



(12) 发明专利

(10) 授权公告号 CN 112637299 B

(45) 授权公告日 2022.04.26

(21) 申请号 202011481240.X

CN 109995677 A,2019.07.09

(22) 申请日 2020.12.15

CN 102571849 A,2012.07.11

(65) 同一申请的已公布的文献号

CN 102891773 A,2013.01.23

申请公布号 CN 112637299 A

US 2016269308 A1,2016.09.15

US 2018254998 A1,2018.09.06

(43) 申请公布日 2021.04.09

江坤等.跨域流媒体边缘云资源调整策略.

(73) 专利权人 中国联合网络通信集团有限公司

《微型机与应用》.2017,(第21期),

地址 100033 北京市西城区金融大街21号

RGS/NFV-REL003ed112.GROUP

专利权人 联通云数据有限公司

SPECIFICATION

(72) 发明人 李朝霞 邢鑫 康楠 李铭轩

Network Functions Virtualisation (NFV);

Reliability;Report on Models and Features

(74) 专利代理机构 北京同立钧成知识产权代理

for End-to-End Reliability

有限公司 11205

Disclaimer The present document has been

代理人 王征 臧建明

produced and approved by the Network

(51) Int.Cl.

Functions Virtualisation (NFV) ETSI

H04L 67/10 (2022.01)

Industry Specification Group (ISG) and

H04L 67/1074 (2022.01)

represents the views of those members who

H04L 67/60 (2022.01)

participated in this ISG.

(56) 对比文件

It does not necessarily represent the

CN 109582234 A,2019.04.05

views of the entire ETSI membership..

CN 105516242 A,2016.04.20

《ETSI GS NFV-REL 003》.2016,

CN 110858161 A,2020.03.03

审查员 房黎黎

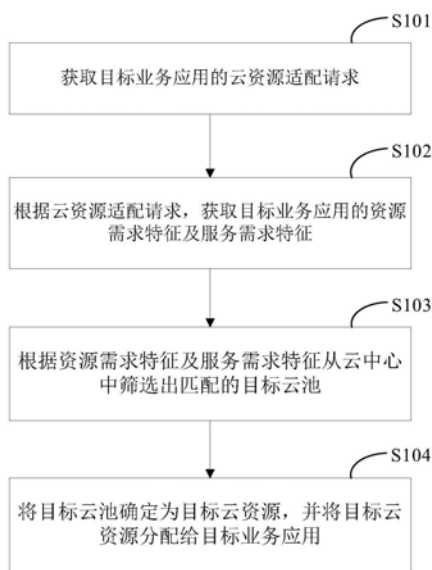
权利要求书2页 说明书15页 附图4页

(54) 发明名称

云资源分配方法、装置、设备、介质及产品

(57) 摘要

本发明提供了一种云资源分配方法、装置、设备、介质及产品,该方法包括:获取目标业务应用的云资源适配请求;根据所述云资源适配请求,获取所述目标业务应用的资源需求特征及服务需求特征;根据所述资源需求特征及所述服务需求特征从云中心中筛选出匹配的目标云池;将所述目标云池确定为目标云资源,并将所述目标云资源分配给目标业务应用。由于全面考虑了业务应用在资源及服务上的需求,所以分配到的目标云资源能够为业务应用提供高质量的服务,进而有效满足业务应用需求。



CN 112637299 B

1. 一种云资源分配方法,其特征在于,包括:
 - 获取目标业务应用的云资源适配请求;
 - 根据所述云资源适配请求,获取所述目标业务应用的资源需求特征及服务需求特征;
 - 根据所述资源需求特征及所述服务需求特征从云中心中筛选出匹配的目标云池;
 - 将所述目标云池确定为目标云资源,并将所述目标云资源分配给目标业务应用;
 - 所述根据所述资源需求特征及所述服务需求特征从云中心中筛选出匹配的目标云池之前,还包括:
 - 确定云中心中多个云池的资源供给数据;
 - 根据预设探测策略确定目标业务应用部署端到每个云池间的第一服务供给数据;
 - 根据所述预设探测策略确定任意两个云池间的第二服务供给数据。
2. 根据权利要求1所述的方法,其特征在于,所述目标业务应用部署端部署第一探针,各云池部署分别部署第二探针,所述根据预设探测策略确定目标业务应用部署端到每个云池间的第一服务供给数据,包括:
 - 控制第一探针监测目标业务应用部署端是否存在当前业务流量即将发送给某云池;
 - 若确定第一探针监测到存在当前业务流量即将发送给该云池,则确定第一探针与第二探针间按照第一探测策略进行当前服务探测;
 - 若确定第一探针监测到不存在当前业务流量即将发送给该云池,则确定第一探针与第二探针间按照第二探测策略进行当前服务探测;
 - 确定当前服务探测结果;
 - 根据预设时间段内获取到的多个当前服务探测结果确定第一服务供给数据。
3. 根据权利要求2所述的方法,其特征在于,所述若确定第一探针监测到存在当前业务流量即将发送给该云池,则确定第一探针与第二探针间按照第一探测策略进行当前服务探测,包括:
 - 若确定第一探针监测到存在当前业务流量即将发送给该云池,则控制探针在所述当前业务流量的数据包的头中加入第一服务测试参数,并将当前业务流量的数据包发送给该云池的第二探针,以控制该云池的第二探针在所述当前业务流量的返回包的包头中加入对应的第二服务测试参数,并将当前业务流量的返回包发送给所述第一探针;
 - 相应地,确定当前服务探测结果,包括:
 - 根据第一服务测试参数及所述第二服务测试参数确定当前服务探测结果。
4. 根据权利要求2所述的方法,其特征在于,所述若确定第一探针监测到不存在当前业务流量即将发送给该云池,则确定第一探针与第二探针间按照第二探测策略进行当前服务探测,包括:
 - 若确定第一探针监测到不存在当前业务流量即将发送给该云池,则控制第一探针向该云池的第二探针发送探测包,以使第二探针根据探测包生成返回包,并将所述返回包发送给所述第一探针;
 - 相应地,确定当前服务探测结果,包括:
 - 根据所述探测包及所述返回包的收发结果确定当前服务探测结果。
5. 根据权利要求1-4任一项所述的方法,其特征在于,所述将所述目标云资源分配给目标业务应用之后,还包括:

周期性监测目标业务应用与目标云资源间的通信路径是否满足SLA要求；

若确定目标业务应用与目标云资源间的通信路径不满足SLA要求，则将目标业务应用与目标云资源间的通信路径调整到最优通信路径。

6. 一种云资源分配装置，其特征在于，包括：

请求获取模块，用于获取目标业务应用的云资源适配请求；

特征获取模块，用于根据所述云资源适配请求，获取所述目标业务应用的资源需求特征及服务需求特征；

云池筛选模块，用于根据所述资源需求特征及所述服务需求特征从云中心中筛选出匹配的目标云池；

云资源分配模块，用于将所述目标云池确定为目标云资源，并将所述目标云资源分配给目标业务应用；

确定模块，用于确定云中心中多个云池的资源供给数据；根据预设探测策略确定目标业务应用部署端到每个云池间的第一服务供给数据；根据预设探测策略确定任意两个云池间的第二服务供给数据。

7. 一种电子设备，其特征在于，包括：

存储器，处理器以及计算机程序；

其中，所述计算机程序存储在所述存储器中，并被配置为由所述处理器执行以实现如权利要求1-5中任一项所述的方法。

8. 一种计算机可读存储介质，其特征在于，其上存储有计算机程序，所述计算机程序被处理器执行以实现如权利要求1-5中任一项所述的方法。

云资源分配方法、装置、设备、介质及产品

技术领域

[0001] 本发明实施例涉及云计算技术领域,尤其涉及一种云资源分配方法、装置、设备、介质及产品。

背景技术

[0002] 云计算是分布式计算的一种,指的是通过网络云将巨大的数据处理程序分解成无数个小程序,然后通过多个服务器组成的系统进行处理和分析这些小程序得到结果并返回给客户。所以云计算是提高信息技术管理和生产效率的重要途径。

[0003] 在云计算中,一般采用端边云的架构。即业务部署端通过边缘云甚至云中心里的云池来处理业务。所以需要为业务应用分配云资源。

[0004] 现有技术中,在为业务应用分配云资源时,只是简单通过云资源的算力来分配,导致分配后的云资源不能为业务应用提供高质量的服务,不能有效满足业务应用需求。

发明内容

[0005] 本发明实施例提供一种云资源分配方法、装置、设备、介质及产品,解决了现有技术中只是简单通过云资源的算力来分配,导致分配后的云资源不能为业务应用提供高质量的服务,不能有效满足业务应用需求的技术问题。

[0006] 第一方面,本发明实施例提供一种云资源分配方法,包括:

[0007] 获取目标业务应用的云资源适配请求;

[0008] 根据所述云资源适配请求,获取所述目标业务应用的资源需求特征及服务需求特征;

[0009] 根据所述资源需求特征及所述服务需求特征从云中心中筛选出匹配的目标云池;

[0010] 将所述目标云池确定为目标云资源,并将所述目标云资源分配给目标业务应用。

[0011] 可选地,如上所述的方法,所述根据所述资源需求特征及所述服务需求特征从云中心中筛选出匹配的目标云池之前,还包括:

[0012] 确定云中心中多个云池的资源供给数据;

[0013] 根据预设探测策略确定目标业务应用部署端到每个云池间的第一服务供给数据;

[0014] 根据所述预设探测策略确定任意两个云池间的第二服务供给数据。

[0015] 可选地,如上所述的方法,所述目标业务应用部署端部署第一探针,各云池部署分别部署第二探针,所述根据预设探测策略确定目标业务应用部署端到每个云池间的第一服务供给数据,包括:

[0016] 控制第一探针监测目标业务应用部署端是否存在当前业务流量即将发送给某云池;

[0017] 若确定第一探针监测到存在当前业务流量即将发送给该云池,则确定第一探针与第二探针间按照第一探测策略进行当前服务探测;

[0018] 若确定第一探针监测到不存在当前业务流量即将发送给该云池,则确定第一探针

与第二探针间按照第二探测策略进行当前服务探测；

[0019] 确定当前服务探测结果；

[0020] 根据预设时间段内获取到的多个当前服务探测结果确定第一服务供给数据。

[0021] 可选地,如上所述的方法,所述若确定第一探针监测到存在当前业务流量即将发送给该云池,则确定第一探针与第二探针间按照第一探测策略进行当前服务探测,包括:

[0022] 若确定第一探针监测到存在当前业务流量即将发送给该云池,则控制探针在所述当前业务流量的数据包的头中加入第一服务测试参数,并将当前业务流量的数据包发送给该云池的第二探针,以控制该云池的第二探针在所述当前业务流量的返回包的包头中加入对应的第二服务测试参数,并将当前业务流量的返回包发送给所述第一探针;

[0023] 相应地,确定当前服务探测结果,包括:

[0024] 根据第一服务测试参数及所述第二服务测试参数确定当前服务探测结果。

[0025] 可选地,如上所述的方法,所述若确定第一探针监测到不存在当前业务流量即将发送给该云池,则确定第一探针与第二探针间按照第二探测策略进行当前服务探测,包括:

[0026] 若确定第一探针监测到不存在当前业务流量即将发送给该云池,则控制第一探针向该云池的第二探针发送探测包,以使第二探针根据探测包生成返回包,并将所述返回包发送给所述第一探针;

[0027] 相应地,确定当前服务探测结果,包括:

[0028] 根据所述探测包及所述返回包的收发结果确定当前服务探测结果。

[0029] 可选地,如上所述的方法,所述将所述目标云资源分配给目标业务应用之后,还包括:

[0030] 周期性监测目标业务应用与目标云资源间的通信路径是否满足SLA要求;

[0031] 若确定目标业务应用与目标云资源间的通信路径不满足SLA要求,则将目标业务应用与目标云资源间的通信路径调整到最优通信路径。

[0032] 第二方面,本发明实施例提供一种云资源分配装置,包括:

[0033] 请求获取模块,用于获取目标业务应用的云资源适配请求;

[0034] 特征获取模块,用于根据所述云资源适配请求,获取所述目标业务应用的资源需求特征及服务需求特征;

[0035] 云池筛选模块,用于根据所述资源需求特征及所述服务需求特征从云中心中筛选出匹配的目标云池;

[0036] 云资源分配模块,用于将所述目标云池确定为目标云资源,并将所述目标云资源分配给目标业务应用。

[0037] 可选地,如上所述的装置,还包括:

[0038] 确定模块,用于确定云中心中多个云池的资源供给数据;根据预设探测策略确定目标业务应用部署端到每个云池间的第一服务供给数据;根据所述预设探测策略确定任意两个云池间的第二服务供给数据。

[0039] 可选地,如上所述的装置,所述目标业务应用部署端部署第一探针,各云池部署分别部署第二探针,所述确定模块,在根据预设探测策略确定目标业务应用部署端到每个云池间的第一服务供给数据时,具体用于:

[0040] 控制第一探针监测目标业务应用部署端是否存在当前业务流量即将发送给某云

池;若确定第一探针监测到存在当前业务流量即将发送给该云池,则确定第一探针与第二探针间按照第一探测策略进行当前服务探测;若确定第一探针监测到不存在当前业务流量即将发送给该云池,则确定第一探针与第二探针间按照第二探测策略进行当前服务探测;确定当前服务探测结果;根据预设时间段内获取到的多个当前服务探测结果确定第一服务供给数据。

[0041] 可选地,如上所述的装置,所述确定模块,在若确定第一探针监测到存在当前业务流量即将发送给该云池,则确定第一探针与第二探针间按照第一探测策略进行当前服务探测时,具体用于:

[0042] 若确定第一探针监测到存在当前业务流量即将发送给该云池,则控制探针在所述当前业务流量的数据包的包头中加入第一服务测试参数,并将当前业务流量的数据包发送给该云池的第二探针,以控制该云池的第二探针在所述当前业务流量的返回包的包头中加入对应的第二服务测试参数,并将当前业务流量的返回包发送给所述第一探针;

[0043] 相应地,所述确定模块,在确定当前服务探测结果时,具体用于:

[0044] 根据第一服务测试参数及所述第二服务测试参数确定当前服务探测结果。

[0045] 可选地,如上所述的装置,所述确定模块,在若确定第一探针监测到不存在当前业务流量即将发送给该云池,则确定第一探针与第二探针间按照第二探测策略进行当前服务探测时,具体用于:

[0046] 若确定第一探针监测到不存在当前业务流量即将发送给该云池,则控制第一探针向该云池的第二探针发送探测包,以使第二探针根据探测包生成返回包,并将所述返回包发送给所述第一探针;

[0047] 相应地,所述确定模块,在确定当前服务探测结果时,具体用于:

[0048] 根据所述探测包及所述返回包的收发结果确定当前服务探测结果。

[0049] 可选地,如上所述的装置,还包括:

[0050] 要求监测模块,用于周期性监测目标业务应用与目标云资源间的通信路径是否满足SLA要求;

[0051] 路径调整模块,用于若确定目标业务应用与目标云资源间的通信路径不满足SLA要求,则将目标业务应用与目标云资源间的通信路径调整到最优通信路径。

[0052] 第三方面,本发明实施例提供一种电子设备,包括:

[0053] 存储器,处理器以及计算机程序;

[0054] 其中,所述计算机程序存储在所述存储器中,并被配置为由所述处理器执行以实现如第一方面中任一项所述的方法。

[0055] 第四方面,本发明实施例提供一种计算机可读存储介质,其上存储有计算机程序,所述计算机程序被处理器执行以实现如第一方面中任一项所述的方法。

[0056] 第五方面,本发明实施例提供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第一方面中任一项所述的方法。

[0057] 本发明实施例提供一种云资源分配方法、装置、设备、介质及产品,通过获取目标业务应用的云资源适配请求;根据所述云资源适配请求,获取所述目标业务应用的资源需求特征及服务需求特征;根据所述资源需求特征及所述服务需求特征从云中心中筛选出匹配的目标云池;将所述目标云池确定为目标云资源,并将所述目标云资源分配给目标业务

应用。由于全面考虑了业务应用在资源及服务上的需求,所以分配到的目标云资源能够为业务应用提供高质量的服务,进而有效满足业务应用需求。

[0058] 应当理解,上述发明内容部分中所描述的内容并非旨在限定本发明的实施例的关键或重要特征,亦非用于限制本发明的范围。本发明的其它特征将通过以下的描述变得容易理解。

附图说明

[0059] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作一简单地介绍,显而易见地,下面描述中的附图是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动性的前提下,还可以根据这些附图获得其他的附图。

[0060] 图1是可以实现本发明实施例的云资源分配方法的一种应用场景图;

[0061] 图2为本发明一实施例提供的云资源分配方法的流程图;

[0062] 图3为本发明另一实施例提供的云资源分配方法的流程图;

[0063] 图4为本发明实施例一提供的云资源分配装置的结构示意图;

[0064] 图5为本发明实施例另一提供的云资源分配装置的结构示意图;

[0065] 图6为本发明一实施例提供的电子设备的结构示意图。

具体实施方式

[0066] 下面将参照附图更详细地描述本发明的实施例。虽然附图中显示了本发明的某些实施例,然而应当理解的是,本发明可以通过各种形式来实现,而且不应该被解释为限于这里阐述的实施例,相反提供这些实施例是为了更加透彻和完整地理解本发明。应当理解的是,本发明的附图及实施例仅用于示例性作用,并非用于限制本发明的保护范围。

[0067] 本发明实施例的说明书和权利要求书及上述附图中的术语“第一”、“第二”、“第三”、“第四”等(如果存在)是用于区别类似的对象,而不必用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换,以便这里描述的本发明实施例如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,术语“包括”和“具有”以及他们的任何变形,意图在于覆盖不排他的包含,例如,包含了一系列步骤或单元的过程、方法、系统、产品或设备不必限于清楚地列出的那些步骤或单元,而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它步骤或单元。

[0068] 为了清楚理解本申请的技术方案,首先对现有技术的方案进行详细介绍。

[0069] 现有技术中,在为业务应用分配云资源时,一般根据业务应用的资源需求在云中心中确定具有匹配资源算力的云池,进而将具有匹配资源算力的云池确定为云资源。例如业务应用的资源需求是需要有GPU资源和CPU资源,则在云中心中筛选出具有GPU资源和CPU资源的云池作为匹配的云资源。但在业务应用中为了保证业务的顺利进行而且用户体验好,单单只考虑资源是不够的,导致分配的云资源不能为业务应用提供高质量的服务,不能有效满足业务应用需求。

[0070] 针对现有技术中的问题,发明人在研究中发现,为了为业务应用提供高质量的服务,不仅需要考虑业务应用的资源需求,而且需要考虑业务应用的服务需求,即在服务等级

协议(简称:SLA)上也满足业务应用。所以本发明实施例中,需要获取目标业务应用的资源需求特征及服务需求特征。根据资源需求特征及服务需求特征从云中心中筛选出匹配的目标云池,进而将目标云池作为目标云资源分配给目标业务应用。由于全面考虑了业务应用在资源及服务上的需求,所以分配到的目标云资源能够为业务应用提供高质量的服务,进而有效满足业务应用需求。

[0071] 所以发明人基于上述的创造性发现,提出了本发明实施例的技术方案。下面对本发明实施例提供的云资源分配方法的应用场景进行介绍。

[0072] 如图1所示,在目标业务应用部署端1有云资源分配的需求时,向电子设备2发送目标业务应用的云资源适配请求,电子设备获取到目标业务应用的云资源适配请求后,从目标业务应用部署端1获取目标业务应用的资源需求特征及服务需求特征。然后从云中心3中获取到多个云池的资源供给数据及服务供给数据,根据资源需求特征及服务需求特征从云中心中筛选出匹配的目标云池,将目标云池确定为目标云资源,并将目标云资源分配给目标业务应用。并且可向目标业务应用部署端1发送目标云资源信息。

[0073] 下面以具体地实施例对本发明的技术方案以及本申请的技术方案如何解决上述技术问题进行详细说明。下面这几个具体的实施例可以相互结合,对于相同或相似的概念或过程可能在某些实施例中不再赘述。下面将结合附图,对本发明的实施例进行描述。

[0074] 实施例一

[0075] 图2为本发明一实施例提供的云资源分配方法的流程图,如图2所示,本实施例的执行主体为云资源分配装置,该云资源分配装置可集成在电子设备中,电子设备可以为计算机,服务器或服务器集群等,则本实施例提供的云资源分配方法包括以下几个步骤。

[0076] 步骤101,获取目标业务应用的云资源适配请求。

[0077] 其中,目标业务应用为进行云资源分配的业务应用。该目标业务应用如可以为基本办公业务应用,还可以为视频业务应用,或其他类型的业务应用,本实施例中对此不作限定。

[0078] 本实施例中,在目标业务应用的部署端有云资源适配需求时,可向电子设备发送云资源适配请求,则电子设备获取到云资源适配请求。或者,在目标业务应用有云资源适配需求时,用户可通过电子设备的该云资源分配方法的应用程序的客户端或者网页触发云资源适配请求。

[0079] 其中,云资源适配请求中包括目标业务应用的标识信息。目标业务应用的标识信息如可以为目标业务应用的名称,编码等唯一表示该目标业务应用的信息。

[0080] 步骤102,根据云资源适配请求,获取目标业务应用的资源需求特征及服务需求特征。

[0081] 本实施例中,可选地,响应于该云资源适配请求,向目标业务应用的部署端发送特征获取请求,则目标业务应用的部署端根据该获取请求获取目标业务应用的资源需求特征及服务需求特征,并将目标业务应用的资源需求特征及服务需求特征发送给电子设备。

[0082] 或者可选地,在电子设备中预先在存储区域存储有目标业务应用的资源需求特征及服务需求特征,则响应于该云资源适配请求,从存储区域获取目标业务应用的资源需求特征及服务需求特征。

[0083] 或者可选地,响应于该云资源适配请求,向用户通过客户端或者网页显示操作界

面,用户可通过操作界面输入目标业务应用的资源需求特征及服务需求特征。

[0084] 其中,目标业务应用的资源需求特征可以包括:处理器类型,内存大小。服务需求特征可以包括:基本时延需求,抖动需求,丢包率要求,但用户峰值带宽要求及体验要求等。

[0085] 如表1所示,分别为目标业务应用为基本办公业务应用及视频业务应用的资源需求特征及服务需求特征的示意。

[0086] 步骤103,根据资源需求特征及服务需求特征从云中心中筛选出匹配的目标云池。

[0087] 本实施例中,首先从云中心中获取多个云池的资源供给数据,第一服务供给数据及第二服务供给数据。

[0088] 其中,云池的资源供给数据表示云池的资源算力。第一服务供给数据表示目标业务应用部署端到该云池的服务供给数据。第二服务供给数据表示该云池与其他任意一个云池间的服务供给数据。

[0089] 其中,每个云池的云池的资源供给数据可以包括:处理器类型(CPU\GPU\FPGA等),用于计算机存储设备(IOPS)的类型,内存大小等。第一服务供给数据可以包括:目标业务应用部署端到该云池的时延,丢包率及抖动。第二服务供给数据包括该云池与其他任意一个云池间的时延,丢包率及抖动。具体的多个云池的资源供给数据,第一服务供给数据及第二服务供给数据资源算力可以表示为表2。

[0090] 其中,表2中的“cloudid”为云池ID。“processor”为处理器类型,不同的取值表示处理器类型不同。“cnumber”表示处理器的核数量。“memory”表示内存大小。“type(sata10/sas20/SSD30)”表示IOPS的类型,取值不同表示类型不同。“delay”“lost”“jitter”分别表示目标业务应用部署端到该云池的时延,丢包率及抖动。“adelay”“alost”“ajitter”分别表示第一个云池分别与其他云池间的时延,丢包率及抖动。同理“bdelay”“blost”“bjitter”分别表示第二个云池分别与其他云池间的时延,丢包率及抖动。

[0091] 表1:目标业务应用的资源需求特征及服务需求特征示意表

目标业务应用	体验要求	处理器	内存	基础RTT要求	抖动要求	丢包率要求	单用户瞬时峰值带宽需求 (Mbps)		
							前提条件1: 服务器负载水平	前提条件2: 基础RTT水平	带宽要求
基础办公	1、判定指标: 操作响应时间、画面更新频率、画面显示质量 2、用户体验要求 (1-5) 在3以上	cpu	128	≤30ms	抖动的概率 <8%	≤0.08%	轻载	5ms	≥18Mbps
		cpu	128					10ms	≥20Mbps
		cpu	128					20ms	≥25Mbps
		cpu	128				中载	5ms	≥22Mbps
		cpu	128					10ms	≥25Mbps
		cpu	128					20ms	≥33Mbps
		cpu	128				重载	5ms	≥29Mbps
		cpu	128					10ms	≥33Mbps
		cpu	128					20ms	≥50Mbps
		[0092] 视频业务系统	1、前置条件: 满足基础办公要求 2、判定指标: 视频有效帧率 3、用户体验要求 (1-5) 在3以上				gpu	128	≤25ms
cpu	32			≤15ms	抖动的概率 <8%	≤0.08%	10ms, ≥25Mbps		

[0093] 具体地,本实施例中,将资源需求特征与多个云池的资源供给数据进行匹配,并将服务需求特征与多个云池的第一服务供给数据和/或第二服务供给数据进行匹配,获取到资源供给数据及第一服务供给数据和/或第二服务供给数据均与目标业务应用匹配的云池作为目标云池。

[0094] 需要说明的是,若目标业务应用需要的云资源的个数为N个,而在多个云池中匹配出的云池为M个。M多于N,则对该M个云池继续进行筛选,确定出M个云池中彼此的网络状态较好的N个云池作为目标云池。

[0095] 或者可将M个云池的资源供给数据,第一服务供给数据及第二服务供给数据均发送给目标业务应用部署端,由目标业务应用部署端从M个云池中筛选出N个云池作为目标云池。

[0096] 表2:多个云池的资源供给数据,第一服务供给数据及第二服务供给数据资源算力的示意表

[0097]

	cbudid	processor	number	memory	type (sata10/sas20/SSD30)	delay (ms)	bst	jit	ade	abs	ajit	bde	bbs	bjit	...
C1	2d1bdab1-f615-490c-927f-4343d4d2bc6e	10	32	128	10	5	10*-9	8				25	10*-9	12	
C2	af71c9e4-9618-4ae3-b4c7-3966922d03c1	20	16	256	10	10	10*-12	8	10	10*-12	8				
C3	aa472dfcb88-41a7-b0e3-ee9b5a5915b6	30	32	512	10	18	10*-9	10	18	10*-9	10	18	10*-9	10	
C4	1ddc6441-6026-4880-ab e8-8943b03c86a1	10	64	256	20	8	10*-12	12	8	10*-12	12	8	10*-12	12	
C5	6c187b10-b5f3-4204-abf5-8591de682c25	10	32	384	30	15	10*-6	8	15	10*-6	8	15	10*-6	8	
C6	756570e3-5568-44dc-95c0-403717b9bfed	10	32	256	20	20	10*-9	8	13	10*-9	8	5	10*-9	8	
C7	01c48672-02b9-42a1-b95e-c34ae25329b0	10	32	256	10	34	10*-8	10	15	10*-12	8	10	10*-12	8	
C8	3c0819fcd60-4812-8c6f-9be3694893f1	10	32	256	20	19	10*-10	8	28	10*-12	10	18	10*-9	10	
C9	6cc70bec-7c9d-4f05-9684-07d9228284f7	20	32	256	30	22	10*-7	11	7	10*-12	12	8	10*-12	12	
C10	ebb3bc61-aa38-43f1-84b4-695919d745f1	10	32	256	20	29	10*-7	8	12	10*-6	8	15	10*-6	8	

[0098] 步骤104,将目标云池确定为目标云资源,并将目标云资源分配给目标业务应用。

[0099] 本实施例中,将目标云池作为目标云资源,并可将目标云资源的资源算力及网络状态感知数据发送给目标业务应用部署端。

[0100] 本实施例中,将目标云资源分配给目标业务应用,即建立目标业务应用部署端与目标云资源间的通信路径。以通过目标云资源对目标业务应用进行处理。

[0101] 本实施例提供的云资源分配方法,通过获取目标业务应用的云资源适配请求;根据云资源适配请求,获取目标业务应用的资源需求特征及服务需求特征;根据资源需求特征及服务需求特征从云中心中筛选出匹配的目标云池;将目标云池确定为目标云资源,并将目标云资源分配给目标业务应用。由于全面考虑了业务应用在资源及服务上的需求,所以分配到的目标云资源能够为业务应用提供高质量的服务,进而有效满足业务应用需求。

[0102] 实施例二

[0103] 图3为本发明另一实施例提供的云资源分配方法的流程图,如图3所示,本实施例提供的云资源分配方法,是在本发明实施例一提供的云资源分配方法的基础上,对步骤103的进一步细化,并且还包含了其他步骤,则本实施例提供的云资源分配方法包括以下步骤。

[0104] 步骤201,确定云中心中多个云池的资源供给数据。

[0105] 具体地,本实施例中,可向云中心发送云池资源供给数据获取请求,云中心根据云

池资源供给数据获取请求获取每个云池的资源供给数据,并将云池的供给数据发送给电子设备。

[0106] 其中,每个云池的资源供给数据如表2所示,可包括:处理器类型(CPU\GPU\FPGA等),处理器的核数量,用于计算机存储设备(IOPS)的类型,内存大小等。

[0107] 步骤202,根据预设探测策略确定目标业务应用部署端到每个云池间的第一服务供给数据。

[0108] 作为一种可选实施方式,本实施例中,目标业务应用部署端部署第一探针,各云池部署分别部署第二探针,相应地,步骤202包括以下步骤:

[0109] 步骤2021,控制第一探针监测目标业务应用部署端是否存在当前业务流量即将发送给某云池。

[0110] 本实施例中,将第一探针部署在目标业务应用部署端,若目标业务应用部署端与某云池有业务流量,即目标业务应用部署端即将向该云池发送业务数据包,则第一探针能够监测到目标业务应用部署端存在当前业务流量即将发送给该云池。若目标业务应用部署端与某云池不存在业务流量,则第一探针监测不到目标业务应用部署端存在当前业务流量即将发送给该云池。

[0111] 步骤2022,若确定第一探针监测到存在当前业务流量即将发送给该云池,则确定第一探针与第二探针间按照第一探测策略进行当前服务探测。

[0112] 可选地,本实施例中,步骤2022具体包括:

[0113] 若确定第一探针监测到存在当前业务流量即将发送给该云池,则控制探针在当前业务流量的数据包的包头中加入第一服务测试参数,并将当前业务流量的数据包发送给该云池的第二探针,以控制该云池的第二探针在当前业务流量的返回包的包头中加入对应的第二服务测试参数,并将当前业务流量的返回包发送给第一探针。

[0114] 具体地,本实施例中,在控制探针在当前业务流量的数据包的包头中加入第一服务测试参数。例如在当前业务流量的数据包的包头的IP header和payload信息之间加入第一服务测试参数。

[0115] 其中,第一服务测试参数可以包括:数据包的编号,发送时间戳。数据包的编号可表示:Seq_number1,发送时间戳可表示:timestamp1。

[0116] 本实施例中,将携带有第一服务测试参数的当前业务流量的数据包发送给云池后,云池针对该当前业务流量的数据包进行处理后,生成对应的返回包,控制该云池的第二探针在当前业务流量的返回包的包头中加入对应的第二服务测试参数。相应地,在当前业务流量的返回包的包头IP header和payload之间加入第二服务测试参数。

[0117] 其中,第二服务测试参数可以包括:返回包的编号,发送时间戳。返回包的编号可表示:Seq_number2,发送时间戳可表示:timestamp2。

[0118] 步骤2023,若确定第一探针监测到不存在当前业务流量即将发送给该云池,则确定第一探针与第二探针间按照第二探测策略进行当前服务探测。

[0119] 可选地,本实施例中,步骤2023具体包括:

[0120] 若确定第一探针监测到不存在当前业务流量即将发送给该云池,则控制第一探针向该云池的第二探针发送探测包,以使第二探针根据探测包生成返回包,并将返回包发送给第一探针。

[0121] 具体地,本实施例中,若确定第一探针未监测到存在当前业务流量即将发送给该云池,则第一探针无法利用目标业务应用部署端与该云池之间的业务流量进行探测,所以控制第一探针向该云池的第二探针发送探测包,探测包可以为ICMP或UDP探测包,获取发送探测包的时间戳及探测包大小。第二探针接收到该探测包后,生成返回包,并监测到第一探针接收到返回包后,获取返回包的发送时间戳及返回包的大小。

[0122] 步骤2024,确定当前服务探测结果。

[0123] 若执行步骤2022,则相应地,步骤2024具体包括:

[0124] 根据第一服务测试参数及第二服务测试参数确定当前服务探测结果。

[0125] 具体地,本实施例中,由于在第一服务测试参数中包括数据包的编号,在第二服务测试参数中包括返回包的编号,所以可以编号关系确定为对应的数据包和返回包,然后根据发送时间戳确定当前服务探测结果中的丢包情况,抖动情况及时延。

[0126] 若执行步骤2023,则相应地,步骤2024具体包括:

[0127] 根据探测包及返回包的收发结果确定当前服务探测结果。

[0128] 具体地,本实施例中,根据探测包和返回包的大小确认当前服务探测结果中的丢包情况。根据探测包的发送时间戳,及返回包的发送时间戳确认当前服务探测结果中的时延及抖动情况。

[0129] 步骤2025,根据预设时间段内获取到的多个当前服务探测结果确定第一服务供给数据。

[0130] 本实施例中,针对每各当前服务探测结果均有丢包情况,时延及抖动情况,所以根据预设时间段内的多个当前服务探测结果中的丢包情况可确定第一服务供给数据中的丢包率。根据预设时间段内的多个当前服务探测结果中的时延确定第一服务供给数据中的时延,根据预设时间段内的多个当前服务探测结果中的抖动确定第一服务供给数据中的抖动。

[0131] 在确定第一服务供给数据中的时延时,可计算预设时间段内多个时延的平均值。在确定第一服务供给数据中的抖动时,可计算预设时间段内多个抖动的平均值。

[0132] 本实施例提供的云资源分配方法,在根据预设探测策略确定目标业务应用部署端到每个云池间的第一服务供给数据时,控制第一探针监测目标业务应用部署端是否存在当前业务流量即将发送给某云池;若确定第一探针监测到存在当前业务流量即将发送给该云池,则控制探针在当前业务流量的数据包的包头中加入第一服务测试参数,并将当前业务流量的数据包发送给该云池的第二探针,以控制该云池的第二探针在当前业务流量的返回包的包头中加入对应的第二服务测试参数,并将当前业务流量的返回包发送给第一探针,确定当前服务探测结果;根据预设时间段内获取到的多个当前服务探测结果确定第一服务供给数据。能够控制探针对目标业务应用部署端与云池之间进行服务探测时,借助业务流量完成服务探测过程,避免云池段由于流量极大,服务探测过程占据带宽进而影响正常业务流的问题。

[0133] 步骤203,根据预设探测策略确定任意两个云池间的第二服务供给数据。

[0134] 需要说明的是,步骤203中,根据预设探测策略确定任意两个云池间的第二服务供给数据的实现方式与步骤202中根据预设探测策略确定目标业务应用部署端到每个云池间的第一服务供给数据的实现方式类似,在此不再一一赘述。

[0135] 步骤204,获取目标业务应用的云资源适配请求。

[0136] 步骤205,根据云资源适配请求,获取目标业务应用的资源需求特征及服务需求特征。

[0137] 本实施例中,步骤204-步骤205的实现方式与本发明实施例一中的步骤101-步骤102的实现方式类似,在此不再一一赘述。

[0138] 步骤206,根据资源需求特征及服务需求特征从云中心中筛选出匹配的目标云池。

[0139] 本实施例中,首先根据资源需求特征确定资源需求的处理器类型。若为一种处理器类型,则在云中心中筛选出与资源需求特征相匹配的至少一个云池。然后根据至少一个云池的第一服务供给数据及式(1)计算第一服务匹配参数。

$$[0140] \quad d = (\text{delay} - 25) + \lg(\text{lost}) + (\text{jitter} - 8) \quad (1)$$

[0141] 其中,d为第一服务匹配参数,delay为目标业务应用部署端与相匹配的某个云池间的时延,lost为目标业务应用部署端与相匹配的某个云池间的丢包率,jitter为目标业务应用部署端与相匹配的某个云池间的抖动。

[0142] 本实施例中,按照第一服务匹配参数有小到大的顺序进行排序,以确定第一服务匹配参数有小到大的相匹配的云池列表。将该云池列表推送给目标业务应用部署端,以供用户在云池列表中进行选择。或者也可以根据预先配置的该处理器类型的云池个数,在云池列表中从前面选择该个数的云池作为目标云池。

[0143] 本实施例中,若根据资源需求特征确定资源需求的处理器类型为多个。则在选择完第一种处理器类型的云池作为目标云池后,在获取第二种处理器类型的云池,根据第二服务供给数据及式(2),计算目标云池与第二种处理器类型的云池之间的第二服务匹配参数。

$$[0144] \quad S = c.\text{adelay} + \lg(c.\text{alost}) + c.\text{ajitter} \quad (2)$$

[0145] 其中,S为第二服务匹配参数,c.adelay为目标云池与某一第二种处理器类型的云池之间的时延,c.alost为目标云池与某一第二种处理器类型的云池之间的丢包率,c.ajitter为目标云池与某一第二种处理器类型的云池之间的抖动。

[0146] 本实施例中,按照第二服务匹配参数有小到大的顺序进行排序,以确定第二服务匹配参数有小到大的相匹配的云池列表。根据预先配置的该处理器类型的云池个数,在该云池列表中从前面选择该个数的云池作为该处理器类型的目标云池。

[0147] 可以理解的是,按照第二处理器类型的目标云池的筛选方式来筛选其他处理器类型的目标云池。

[0148] 步骤207,将目标云池确定为目标云资源,并将目标云资源分配给目标业务应用。

[0149] 本实施例中,步骤207的实现方式与本发明实施例一中的步骤104的实现方式类似,在此不再一一赘述。

[0150] 步骤208,周期性监测目标业务应用与目标云资源间的通信路径是否满足SLA要求。

[0151] 本实施例中,由于需要云资源的业务应用有很多。而且目标业务应用需要依赖于多个目标云池,且多个目标云池之间会有同步要求或协作要求。所以需要使目标业务应用与多个目标云池之间的通信路径实时满足SLA要求。所以根据目标业务应用的SLA要求,确定目标业务应用与目标云资源间的当前通信路径是否满足SLA要求。

[0152] 步骤209,若确定目标业务应用与目标云资源间的通信路径不满足SLA要求,则将目标业务应用与目标云资源间的通信路径调整到最优通信路径。

[0153] 本实施例中,在确定目标业务应用与目标云资源间的通信路径不满足SLA要求时,需要通过最优通信路径确定策略确定出当前的最有通信路径。

[0154] 其中,在目标业务应用部署端与目标云资源间若存在边缘云,则首先根据分布式最优路径确定策略确定出目标业务应用部署端与边缘云之间的最优路径。在目标业务应用部署端于边缘云之间具有两个隧道。示例性的,可以一条专线一条internet加密隧道,或者两条internet加密隧道。若确定目标业务应用部署端与边缘云之间当前的一条隧道断开,那么切换到另外一条隧道上作为目标业务应用部署端与边缘云之间的最优路径。

[0155] 若在目标业务应用部署端与目标云资源间不存在边缘云,则直接根据SLA要求确定最优通信路径。具体地,由于SLA要求主要有:时延,丢包率,最小最大带宽利用率,路径必须含网元,路径必须避网元等要求。所以在确定最优通信路径时,对SLA要求的几个参数综合考虑。

[0156] 首先在云池之间产生网络向量图,图中每个点都是云池对应的网元,点和点的连接就是物理链路,每个连接有4个参数对应的权重。4个参数分别为:时延,丢包率,最小最大带宽利用率,抖动。根据四个参数对应的权重确定该连接对应的该段物理链路是否为最优链路。其中,连接的权重越大,说明综合考虑到:时延,丢包率,最小最大带宽利用率,抖动后,越适合作为最有路径中的某一段物理链路。并且在确定每段最优物理链路时,考虑路径必须含网元和路径必须避开网元的约束。将每段最优物理链路结合后形成最优通信路径。

[0157] 本实施例中,在确定出最优通信链路后,将目标业务应用与目标云资源间的通信路径调整到最优通信路径。具体地,屏蔽异构网元(例如:路由器,交换机,虚拟网元)的差异性,由路径计算微服务通过REST接口下发最优通信路径到网络编排器,继而下发到对应的目标网元。目标网元对应目标云池。进而各目标网元根据最优通信路径进行当前通信路径的调整。

[0158] 本实施例提供的云资源分配方法,通过周期性监测目标业务应用与目标云资源间的通信路径是否满足SLA要求;若确定目标业务应用与目标云资源间的通信路径不满足SLA要求,则将目标业务应用与目标云资源间的通信路径调整到最优通信路径,能够保证目标业务应用与目标云资源间的通信路径随时为最优通信路径,进而满足目标业务应用的SLA要求。

[0159] 实施例三

[0160] 图4为本发明实施例一提供的云资源分配装置的结构示意图,如图4所示,本实施例提供的云资源分配装置30包括:请求获取模块31,特征获取模块32,云池筛选模块33,云资源分配模块34。

[0161] 其中,请求获取模块31,用于获取目标业务应用的云资源适配请求。特征获取模块32,用于根据云资源适配请求,获取目标业务应用的资源需求特征及服务需求特征。云池筛选模块33,用于根据资源需求特征及服务需求特征从云中心中筛选出匹配的目标云池。云资源分配模块34,用于将目标云池确定为目标云资源,并将目标云资源分配给目标业务应用。

[0162] 本实施例提供的云资源分配装置可以执行图2所示方法实施例的技术方案,其实

现原理和技术效果类似,此处不再赘述。

[0163] 实施例四

[0164] 图5为本发明实施例另一提供的云资源分配装置的结构示意图,如图5所示,本实施例提供的云资源分配装置40在本发明实施例一提供的云资源分配装置30基础上,进一步地,还包括:确定模块41,要求监测模块42,及路径调整模块43。

[0165] 可选地,确定模块41,用于确定云中心中多个云池的资源供给数据;根据预设探测策略确定目标业务应用部署端到每个云池间的第一服务供给数据;根据预设探测策略确定任意两个云池间的第二服务供给数据。

[0166] 可选地,目标业务应用部署端部署第一探针,各云池部署分别部署第二探针,确定模块41,在根据预设探测策略确定目标业务应用部署端到每个云池间的第一服务供给数据时,具体用于:

[0167] 控制第一探针监测目标业务应用部署端是否存在当前业务流量即将发送给某云池;若确定第一探针监测到存在当前业务流量即将发送给该云池,则确定第一探针与第二探针间按照第一探测策略进行当前服务探测;若确定第一探针监测到不存在当前业务流量即将发送给该云池,则确定第一探针与第二探针间按照第二探测策略进行当前服务探测;确定当前服务探测结果;根据预设时间段内获取到的多个当前服务探测结果确定第一服务供给数据。

[0168] 可选地,确定模块41,在若确定第一探针监测到存在当前业务流量即将发送给该云池,则确定第一探针与第二探针间按照第一探测策略进行当前服务探测时,具体用于:

[0169] 若确定第一探针监测到存在当前业务流量即将发送给该云池,则控制探针在当前业务流量的数据包的头中加入第一服务测试参数,并将当前业务流量的数据包发送给该云池的第二探针,以控制该云池的第二探针在当前业务流量的返回包的包头中加入对应的第二服务测试参数,并将当前业务流量的返回包发送给第一探针;

[0170] 相应地,确定模块41,在确定当前服务探测结果时,具体用于:

[0171] 根据第一服务测试参数及第二服务测试参数确定当前服务探测结果。

[0172] 可选地,如上的装置,确定模块,在若确定第一探针监测到不存在当前业务流量即将发送给该云池,则确定第一探针与第二探针间按照第二探测策略进行当前服务探测时,具体用于:

[0173] 若确定第一探针监测到不存在当前业务流量即将发送给该云池,则控制第一探针向该云池的第二探针发送探测包,以使第二探针根据探测包生成返回包,并将返回包发送给第一探针;

[0174] 相应地,确定模块41,在确定当前服务探测结果时,具体用于:

[0175] 根据探测包及返回包的收发结果确定当前服务探测结果。

[0176] 可选地,要求监测模块42,用于周期性监测目标业务应用与目标云资源间的通信路径是否满足SLA要求。路径调整模块43,用于若确定目标业务应用与目标云资源间的通信路径不满足SLA要求,则将目标业务应用与目标云资源间的通信路径调整到最优通信路径。

[0177] 本实施例提供的云资源分配装置可以执行图3所示方法实施例的技术方案,其实现原理和技术效果类似,此处不再赘述。

[0178] 实施例五

[0179] 图6为本发明一实施例提供的电子设备的结构示意图,如图6所示,本发明实施例提供的电子设备50,包括:存储器51,处理器52以及计算机程序。

[0180] 其中,计算机程序存储在存储器51中,并被配置为由处理器52执行以实现本发明实施例一或实施例二中的方法。

[0181] 相关说明可以对应参见图2至图3的步骤所对应的相关描述和效果进行理解,此处不做过多赘述。

[0182] 其中,本实施例中,存储器51和处理器52通过总线连接。

[0183] 本发明实施例还提供一种计算机可读存储介质,其上存储有计算机程序,计算机程序被处理器执行以实现本发明实施例一或实施例二中的方法。

[0184] 本发明实施例还提供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实施例一或实施例二中的方法。

[0185] 在本发明所提供的几个实施例中,应该理解到,所揭露的装置和方法,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,模块的划分,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,例如多个模块或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另一点,所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口,装置或模块的间接耦合或通信连接,可以是电性,机械或其它的形式。

[0186] 作为分离部件说明的模块可以是或者也可以不是物理上分开的,作为模块显示的部件可以是或者也可以不是物理模块,即可以位于一个地方,或者也可以分布到多个网络模块上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。

[0187] 另外,在本发明各个实施例中的各功能模块可以集成在一个处理模块中,也可以是各个模块单独物理存在,也可以两个或两个以上模块集成在一个模块中。上述集成的模块既可以采用硬件的形式实现,也可以采用硬件加软件功能模块的形式实现。

[0188] 用于实施本发明的方法的程序代码可以采用一个或多个编程语言的任何组合来编写。这些程序代码可以提供给通用计算机、专用计算机或其他可编程数据处理装置的处理器或控制器,使得程序代码当由处理器或控制器执行时使流程图和/或框图中所规定的功能/操作被实施。程序代码可以完全在机器上执行、部分地在机器上执行,作为独立软件包部分地在机器上执行且部分地在远程机器上执行或完全在远程机器或服务器上执行。

[0189] 在本发明的上下文中,机器可读介质可以是有形的介质,其可以包含或存储以供指令执行系统、装置或设备使用或与指令执行系统、装置或设备结合地使用的程序。机器可读介质可以是机器可读信号介质或机器可读储存介质。机器可读介质可以包括但不限于电子的、磁性的、光学的、电磁的、红外的、或半导体系统、装置或设备,或者上述内容的任何合适组合。机器可读存储介质的更具体示例会包括基于一个或多个线的电气连接、便携式计算机盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦除可编程只读存储器(EPROM或快闪存储器)、光纤、便捷式紧凑盘只读存储器(CD-ROM)、光学储存设备、磁储存设备、或上述内容的任何合适组合。

[0190] 此外,虽然采用特定次序描绘了各操作,但是这应当理解为要求这样操作以所示出的特定次序或以顺序次序执行,或者要求所有图示的操作应被执行以取得期望的结果。在一定环境下,多任务和并行处理可能是有利的。同样地,虽然在上面论述中包含了若干具

体实现细节,但是这些不应当被解释为对本公开的范围的限制。在单独的实施例的上下文中描述的某些特征还可以组合地实现在单个实现中。相反地,在单个实现的上下文中描述的各种特征也可以单独地或以任何合适的子组合的方式实现在多个实现中。

[0191] 尽管已经采用特定于结构特征和/或方法逻辑动作的语言描述了本主题,但是应当理解所附权利要求书中所限定的主题未必局限于上面描述的特定特征或动作。相反,上面所描述的特定特征和动作仅仅是实现权利要求书的示例形式。

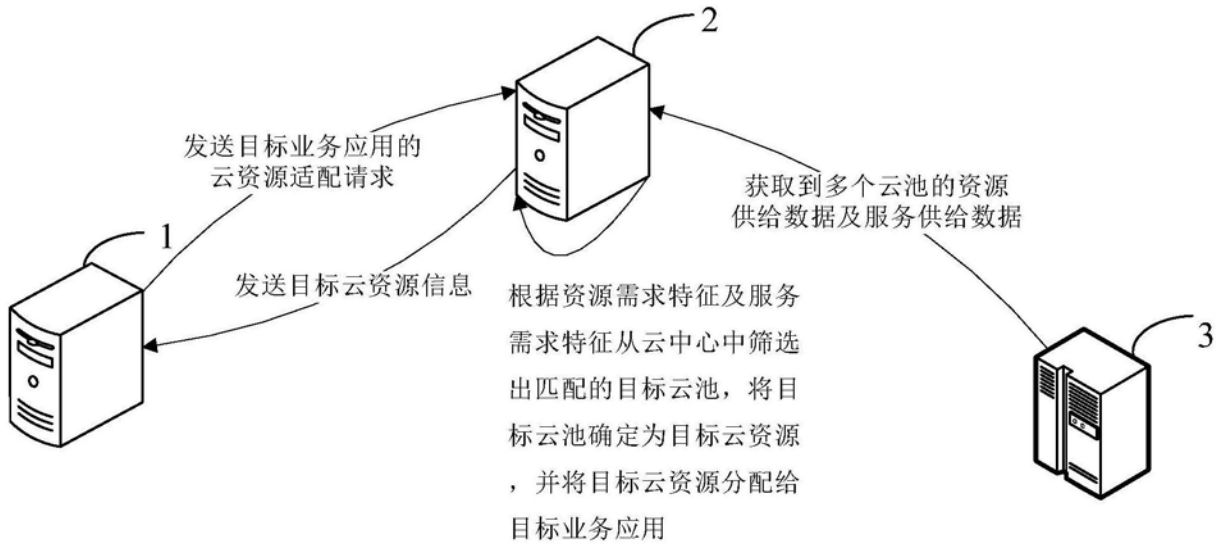


图1

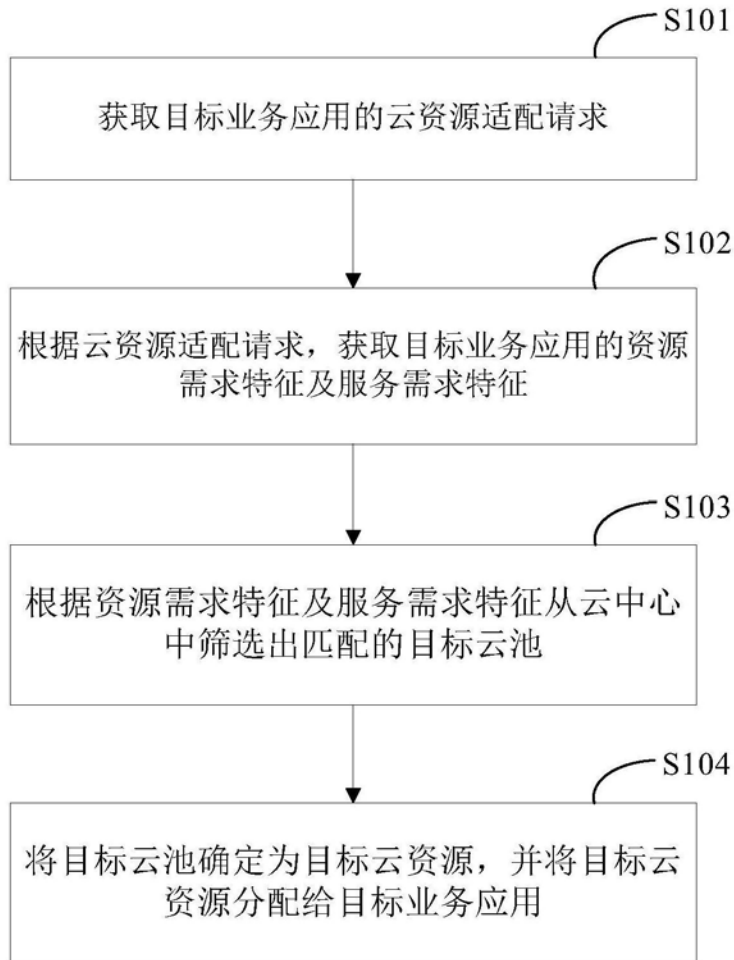


图2

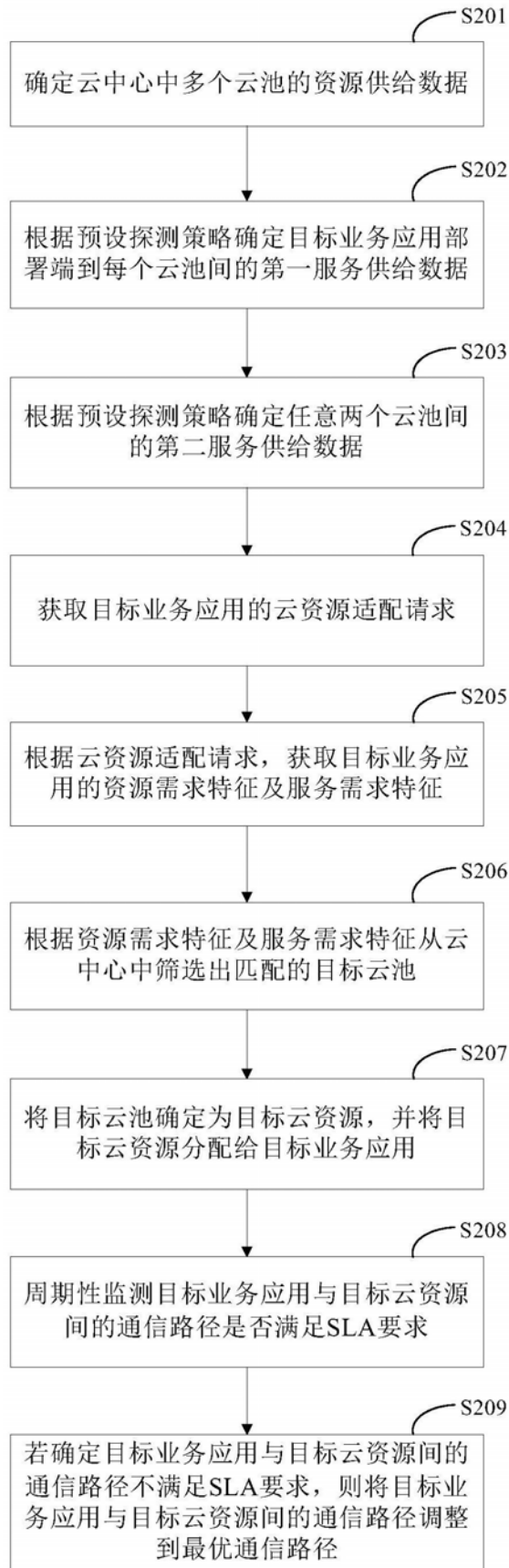


图3

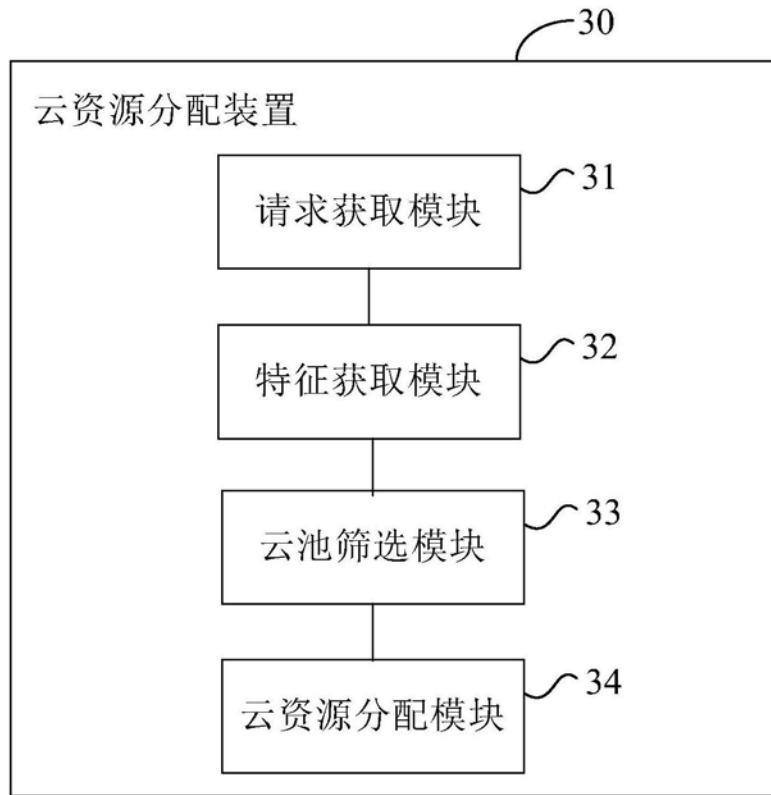


图4

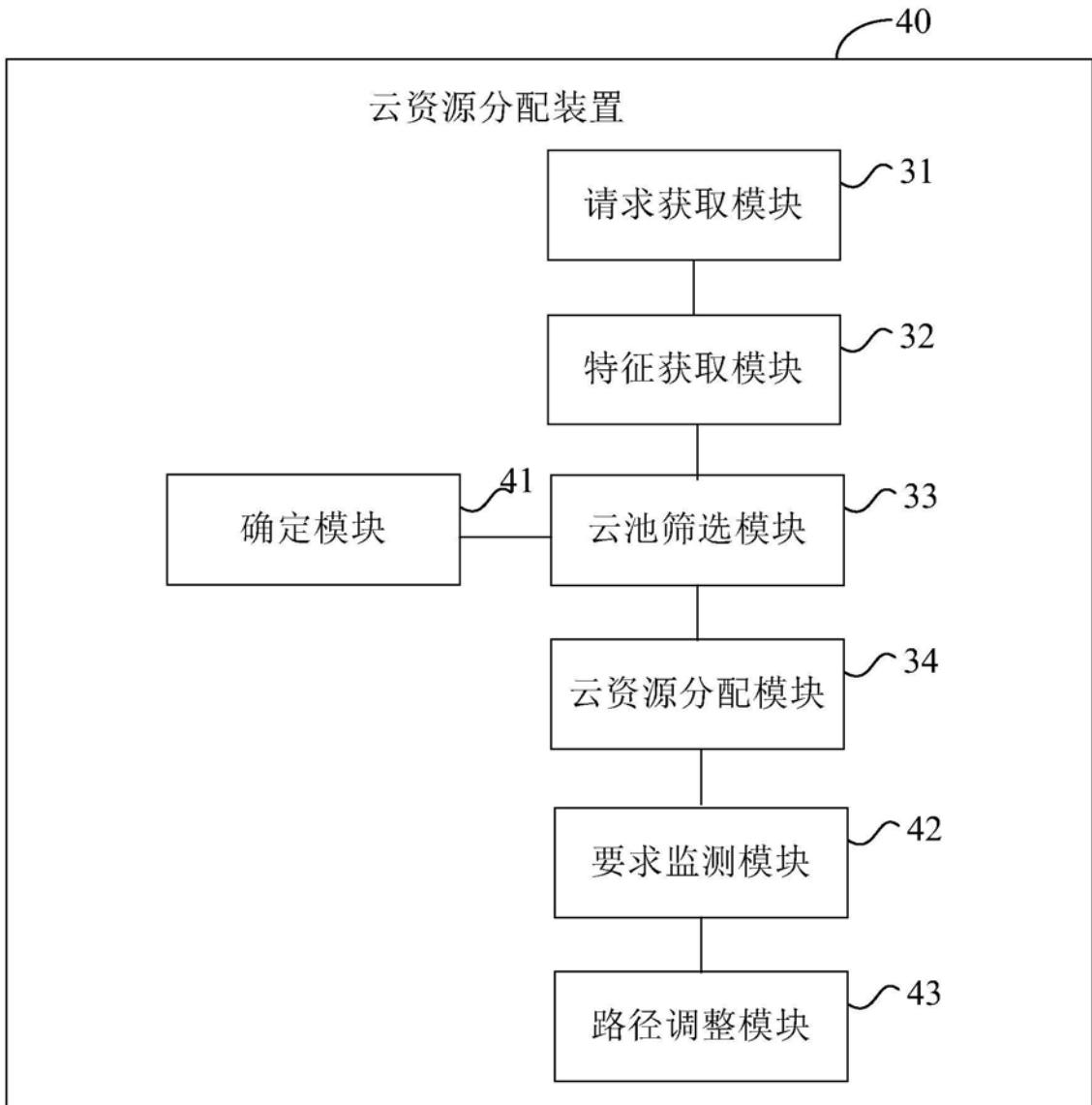


图5

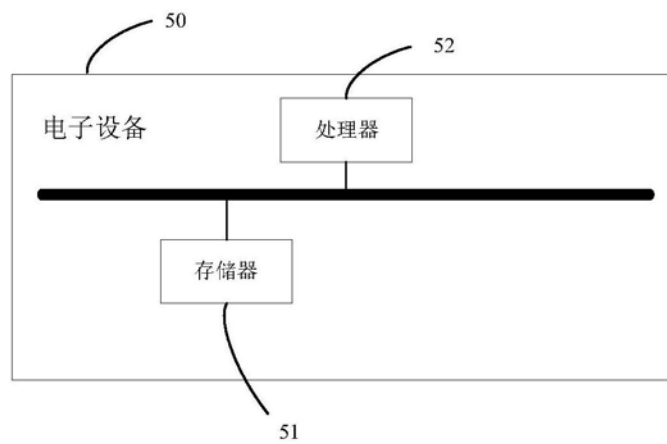


图6