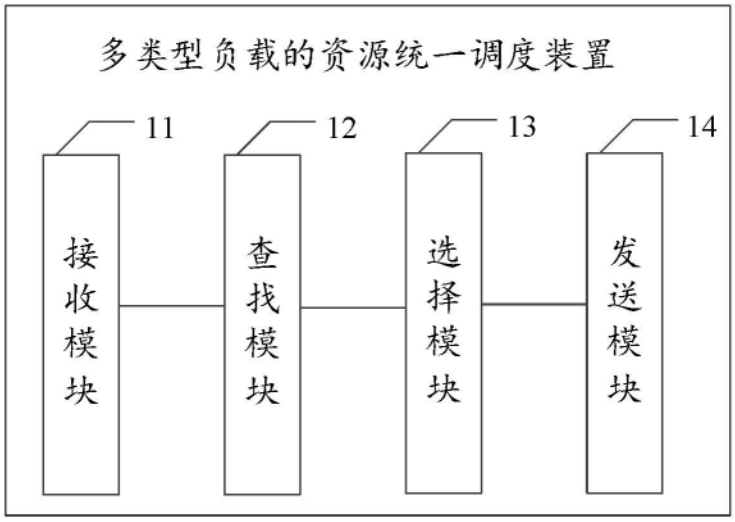

多类型负载的资源统一调度装置
11
12
13
14
接收模块
查找模块
选择模块
发送模块

图5

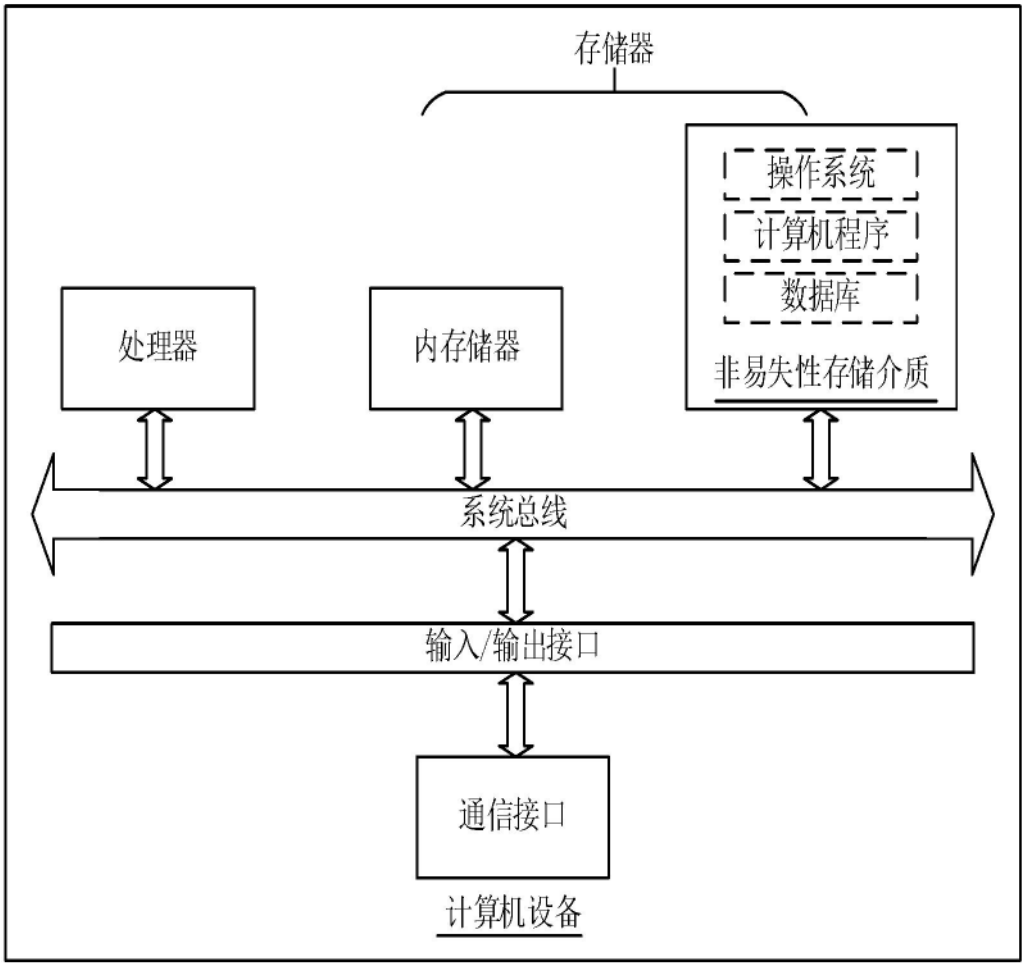


图6