(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国 际 局

(43) 国际公布日
2020 年 7 月 16 日 (16.07.2020)

WIPO | PCT

(10) 国际公布号
WO 2020/143164 A1

(51) 国际专利分类号:
G06F 9/455 (2006.01)

(21) 国际申请号: PCT/CN2019/091499

(22) 国际申请日: 2019 年 6 月 17 日 (17.06.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910017051.8 2019年1月8日 (08.01.2019) CN

(71) 申请人: 平安科技（深圳）有限公司 (PING AN TECHNOLOGY(SHENZHEN)CO., LTD.) [CN/CN]; 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。

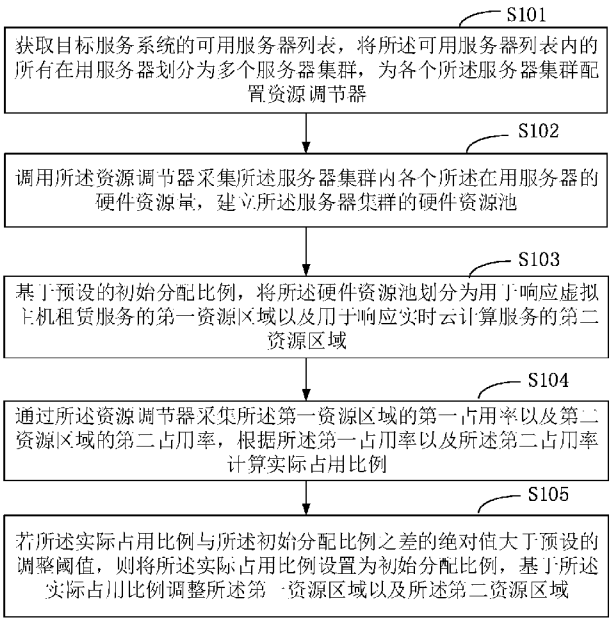
(72) 发明人: 易仁杰 (YI, Renjie); 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。 张伟新 (ZHANG, Weixin); 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。

(74) 代理人: 深圳中一联合知识产权代理有限公司 (SHENZHEN ZHONGYI UNION INTELLECTUAL PROPERTY AGENCY CO., LTD.); 中国广东省深圳市福田区园岭街道深南中路1014号报春大厦9楼（5号信箱）, Guangdong 518028 (CN)。

(81) 指定国 (除另有指明，要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS,

(54) **Title:** NETWORK RESOURCE ALLOCATION METHOD AND DEVICE

(54) **发明名称:** 一种网络资源的分配方法及设备



S101 Acquire an available server list of a target service system, divide all in-use servers in the available server list into a plurality of server clusters, and configure a resource regulator for each of the server clusters

S102 Call the resource regulator to collect a hardware resource quantity of each of the in-use servers in the server cluster, and establish a hardware resource pool of the server cluster

S103 Divide, based on a pre-set initial allocation proportion, the hardware resource pool into a first resource region for responding to a virtual host rental service and a second resource region for responding to a real-time cloud computing service

S104 Collect a first occupation rate of the first resource region and a second occupation rate of the second resource region by means of the resource regulator, and calculate an actual occupation proportion according to the first occupation rate and the second occupation rate

S105 If an absolute value of the difference between the actual occupation proportion and the initial allocation proportion is greater than a pre-set adjustment threshold value, set the actual occupation proportion to be the initial allocation proportion, and adjust the first resource region and the second resource region based on the actual occupation proportion

图 1

(57) **Abstract:** The present application is applicable to the technical field of the Internet, and provides a network resource allocation method and device. The method comprises: acquiring an available server list of a target service system, dividing the available server list into a plurality of server clusters, and configuring a resource regulator for each of the server clusters; calling the resource regulator to collect a hardware resource quantity of each in-use server in the server cluster, and establishing a hardware resource pool of the server cluster; dividing the hardware resource pool into a first resource region and a second resource region based on a pre-set initial

WO 2020/143164 A1

[见续页]

JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国(除另有指明，要求每一种可提供的地区保护)：ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

allocation proportion; calculating an actual occupation proportion according to a first occupation rate and a second occupation rate; and if an absolute value of the difference between the two is greater than a pre-set adjustment threshold value, setting the actual occupation proportion to be the initial allocation proportion, and adjusting the first resource region and the second resource region. In the present application, hardware resources in a server cluster are managed by means of a resource regulator, and two resource regions of service are dynamically allocated, such that the resource utilization rate is improved.

(57) 摘要：本申请适用于互联网技术领域，提供了一种网络资源的分配方法及设备，包括：获取目标服务系统的可用服务器列表，划分为多个服务器集群，为各个服务器集群配置资源调节器；调用资源调节器采集服务器集群内各个在用服务器的硬件资源量，建立服务器集群的硬件资源池；基于预设的初始分配比例，将硬件资源池划分为第一资源区域以及第二资源区域；根据第一占用率以及第二占用率计算实际占用比例；若两者之差的绝对值大于预设的调整阈值，则将实际占用比例设置为初始分配比例，调整第一资源区域以及第二资源区域。本申请通过资源调节器管理服务器集群内的硬件资源，并动态分配两个服务的资源区域，提高了资源的利用率。

一种网络资源的分配方法及设备

本申请申明享有2019年01月08日递交的申请号为201910017051.8、名称为“一种网络资源的分配方法及设备”中国专利申请的优先权，该中国专利申请的整体内容以参考的方式结合在本申请中。

技术领域

本申请属于互联网技术领域，尤其涉及一种网络资源的分配方法及设备。

背景技术

随着云服务技术的不断发展，越来越多用户通过云服务器来实现存储区域的扩容以及本地运算能力补强等功能，例如通过租赁云服务器来搭建不同的系统，从而完成对应的功能测试，或者通过云服务器系统的快速实施运算能力来进行程序编译或压测等。在无需购置额外的硬件资源的情况下，通过云服务实现资源云共享。

现有的云服务器技术中，虚拟机租赁以及实时计算服务是通过两套独立的服务器集群来响应用户的服务请求，而由于用户的需求是动态变化的，部分时刻云服务器租赁请求较少，而该时刻实时计算服务的请求较多，上述两套系统相互独立的情况，则不利于资源分配，从而降低了系统的资源利用率，无法满足用户的服务需求。

技术问题

有鉴于此，本申请实施例提供了一种网络资源的分配方法及设备，以解决现有的云服务器技术中，虚拟机租赁以及实时计算服务是通过两套独立的服务器集群来响应用户的服务请求，系统的资源利用率低，无法满足用户的服务需求的问题。

技术解决方案

本申请实施例的第一方面提供了一种网络资源的分配方法，包括：

获取目标服务系统的可用服务器列表，将所述可用服务器列表内的所有在用服务器划分为多个服务器集群，为各个所述服务器集群配置资源调节器；

调用所述资源调节器采集所述服务器集群内各个所述在用服务器的硬件资源量，建立所述服务器集群的硬件资源池；

基于预设的初始分配比例，将所述硬件资源池划分为用于响应虚拟主机租赁服务的第一资源区域以及用于响应实时云计算服务的第二资源区域；

通过所述资源调节器采集所述第一资源区域的第一占用率以及第二资源区域的第二占用率，根据所述第一占用率以及所述第二占用率计算实际占用比例；

若所述实际占用比例与所述初始分配比例之差的绝对值大于预设的调整阈值，则将所述实际占用比例设置为初始分配比例，基于所述实际占用比例调整所述第一资源区域以及所述第二资源区域。

有益效果

本申请实施例通过将目标服务系统内的所有在用服务器划分为多个服务器集群，并为每个服务器集群配置对应的资源调节器，继而可以通过资源调节器根据初始分配比例对服务器集群内的硬件资源进行划分，在运行过程中，资源调节器可以采集各个资源区域的实际占用率，并根据第一占用率以及第二占用率确定实际占用比例，从而判断当前的资源分配是否合理，若实际占用比例与初始分配比例不一致，则进行动态调整，从而实现了对网络资源进行动态分配的目的。与现有的网络资源的分配方式相比，用于虚拟主机租赁的服务器以及用于实时云计算的服务器布放于同一服务器集群内，甚至同一服务器也能够响应两种不同的服务请求，通过资源调节器管理服务器集群内的硬件资源，并动态分配两个服务的资源区域，提高了资源的利用率。

附图说明

图 1 是本申请第一实施例提供的一种网络资源的分配方法的实现流程图；

图 2 是本申请第二实施例提供的一种网络资源的分配方法 S104 具体实现流程图；

图 3 是本申请第三实施例提供的一种网络资源的分配方法 S101 具体实现流程图；

图 4 是本申请第四实施例提供的一种网络资源的分配方法具体实现流程图；

图 5 是本申请第五实施例提供的一种网络资源的分配方法具体实现流程图；

图 6 是本申请一实施例提供的一种网络资源的分配设备的结构框图；

图 7 是本申请另一实施例提供的一种终端设备的示意图。

本发明的实施方式

在本申请实施例中，流程的执行主体为终端设备。该终端设备包括但不限于：服务器、计算机、智能手机以及平板电脑等能够执行网络资源的分配操作的设备。特别地，该网络资源的分配设备可以为一业务服务系统的上位机设备，用于管理该业务服务系统内所有服务器的运行以及异常识别等，对该业务服务系统的拓扑结构进行管理。图 1 示出了本申请第一实施例提供的网络资源的分配方法的实现流程图，详述如下：

在 S101 中，获取目标服务系统的可用服务器列表，将所述可用服务器列表内的所有在用服务器划分为多个服务器集群，为各个所述服务器集群配置资源调节器。

在本实施例中，业务服务系统内包含多个在用服务器，通过在用服务器响应各个用户

终端发起的业务请求，业务服务系统可以根据地理区块来对在用服务器进行划分，即用户终端发送业务请求时会携带有该用户终端所处的位置信息，继而网关设备会基于该位置信息确定对应的在用服务器，重新对该业务请求进行重定向，从而能够定向发送给该地理区块所处理的在用服务器；当然，业务服务系统还可以根据各个在用服务器的负载情况来分配服务请求。业务服务系统可以以预设的检测周期获取各个服务器当前的运行状态，将正常运行且存在空闲硬件资源的在用服务器添加到可用服务器列表内，即表示该可用服务器列表内的在用服务器可以用于响应用户终端发起的业务请求。

在本实施例中，终端设备，即硬件资源的分配设备，可以从目标服务系统的数据库内获取得到该可用服务器列表，从而基于目标服务系统对于在用服务器的预设划分方式，将所有在用服务器划分为多个服务器集群，每个服务器集群中至少包含一个在用服务器。如上所述，若目标服务器系统是基于地理区块对在用服务器进行划分的，则终端设备可以基于地理区块将属于同一地理区块或相邻地理区块的在用服务器集成于同一个服务器集群内。若目标服务系统是基于负载情况对在用服务器进行划分的，则终端设备可以根据各个在用服务器当前的负载情况划分为多个服务器集群。若目标服务系统是基于服务类型将业务请求分配给不同在用服务器的，则终端设备可以将用于响应不同类型的服务器各选一个，集成于一个服务器集群内，从而该服务器集群可以响应所有服务类型的业务请求。

在本实施例中，终端设备可以用于管理多个不同业务服务系统的网络资源，若某一业务服务系统满足分配触发条件，则终端设备会对该业务服务系统识别为目标服务系统，并对该目标服务系统内的在用服务器进行网络资源的分配操作。例如，若某一业务服务系统进行扩容操作，则加入了新的在用服务器，则终端设备会对该业务服务系统内的在用服务器重新进行网络资源划分。又例如，终端设备会设置有一检测周期，若终端设备检测到当前时刻到达预设的检测周期的时间节点，则触发 S1 的操作。

在本实施例中，为了便于对不同集群进行并发管理，从而提高网络资源的管理及时性以及准确性，终端设备会为每个服务器集群配置一个资源调节器。该资源调节器可以为一独立的实体装置，也可以为一虚拟装置。若该资源调节器为一虚拟装置，则终端设备可以在服务器集群中选取一个在用服务器作为目标服务器，例如根据各个在用服务器的硬件资源参数，确定出处理能力评分最优的一个在用服务器作为目标服务器。终端设备会控制该目标服务器，并在该目标服务器的主线程下创建一条子线程，并基于预设的资源调节程序将该子线程设置为用于运行资源调节进程的虚拟装置，完成资源调节器的配置操作。

在 S102 中，调用所述资源调节器采集所述服务器集群内各个所述在用服务器的硬件资源量，建立所述服务器集群的硬件资源池。

在本实施例中，终端设备在为了方便对服务器集群的网络资源进行分配，会将该服务器集群内所包含的硬件资源识别为一个整体，从而终端设备只需对硬件资源池这一统一的整体进行资源划分，而无需对单一的服务器进行资源划分，能够提高管理的效率。基于此，终端设备需要获取各个在用服务器的硬件资源量，为每个在用服务器的硬件资源量配置一个关于服务器集群的资源标识码，并建立一个资源标识码的寻址对应表，从而可以根据寻址对应表确定资源标识码所对应的在用服务器以及在该在用服务器的资源地址。终端设备将所有资源标识码所构成的硬件资源识别为关于该服务器集群的硬件资源池。

在本实施例中，该硬件资源包括有内存资源、带宽资源、线程资源等与硬件处理能力相关的资源，由于上述硬件资源均可以被其他用户终端进行在线占用，即属于该业务服务系统的可用网络资源。终端设备可以为不同类型的硬件资源配置对应的硬件资源池。

在 S103 中，基于预设的初始分配比例，将所述硬件资源池划分为用于响应虚拟主机租赁服务的第一资源区域以及用于响应实时云计算服务的第二资源区域。

在本实施例中，虚拟主机服务主要为租赁服务，即用户可以在目标服务系统中选取一个目标服务器，在该目标服务器中创建一个虚拟主机，并搭建相应的系统，在申请时会设置资源需求，并在租赁期内，该虚拟主机的资源将被一直占用，无法响应其他服务，是无法时分复用的占用方式。而实时云计算服务，是用户将所需运行的大数据分析、程序代码、测试实例等上传至云服务的服务器，通过云服务的服务器运行上述对象，并输出对应的运行结果，通过 OBS、EFS 或数据库等方式将运行结果进行存储，运行完毕后，所占用的资源将会被释放，是一个可时分复用的资源占用方式。基于上述两类不同的服务类型，需要将服务器集群的硬件资源进行划分，即云服务系统内的各个服务器集群既可响应虚拟主机服务又可以响应实时云计算服务，两类服务在物理层面上是统一管理的，从而便于实现资源的弹性分配。

在本实施例中，终端设备配置有一初始分配比例，即用于响应响应云计算服务的网络资源与用于响应虚拟主机租赁服务的网络资源之间的比值。该初始分配比例可以由用户手动进行配置。继而终端设备可以根据该初始分配比例，将硬件资源池的硬件资源进行划分，得到第一资源区域以及第二资源区域。

可选地，终端设备可以获取历史业务记录，基于历史业务记录统计虚拟主机的服务次数以及云计算服务的服务次数，并确定虚拟主机服务的平均占用资源量以及云计算服务的平均占用资源量，基于上述四个参数，即虚拟主机租赁的次数、虚拟主机租赁的平均占用资源量、云计算服务的服务次数以及云计算服务的平均占用资源量，计算初始比例。

可选地，终端设备可以将各个历史业务记录导入到调整好的 RNN 神经网络中，通过神

经网络确定该初始比例。其中，历史业务记录包括有服务类型，即用于确定该历史业务记录属于虚拟主机租赁服务抑或是云计算服务，服务时长、资源占用量等信息，从而神经网络可以预测到当前时刻目标服务系统中两个服务的比例，提高初始分配比例的准确性。

在 S104 中，通过所述资源调节器采集所述第一资源区域的第一占用率以及第二资源区域的第二占用率，根据所述第一占用率以及所述第二占用率计算实际占用比例。

在本实施例中，终端设备对硬件资源池进行划分后，即每个服务器内对应区域的硬件资源用于响应何种服务类型已经固定了，当然，若某一服务器中部分硬件资源被划分到第一占用区域，而另一部分的硬件资源被划分至第二占用区域，则该在用服务器可以用于响应云租赁服务以及云计算服务，具体在用服务器响应的服务类型可以基于本地的硬件资源的属性所决定。

在本实施例中，终端设备可以接收资源调节器反馈关于第一资源区域的第一占用率以及关于第二资源区域的第二占用率，从而可以实时确定初始分配比例是否合理，继而根据第一占用率以及第二占用率即可以确定用于响应两个类型服务的实际占用比例。若该实际占用比例与初始比例之间的差值的绝对值小于或等于预设的调整阈值，则表示与预设分配比例较为接近，在可接受的浮动范围内，此时，则无需调整两个网络资源的分配比例，等待下一个采集时刻的到达，再进行调整判断；反之，若两个之间的差值的绝对值大于预设的调整阈值，则需要对硬件资源池进行重新分配，执行 S105 的相关操作。

在 S105 中，若所述实际占用比例与所述初始分配比例之差的绝对值大于预设的调整阈值，则将所述实际占用比例设置为初始分配比例，基于所述实际占用比例调整所述第一资源区域以及所述第二资源区域。

在本实施例中，终端设备在检测到某一服务器集群中实际占用比例与初始分配比例之差的绝对值大于预设的调整阈值时，则需要重新对该服务器集群中的硬件资源池的硬件资源进行重新划分，重新划分的操作如下：终端设备会根据实际占用比例确定所需调整的硬件资源量，并检测第一资源区域以及第二资源区域中包含的空闲硬件资源量，基于调整的硬件资源量以及空闲的硬件资源量，调整两个资源区域所包含的资源量，以使调整的第一资源区域的硬件资源量与第二资源区域的硬件资源量之间的比值满足实际占用比例。

在本实施例中，为了在下一次检测周期到达时能够避免重复调整以及后续扩容操作时能够合理地进行初始资源分配，终端设备会将本次获取得到的实际占用比例设置为初始分配比例，从而能够实现动态调整初始分配比例的目的。

以上可以看出，本申请实施例提供的一种网络资源的分配方法通过将目标服务系统内的所有在用服务器划分为多个服务器集群，并为每个服务器集群配置对应的资源调节器，

继而可以通过资源调节器根据初始分配比例对服务器集群内的硬件资源进行划分，在运行过程中，资源调节器可以采集各个资源区域的实际占用率，并根据第一占用率以及第二占用率确定实际占用比例，从而判断当前的资源分配是否合理，若实际占用比例与初始分配比例不一致，则进行动态调整，从而实现了对网络资源进行动态分配的目的。与现有的网络资源的分配方式相比，用于虚拟主机租赁的服务器以及用于实时云计算的服务器布放于同一服务器集群内，甚至同一服务器也能够响应两种不同的服务请求，通过资源调节器管理服务器集群内的硬件资源，并动态分配两个服务的资源区域，提高了资源的利用率。

图2示出了本申请第二实施例提供的一种网络资源的分配方法S104的具体实现流程图。参见图2，相对于图1所述实施例，本实施例提供的一种网络资源的分配方法S104包括：S1041~S1044，具体详述如下：

进一步地，所述通过所述资源调节器采集所述第一资源区域的第一占用率以及第二资源区域的第二占用率，根据所述第一占用率以及所述第二占用率计算实际占用比例，包括

在S1041中，若满足预设的采集条件，则以预设的采集频率获取所述服务器集群内各个所述在用服务器的多个实际资源占用参数；所述实际资源占用参数包括：用于响应虚拟主机租赁服务的第一占用参量以及用于响应实时云计算服务的第二占用参量。

在本实施例中，终端设备配置有实际资源占用参数的采集条件，若检测到当前时刻或当前目标服务系统的网络状态满足该预设的采集条件，则执行S1041的相关操作。具体地，该采集条件可以为一时间条件，即终端设备配置有一采集周期或多个采集节点，若检测到当前时刻满足该时间条件，则会触发实际资源占用参数的采集流程，判断是否需要对分配比例进行调整。可选地，该采集条件可以为实际占用量，即终端设备根据第一资源区域以及第二资源区域分配的资源量设置有多个资源量节点，若检测到当前时刻某一资源区域的实际占用量到达对应的资源量节点则执行S1041的相关操作。管理员可以通过发送采集指令等手动触发的方式，来执行S1041的相关操作。

在本实施例中，终端设备在确定了当前满足采集条件时，可以通过各个服务器集群的资源调节器获取其对应集群内所有在用服务器的实际资源占用参数，从而实现多线程并发采集，提高实际资源占用参数的采集效率。另一方面，为了提高采集的准确性，特别对于实时云计算服务，其瞬时资源占用率的浮动范围较大，因此，终端设备可以控制资源调节器以预设的采集频率对每个在用服务器获取多个实时资源占用参数，从而能够提高实时资源占用参数的准确性，减少因瞬时浮动而带来的影响。

由于虚拟主机租赁服务以及实时云计算服务所占用的资源是基于整个服务器集群的硬件资源池进行分配的，而具体集群里每个在用服务器的硬件资源分配对于终端设备而言是

透明的，因此为了统计整个集群的实时资源分配情况，终端设备在采集当前时刻的实际占用资源量时，会检测该在用服务器关于两种服务分别占用的实时资源量，从而可以通过集群里所有在用服务器的实际占用资源量确定出整个集群的实际占用比例。

在 S1042 中，根据各个所述在用服务器的多个所述第一占用参量，计算所述第一资源区域的第一占用均值以及第一占用标准差。

在 S1043 中，根据各个所述在用服务器的多个所述第二占用参量，计算所述第二资源区域的第二占用均值以及第二占用标准差。

在本实施例中，终端设备获取了关于服务器集群内各个在用服务器在多个采集周期的实际资源占用率，因此可以根据所有在用服务器在同一采集周期内的第一占用参量，确定该采集周期关于第一资源区域的总的实际占用参量。同样地，对于各个采集周期都可以将所有在用服务器在该周期内的第一占用参量进行叠加，从而求出关于该周期内第一资源区域的总的实际占用参量。因此，终端设备可以计算出第一资源区域在基于多个不同的采集周期对应的第一占用均值以及第一占用标准差。对于第二资源区域也可以通过上述方式进行计算，在此不再赘述。

在 S1044 中，将所述第一占用均值、所述第一占用标注差、所述第二占用均值以及所述第二占用标准差导入实际占用率计算模型，确定所述实际占用比例；所述实际占用率计算模型具体为：

$$ActualRate = \frac{(\overline{VM} + \xi_{VM}) \cdot Countmax}{(\overline{Count} + \xi_{Count}) \cdot VMmax}$$

其中，$ActualRate$ 为所述实际占用比例；$\overline{VM}$ 为第一占用均值；ξ_{VM} 为第一占用标准差；$\overline{Count}$ 为第二占用均值；ξ_{Count} 为第二占用标准差；$VMmax$ 为第一资源区域的硬件资源量；$Countmax$ 为第二资源区域的硬件资源量。

在本实施例中，终端设备将上述计算得到的四个参量导入到实际占用率的计算模型内，确定出关于本次采集操作计算得到实际占用比例。由于该实际占用率不仅考虑了第一占用均值以及第二占用均值，还引入了占用标准差，从而减少浮动率而带来的影响，提高实际占用率的准确性。

在本申请实施例中，在满足采集条件时，获取多个实际资源占用参量，从而能够通过多个实际资源占用参量计算出关于服务器集群的实际资源占用率，从而提高实际资源占用率的准确性。

图 3 示出了本申请第三实施例提供的一种网络资源的分配方法 S101 的具体实现流程图。

参见图 3，相对于图 1 所述的实施例，本实施例提供的一种网络资源的分配方法 S101 包括：S1011~S1013，具体详述如下：

进一步地，所述将所述可用服务器列表内的所有在用服务器划分为多个服务器集群，为各个所述服务器集群配置资源调节器，包括：

在 S1011 中，获取各个所述在用服务器的安装位置，并根据所述安装位置在预设的地图界面上标记各个所述在用服务器。

在本实施例中，终端设备可以基于各个在用服务器的安装位置来进行服务器集群的划分。在该情况下，目标服务系统可以用户终端所述的地理区块，将业务请求分配到该地理区块的服务器集群，从而可以减少中间路由的跳转，从而提高业务响应的速率。

在本实施例中，若在用服务器可以配置有定位模块，则可以通过该定位模块获取本地的安装位置，并将该安装位置反馈给终端设备。若在用服务器并没有配置有定位模块，则可以根据该在用服务器所在的网关地址，基于网关地址确定安装位置。终端设备根据安装位置在预设的地图界面上标记出各个在用服务器，以便对在用服务器进行集群划分。特别地，终端设备可以调用第三方地图应用的 API 接口，在本地显示模块上输出第三方地图应用的界面，并标记出各个在用服务器，从而无需重新编写地图程序，减少所需的开发量。

在 S1012 中，通过预设的集群窗口在所述地图界面上进行遍历框取，将处于同一所述集群窗口内的所有在用服务器识别为隶属于同一所述服务器集群。

在本实施例中，终端设备根据预设的集群窗口，可以通过该集群窗口对在用服务器进行框取，从而同一服务器集群内的服务器所处的安装位置之间的距离会小于或等于预设的距离阈值。具体地，终端设备可以控制集群窗口在地图界面上进行滑动，实现遍历框取，而处于同一集群窗口内的在用服务器则识别为同一服务器集群。

在 S1013 中，为各个所述服务器集群配置所述资源调节器。

在本实施例中，为了便于对不同集群进行并发管理，从而提高网络资源的管理及时性以及准确性，终端设备会为每个服务器集群配置一个资源调节器。

在本申请实施例中，终端设备根据各个在用服务器的安装位置来实现对服务器集群的划分，从而能够提高后续业务请求的分配操作，提高业务请求的分配效率。

图 4 示出了本申请第四实施例提供的一种网络资源的分配方法的具体实现流程图。参见图 4，相对于图 1 至图 3 所述实施例，本实施例提供的一种网络资源的分配方法中在所述根据所述第一占用率以及所述第二占用率计算实际占用比例之后，还包括：S401~S402，具体详述如下：

在 S401 中，将所述初始分配比例以及各个采集时刻获取得到的所述实际占用比例作为

训练样本，将所述训练样本导入多层前馈神经网络，计算出预期分配比例。

在本实施例中，终端设备预设有多层前馈 RNN 神经网路，可以根据预设的初始分配比例以及各个周期的实际占用比例来提高初始分配比例的准确性，因此，终端设备会基于初始分配比例以及实际占用比例生成多个训练样本，并将多个训练样本导入到该 RNN 神经网络内，计算出预期分配比例。若检测到新的服务器集群或接收到扩容请求时，可以根据该预期分配比例来对硬件资源池进行划分。

需要说明的是，由于各个服务器集群的业务请求的数量以及服务类型存在差异，即该预期分配比例对于不同的服务器集群而言也是不同的，从而能够与该服务器集群的用户终端的服务请求情况相匹配。

在 S402 中，若接收到系统扩容指令，则确定扩容服务器的硬件资源量，并根据所述预期分配比例对所述扩容服务器的硬件资源量执行硬件资源划分操作。

在本实施例中，终端设备若接收到系统扩容指令，即该服务器集群添加有新的扩容服务器，需要对该扩容服务器的硬件资源进行划分。因此，终端设备则可以根据预期分配比例以及该扩容服务器的硬件资源量，确定出用于云主机租赁服务的第一扩容资源量以及用于实时云计算服务的第二扩容资源量，并基于上述两个扩容资源量对硬件资源进行划分。

在本申请实施例中，根据初始分配比例以及多个实时占用比例预测出预期分配比例，并基于该预期分配比例对扩容服务期间进行划分操作，从而提高划分操作的准确率。

图 5 示出了本申请第五实施例提供的一种网络资源的分配方法的具体实现流程图。参见图 5，相对于图 1-图 3 所述实施例，本实施例提供的一种网络资源的分配方法还包括：S501~S502，具体详述如下：

在 S501 中，向各个所述资源调节器广播资源上报指令，以使各个所述资源调节器采集对应的所述服务器集群的资源占用率。

在本实施例中，终端设备为了判断各个服务器集群是否需要进行服务器扩容，会向各个资源调节器发送一个资源上报指令，通过各个资源调节器采集关于自身所属的服务器集群内的资源占用情况。终端设备可以设置有触发条件，例如条件触发以及事件触发，若当前满足预设的触发条件，则执行 S501 的相关操作。

在 S502 中，若任一所述资源占用率均大于预设的扩容阈值，则向管理员的终端发送扩容提示信息。

在本实施例中，终端设备若检测到资源占用率小于或等于预设的扩容阈值，则表示该服务器集群的网络资源并未处于饱和状态，可以继续响应用户终端的业务请求；反之，若该资源占用率大于扩容阈值，则表示该服务器已处于饱和状态，则生成一个扩容提示信息，

并发送给管理员的终端，以便管理员对服务器集群进行扩容。

在本申请实施例中，终端设备检测服务器集群的资源占用情况，并自动生成扩容提示信息，从而提高了扩容的响应效率。

应理解，上述实施例中各步骤的序号的大小并不意味着执行顺序的先后，各过程的执行顺序应以其功能和内在逻辑确定，而不应对本申请实施例的实施过程构成任何限定。

图 6 示出了本申请一实施例提供的一种网络资源的分配设备的结构框图，该网络资源的分配设备包括的各单元用于执行图 1 对应的实施例中的各步骤。具体请参阅图 1 与图 1 所对应的实施例中的相关描述。为了便于说明，仅示出了与本实施例相关的部分。

参见图 6，所述网络资源的分配设备包括：

资源调节器配置单元 61，用于获取目标服务系统的可用服务器列表，将所述可用服务器列表内的所有在用服务器划分为多个服务器集群，为各个所述服务器集群配置资源调节器；

硬件资源池建立单元 62，用于调用所述资源调节器采集所述服务器集群内各个所述在用服务器的硬件资源量，建立所述服务器集群的硬件资源池；

硬件资源池划分单元 63，用于基于预设的初始分配比例，将所述硬件资源池划分为用于响应虚拟主机租赁服务的第一资源区域以及用于响应实时云计算服务的第二资源区域；

实际占用比例获取单元 64，用于通过所述资源调节器采集所述第一资源区域的第一占用率以及第二资源区域的第二占用率，根据所述第一占用率以及所述第二占用率计算实际占用比例；

硬件资源调整单元 65，用于若所述实际占用比例与所述初始分配比例之差的绝对值大于预设的调整阈值，则将所述实际占用比例设置为初始分配比例，基于所述实际占用比例调整所述第一资源区域以及所述第二资源区域。

可选地，所述实际占用比例获取单元 64 包括：

采集触发单元，用于若满足预设的采集条件，则以预设的采集频率获取所述服务器集群内各个所述在用服务器的多个实际资源占用参数；所述实际资源占用参数包括：用于响应虚拟主机租赁服务的第一占用参量以及用于响应实时云计算服务的第二占用参量；

第一占用参量计算单元，用于根据各个所述在用服务器的多个所述第一占用参量，计算所述第一资源区域的第一占用均值以及第一占用标准差；

第二占用参量计算单元，用于根据各个所述在用服务器的多个所述第二占用参量，计算所述第二资源区域的第二占用均值以及第二占用标准差；

实际占用比例计算单元，用于将所述第一占用均值、所述第一占用标注差、所述第二

占用均值以及所述第二占用标准差导入实际占用率计算模型，确定所述实际占用比例；所述实际占用率计算模型具体为：

$$ActualRate=\frac{(\overline{VM}+\xi_{VM})\cdot Countmax}{(\overline{Count}+\xi_{Count})\cdot VMmax}$$

其中，$ActualRate$ 为所述实际占用比例；$\overline{VM}$ 为第一占用均值；ξ_{VM} 为第一占用标准差；$\overline{Count}$ 为第二占用均值；ξ_{Count} 为第二占用标准差；$VMmax$ 为第一资源区域的硬件资源量；$Countmax$ 为第二资源区域的硬件资源量。

可选地，所述资源调节器配置单元 61 包括：

安装位置获取单元，用于获取各个所述在用服务器的安装位置，并根据所述安装位置在预设的地图界面上标记各个所述在用服务器；

服务器集群识别单元，用于通过预设的集群窗口在所述地图界面上进行遍历框取，将处于同一所述集群窗口内的所有在用服务器识别为隶属于同一所述服务器集群；

资源调节器创建单元，用于为各个所述服务器集群配置所述资源调节器。

可选地，所述网络资源的分配设备还包括：

预期分配比例计算单元，用于将所述初始分配比例以及各个采集时刻获取得到的所述实际占用比例作为训练样本，将所述训练样本导入多层前馈神经网络，计算出预期分配比例；

扩容操作响应单元，用于若接收到系统扩容指令，则确定扩容服务器的硬件资源量，并根据所述预期分配比例对所述扩容服务器的硬件资源量执行硬件资源划分操作。

可选地，所述网络资源的分配设备还包括：

资源占用率采集单元，用于向各个所述资源调节器广播资源上报指令，以使各个所述资源调节器采集对应的所述服务器集群的资源占用率；

扩容提示信息发送单元，用于若任一所述资源占用率均大于预设的扩容阈值，则向管理员的终端发送扩容提示信息。

因此，本申请实施例提供的网络资源的分配设备中，用于虚拟主机租赁的服务器以及用于实时云计算的服务器布放于同一服务器集群内，甚至同一服务器也能够响应两种不同的服务请求，通过资源调节器管理服务器集群内的硬件资源，并动态分配两个服务的资源区域，提高了资源的利用率。

图 7 是本申请另一实施例提供的一种终端设备的示意图。如图 7 所示，该实施例的终端设备 7 包括：处理器 70、存储器 71 以及存储在所述存储器 71 中并可在所述处理器 70

上运行的计算机可读指令 72，例如网络资源的分配程序。所述处理器 70 执行所述计算机可读指令 72 时实现上述各个网络资源的分配方法实施例中的步骤，例如图 1 所示的 S101 至 S105。或者，所述处理器 70 执行所述计算机可读指令 72 时实现上述各装置实施例中各单元的功能，例如图 6 所示模块 61 至 65 功能。

示例性的，所述计算机可读指令 72 可以被分割成一个或多个单元，所述一个或者多个单元被存储在所述存储器 71 中，并由所述处理器 70 执行，以完成本申请。所述一个或多个单元可以是能够完成特定功能的一系列计算机可读指令指令段，该指令段用于描述所述计算机可读指令 72 在所述终端设备 7 中的执行过程。例如，所述计算机可读指令 72 可以被分割成资源调节器配置单元、硬件资源池建立单元、硬件资源池划分单元、实际占用比例获取单元以及硬件资源调整单元，各单元具体功能如上所述。

所述终端设备 7 可以是桌上型计算机、笔记本、掌上电脑及云端服务器等计算设备。所述终端设备可包括，但不仅限于，处理器 70、存储器 71。本领域技术人员可以理解，图 7 仅仅是终端设备 7 的示例，并不构成对终端设备 7 的限定，可以包括比图示更多或更少的部件，或者组合某些部件，或者不同的部件，例如所述终端设备还可以包括输入输出设备、网络接入设备、总线等。所称处理器 70 可以是中央处理单元（Central Processing Unit，CPU），还可以是其他通用处理器、数字信号处理器（Digital Signal Processor，DSP）、专用集成电路（Application Specific Integrated Circuit，ASIC）、现成可编程门阵列（Field-Programmable Gate Array，FPGA）或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。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。所述存储器 71 可以是所述终端设备 7 的内部存储单元，例如终端设备 7 的硬盘或内存。所述存储器 71 也可以是所述终端设备 7 的外部存储设备，例如所述终端设备 7 上配备的插接式硬盘，智能存储卡（Smart Media Card，SMC），安全数字（Secure Digital，SD）卡，闪存卡（Flash Card）等。进一步地，所述存储器 71 还可以既包括所述终端设备 7 的内部存储单元也包括外部存储设备。所述存储器 71 用于存储所述计算机可读指令以及所述终端设备所需的其他程序和数据。所述存储器 71 还可以用于暂时地存储已经输出或者将要输出的数据。

以上所述实施例仅用以说明本申请的技术方案，而非对其限制；尽管参照前述实施例对本申请进行了详细的说明，本领域的普通技术人员应当理解：其依然可以对前述各实施例所记载的技术方案进行修改，或者对其中部分技术特征进行等同替换；而这些修改或者替换，并不使相应技术方案的本质脱离本申请各实施例技术方案的精神和范围，均应包含在本申请的保护范围之内。

权利要求书

1.一种网络资源的分配方法，其特征在于，包括：

获取目标服务系统的可用服务器列表，将所述可用服务器列表内的所有在用服务器划分为多个服务器集群，为各个所述服务器集群配置资源调节器；

调用所述资源调节器采集所述服务器集群内各个所述在用服务器的硬件资源量，建立所述服务器集群的硬件资源池；

基于预设的初始分配比例，将所述硬件资源池划分为用于响应虚拟主机租赁服务的第一资源区域以及用于响应实时云计算服务的第二资源区域；

通过所述资源调节器采集所述第一资源区域的第一占用率以及第二资源区域的第二占用率，根据所述第一占用率以及所述第二占用率计算实际占用比例；

若所述实际占用比例与所述初始分配比例之差的绝对值大于预设的调整阈值，则将所述实际占用比例设置为初始分配比例，基于所述实际占用比例调整所述第一资源区域以及所述第二资源区域。

2. 根据权利要求 1 所述的分配方法，其特征在于，所述通过所述资源调节器采集所述第一资源区域的第一占用率以及第二资源区域的第二占用率，根据所述第一占用率以及所述第二占用率计算实际占用比例，包括：

若满足预设的采集条件，则以预设的采集频率获取所述服务器集群内各个所述在用服务器的多个实际资源占用参数；所述实际资源占用参数包括：用于响应虚拟主机租赁服务的第一占用参量以及用于响应实时云计算服务的第二占用参量；

根据各个所述在用服务器的多个所述第一占用参量，计算所述第一资源区域的第一占用均值以及第一占用标准差；

根据各个所述在用服务器的多个所述第二占用参量，计算所述第二资源区域的第二占用均值以及第二占用标准差；

将所述第一占用均值、所述第一占用标注差、所述第二占用均值以及所述第二占用标准差导入实际占用率计算模型，确定所述实际占用比例；所述实际占用率计算模型具体为：

$$ActualRate=\frac{(\overline{VM}+\xi_{VM})\cdot Countmax}{(\overline{Count}+\xi_{Count})\cdot VMmax}$$

其中，$ActualRate$ 为所述实际占用比例；$\overline{VM}$ 为第一占用均值；ξ_{VM} 为第一占用标准差；$\overline{Count}$ 为第二占用均值；ξ_{Count} 为第二占用标准差；$VMmax$ 为第一资源区域的硬件资源量；

Countmax 为第二资源区域的硬件资源量。

3. 根据权利要求 1 所述的分配方法，其特征在于，所述将所述可用服务器列表内的所有在用服务器划分为多个服务器集群，为各个所述服务器集群配置资源调节器，包括：

获取各个所述在用服务器的安装位置，并根据所述安装位置在预设的地图界面上标记各个所述在用服务器；

通过预设的集群窗口在所述地图界面上进行遍历框取，将处于同一所述集群窗口内的所有在用服务器识别为隶属于同一所述服务器集群；

为各个所述服务器集群配置所述资源调节器。

4. 根据权利要求 1-3 任一项所述的分配方法，其特征在于，在所述根据所述第一占用率以及所述第二占用率计算实际占用比例之后，还包括：

将所述初始分配比例以及各个采集时刻获取得到的所述实际占用比例作为训练样本，将所述训练样本导入多层前馈神经网络，计算出预期分配比例；

若接收到系统扩容指令，则确定扩容服务器的硬件资源量，并根据所述预期分配比例对所述扩容服务器的硬件资源量执行硬件资源划分操作。

5. 根据权利要求 1-3 任一项所述的分配方法，其特征在于，还包括：

向各个所述资源调节器广播资源上报指令，以使各个所述资源调节器采集对应的所述服务器集群的资源占用率；

若任一所述资源占用率均大于预设的扩容阈值，则向管理员的终端发送扩容提示信息。

6. 一种网络资源的分配设备，其特征在于，包括：

资源调节器配置单元，用于获取目标服务系统的可用服务器列表，将所述可用服务器列表内的所有在用服务器划分为多个服务器集群，为各个所述服务器集群配置资源调节器；

硬件资源池建立单元，用于调用所述资源调节器采集所述服务器集群内各个所述在用服务器的硬件资源量，建立所述服务器集群的硬件资源池；

硬件资源池划分单元，用于基于预设的初始分配比例，将所述硬件资源池划分为用于响应虚拟主机租赁服务的第一资源区域以及用于响应实时云计算服务的第二资源区域；

实际占用比例获取单元，用于通过所述资源调节器采集所述第一资源区域的第一占用率以及第二资源区域的第二占用率，根据所述第一占用率以及所述第二占用率计算实际占用比例；

硬件资源调整单元，用于若所述实际占用比例与所述初始分配比例之差的绝对值大于预设的调整阈值，则将所述实际占用比例设置为初始分配比例，基于所述实际占用比例调整所述第一资源区域以及所述第二资源区域。

7. 根据权利要求 6 所述的分配设备，其特征在于，所述实际占用比例获取单元包括：

采集触发单元，用于若满足预设的采集条件，则以预设的采集频率获取所述服务器集群内各个所述在用服务器的多个实际资源占用参数；所述实际资源占用参数包括：用于响应虚拟主机租赁服务的第一占用参量以及用于响应实时云计算服务的第二占用参量；

第一占用参量计算单元，用于根据各个所述在用服务器的多个所述第一占用参量，计算所述第一资源区域的第一占用均值以及第一占用标准差；

第二占用参量计算单元，用于根据各个所述在用服务器的多个所述第二占用参量，计算所述第二资源区域的第二占用均值以及第二占用标准差；

实际占用比例计算单元，用于将所述第一占用均值、所述第一占用标注差、所述第二占用均值以及所述第二占用标准差导入实际占用率计算模型，确定所述实际占用比例；所述实际占用率计算模型具体为：

$$ActualRate = \frac{(\overline{VM} + \xi_{VM}) \cdot Countmax}{(\overline{Count} + \xi_{Count}) \cdot VMmax}$$

其中，$ActualRate$ 为所述实际占用比例；$\overline{VM}$ 为第一占用均值；ξ_{VM} 为第一占用标准差；$\overline{Count}$ 为第二占用均值；ξ_{Count} 为第二占用标准差；$VMmax$ 为第一资源区域的硬件资源量；$Countmax$ 为第二资源区域的硬件资源量。

8. 根据权利要求 6 所述的分配设备，其特征在于，所述资源调节器配置单元包括：

安装位置获取单元，用于获取各个所述在用服务器的安装位置，并根据所述安装位置在预设的地图界面上标记各个所述在用服务器；

服务器集群识别单元，用于通过预设的集群窗口在所述地图界面上进行遍历框取，将处于同一所述集群窗口内的所有在用服务器识别为隶属于同一所述服务器集群；

资源调节器创建单元，用于为各个所述服务器集群配置所述资源调节器。

9. 根据权利要求 6-8 任一项所述的分配设备，其特征在于，所述资源调节器配置单元包括：

预期分配比例计算单元，用于将所述初始分配比例以及各个采集时刻获取得到的所述实际占用比例作为训练样本，将所述训练样本导入多层前馈神经网络，计算出预期分配比例；

扩容操作响应单元，用于若接收到系统扩容指令，则确定扩容服务器的硬件资源量，并根据所述预期分配比例对所述扩容服务器的硬件资源量执行硬件资源划分操作。

10. 根据权利要求 6-8 任一项所述的分配设备，其特征在于，所述网络资源的分配设备

还包括：

资源占用率采集单元，用于向各个所述资源调节器广播资源上报指令，以使各个所述资源调节器采集对应的所述服务器集群的资源占用率；

扩容提示信息发送单元，用于若任一所述资源占用率均大于预设的扩容阈值，则向管理员的终端发送扩容提示信息。

11. 一种终端设备，其特征在于，包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机可读指令，所述处理器执行所述计算机可读指令时实现如下步骤：

获取目标服务系统的可用服务器列表，将所述可用服务器列表内的所有在用服务器划分为多个服务器集群，为各个所述服务器集群配置资源调节器；

调用所述资源调节器采集所述服务器集群内各个所述在用服务器的硬件资源量，建立所述服务器集群的硬件资源池；

基于预设的初始分配比例，将所述硬件资源池划分为用于响应虚拟主机租赁服务的第一资源区域以及用于响应实时云计算服务的第二资源区域；

通过所述资源调节器采集所述第一资源区域的第一占用率以及第二资源区域的第二占用率，根据所述第一占用率以及所述第二占用率计算实际占用比例；

若所述实际占用比例与所述初始分配比例之差的绝对值大于预设的调整阈值，则将所述实际占用比例设置为初始分配比例，基于所述实际占用比例调整所述第一资源区域以及所述第二资源区域。

12. 根据权利要求 11 所述的终端设备，其特征在于，所述通过所述资源调节器采集所述第一资源区域的第一占用率以及第二资源区域的第二占用率，根据所述第一占用率以及所述第二占用率计算实际占用比例，包括：

若满足预设的采集条件，则以预设的采集频率获取所述服务器集群内各个所述在用服务器的多个实际资源占用参数；所述实际资源占用参数包括：用于响应虚拟主机租赁服务的第一占用参量以及用于响应实时云计算服务的第二占用参量；

根据各个所述在用服务器的多个所述第一占用参量，计算所述第一资源区域的第一占用均值以及第一占用标准差；

根据各个所述在用服务器的多个所述第二占用参量，计算所述第二资源区域的第二占用均值以及第二占用标准差；

将所述第一占用均值、所述第一占用标注差、所述第二占用均值以及所述第二占用标准差导入实际占用率计算模型，确定所述实际占用比例；所述实际占用率计算模型具体为：

$$ActualRate = \frac{(\overline{VM} + \xi_{VM}) \cdot Countmax}{(\overline{Count} + \xi_{Count}) \cdot VMmax}$$

其中，*ActualRate* 为所述实际占用比例；$\overline{VM}$ 为第一占用均值；ξ_{VM} 为第一占用标准差；$\overline{Count}$ 为第二占用均值；ξ_{Count} 为第二占用标准差；*VMmax* 为第一资源区域的硬件资源量；*Countmax* 为第二资源区域的硬件资源量。

13. 根据权利要求 11 所述的终端设备，其特征在于，所述将所述可用服务器列表内的所有在用服务器划分为多个服务器集群，为各个所述服务器集群配置资源调节器，包括：

获取各个所述在用服务器的安装位置，并根据所述安装位置在预设的地图界面上标记各个所述在用服务器；

通过预设的集群窗口在所述地图界面上进行遍历框取，将处于同一所述集群窗口内的所有在用服务器识别为隶属于同一所述服务器集群；

为各个所述服务器集群配置所述资源调节器。

14. 根据权利要求 11-13 任一项所述的终端设备，其特征在于，在所述根据所述第一占用率以及所述第二占用率计算实际占用比例之后，所述处理器执行所述计算机可读指令时还实现如下步骤：

将所述初始分配比例以及各个采集时刻获取得到的所述实际占用比例作为训练样本，将所述训练样本导入多层前馈神经网络，计算出预期分配比例；

若接收到系统扩容指令，则确定扩容服务器的硬件资源量，并根据所述预期分配比例对所述扩容服务器的硬件资源量执行硬件资源划分操作。

15. 根据权利要求 11-13 任一项所述的终端设备，其特征在于，所述处理器执行所述计算机可读指令时还实现如下步骤：

向各个所述资源调节器广播资源上报指令，以使各个所述资源调节器采集对应的所述服务器集群的资源占用率；

若任一所述资源占用率均大于预设的扩容阈值，则向管理员的终端发送扩容提示信息。

16. 一种计算机非易失性可读存储介质，所述计算机非易失性可读存储介质存储有计算机可读指令，其特征在于，所述计算机可读指令被处理器执行时实现如下步骤：

获取目标服务系统的可用服务器列表，将所述可用服务器列表内的所有在用服务器划分为多个服务器集群，为各个所述服务器集群配置资源调节器；

调用所述资源调节器采集所述服务器集群内各个所述在用服务器的硬件资源量，建立所述服务器集群的硬件资源池；

基于预设的初始分配比例，将所述硬件资源池划分为用于响应虚拟主机租赁服务的第一资源区域以及用于响应实时云计算服务的第二资源区域；

通过所述资源调节器采集所述第一资源区域的第一占用率以及第二资源区域的第二占用率，根据所述第一占用率以及所述第二占用率计算实际占用比例；

若所述实际占用比例与所述初始分配比例之差的绝对值大于预设的调整阈值，则将所述实际占用比例设置为初始分配比例，基于所述实际占用比例调整所述第一资源区域以及所述第二资源区域。

17. 根据权利要求 16 所述的计算机非易失性可读存储介质，其特征在于，所述通过所述资源调节器采集所述第一资源区域的第一占用率以及第二资源区域的第二占用率，根据所述第一占用率以及所述第二占用率计算实际占用比例，包括：

若满足预设的采集条件，则以预设的采集频率获取所述服务器集群内各个所述在用服务器的多个实际资源占用参数；所述实际资源占用参数包括：用于响应虚拟主机租赁服务的第一占用参量以及用于响应实时云计算服务的第二占用参量；

根据各个所述在用服务器的多个所述第一占用参量，计算所述第一资源区域的第一占用均值以及第一占用标准差；

根据各个所述在用服务器的多个所述第二占用参量，计算所述第二资源区域的第二占用均值以及第二占用标准差；

将所述第一占用均值、所述第一占用标注差、所述第二占用均值以及所述第二占用标准差导入实际占用率计算模型，确定所述实际占用比例；所述实际占用率计算模型具体为：

$$ActualRate = \frac{(\overline{VM} + \xi_{VM}) \cdot Countmax}{(\overline{Count} + \xi_{Count}) \cdot VMmax}$$

其中，$ActualRate$ 为所述实际占用比例；$\overline{VM}$ 为第一占用均值；ξ_{VM} 为第一占用标准差；$\overline{Count}$ 为第二占用均值；ξ_{Count} 为第二占用标准差；$VMmax$ 为第一资源区域的硬件资源量；$Countmax$ 为第二资源区域的硬件资源量。

18. 根据权利要求 16 所述的计算机非易失性可读存储介质，其特征在于，所述将所述可用服务器列表内的所有在用服务器划分为多个服务器集群，为各个所述服务器集群配置资源调节器，包括：

获取各个所述在用服务器的安装位置，并根据所述安装位置在预设的地图界面上标记各个所述在用服务器；

通过预设的集群窗口在所述地图界面上进行遍历框取，将处于同一所述集群窗口内的

所有在用服务器识别为隶属于同一所述服务器集群；

为各个所述服务器集群配置所述资源调节器。

19. 根据权利要求 16-18 任一项所述的计算机非易失性可读存储介质，其特征在于，在所述根据所述第一占用率以及所述第二占用率计算实际占用比例之后，所述处理器执行所述计算机可读指令时还实现如下步骤：

将所述初始分配比例以及各个采集时刻获取得到的所述实际占用比例作为训练样本，将所述训练样本导入多层前馈神经网络，计算出预期分配比例；

若接收到系统扩容指令，则确定扩容服务器的硬件资源量，并根据所述预期分配比例对所述扩容服务器的硬件资源量执行硬件资源划分操作。

20. 如权利要求 16-18 任一项所述的计算机非易失性可读存储介质，其特征在于，所述处理器执行所述计算机可读指令时还实现如下步骤：

向各个所述资源调节器广播资源上报指令，以使各个所述资源调节器采集对应的所述服务器集群的资源占用率；

若任一所述资源占用率均大于预设的扩容阈值，则向管理员的终端发送扩容提示信息。

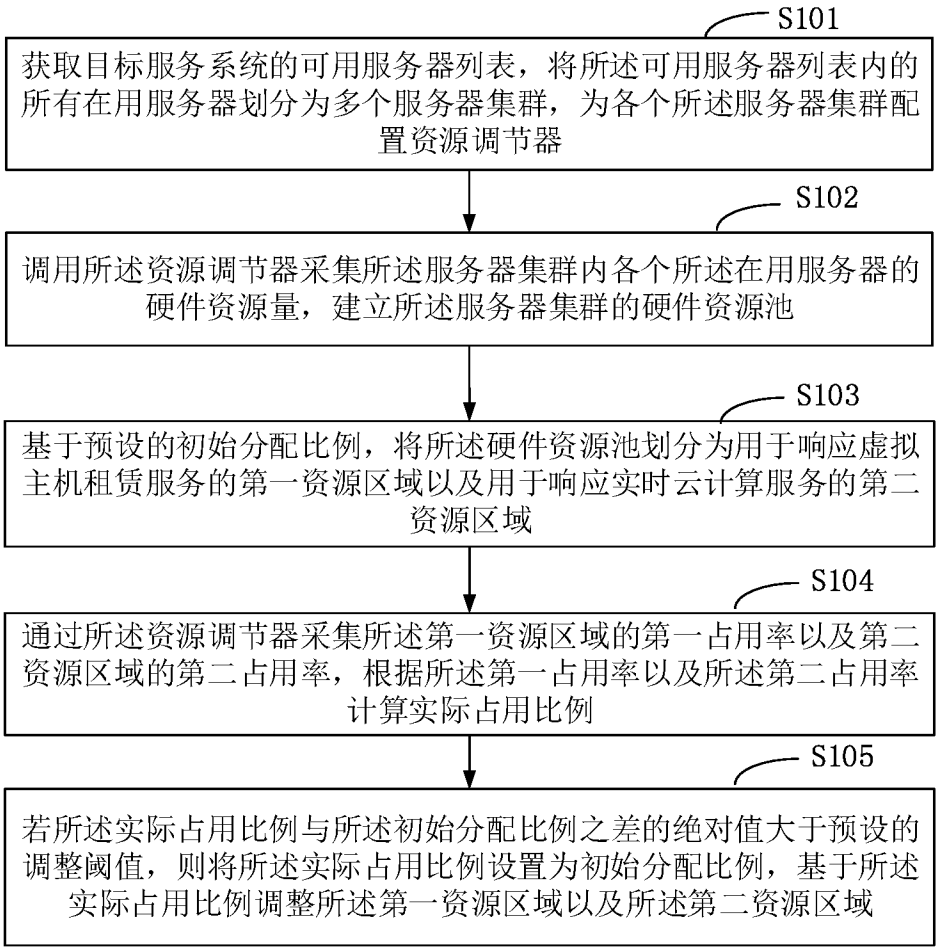

S101
获取目标服务系统的可用服务器列表，将所述可用服务器列表内的所有在用服务器划分为多个服务器集群，为各个所述服务器集群配置资源调节器
S102
调用所述资源调节器采集所述服务器集群内各个所述在用服务器的硬件资源量，建立所述服务器集群的硬件资源池
S103
基于预设的初始分配比例，将所述硬件资源池划分为用于响应虚拟主机租赁服务的第一资源区域以及用于响应实时云计算服务的第二资源区域
S104
通过所述资源调节器采集所述第一资源区域的第一占用率以及第二资源区域的第二占用率，根据所述第一占用率以及所述第二占用率计算实际占用比例
S105
若所述实际占用比例与所述初始分配比例之差的绝对值大于预设的调整阈值，则将所述实际占用比例设置为初始分配比例，基于所述实际占用比例调整所述第一资源区域以及所述第二资源区域

图 1

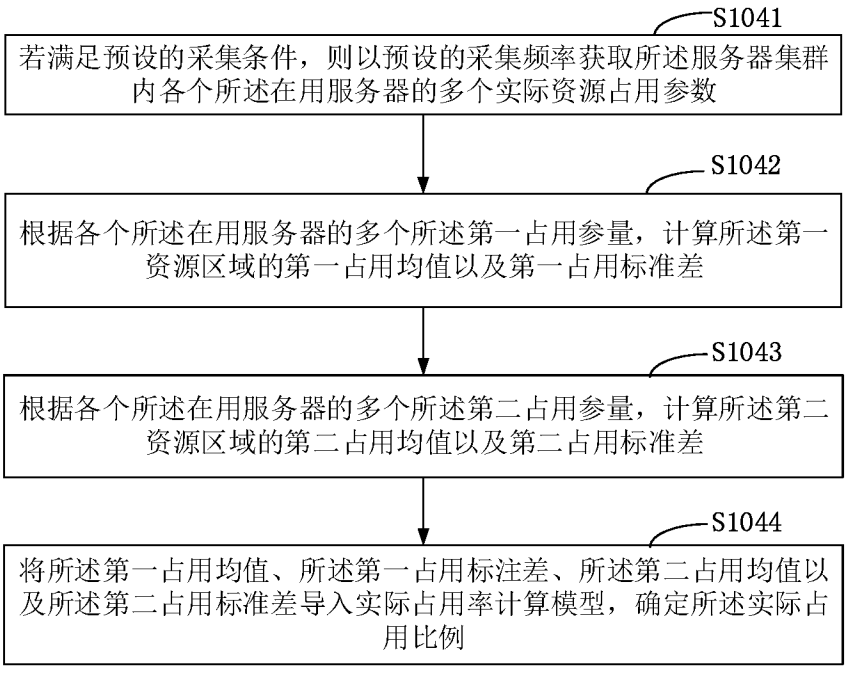

S1041
若满足预设的采集条件，则以预设的采集频率获取所述服务器集群内各个所述在用服务器的多个实际资源占用参数
S1042
根据各个所述在用服务器的多个所述第一占用参量，计算所述第一资源区域的第一占用均值以及第一占用标准差
S1043
根据各个所述在用服务器的多个所述第二占用参量，计算所述第二资源区域的第二占用均值以及第二占用标准差
S1044
将所述第一占用均值、所述第一占用标注差、所述第二占用均值以及所述第二占用标准差导入实际占用率计算模型，确定所述实际占用比例

图 2

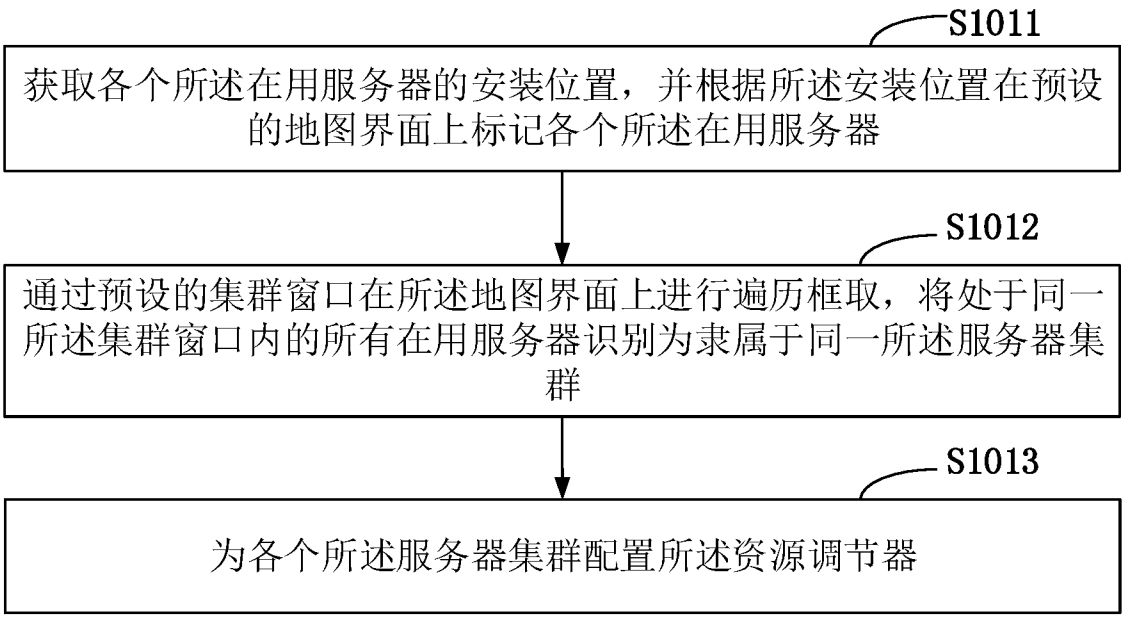

S1011
获取各个所述在用服务器的安装位置，并根据所述安装位置在预设的地图界面上标记各个所述在用服务器
S1012
通过预设的集群窗口在所述地图界面上进行遍历框取，将处于同一所述集群窗口内的所有在用服务器识别为隶属于同一所述服务器集群
S1013
为各个所述服务器集群配置所述资源调节器

图 3

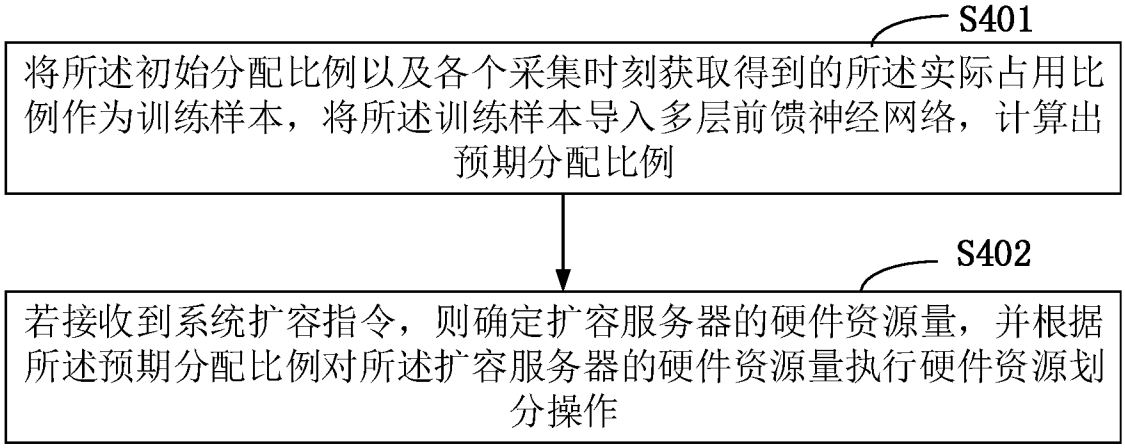

S401
将所述初始分配比例以及各个采集时刻获取得到的所述实际占用比例作为训练样本，将所述训练样本导入多层前馈神经网络，计算出预期分配比例
S402
若接收到系统扩容指令，则确定扩容服务器的硬件资源量，并根据所述预期分配比例对所述扩容服务器的硬件资源量执行硬件资源划分操作

图 4

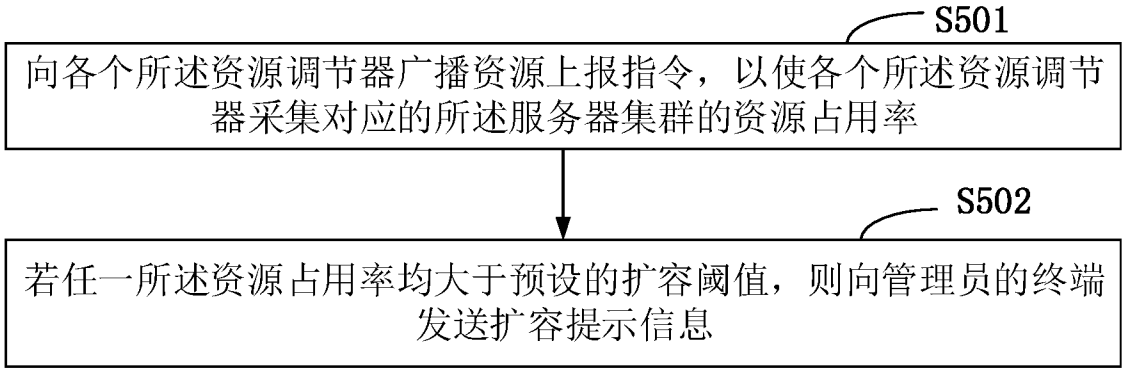

S501
向各个所述资源调节器广播资源上报指令，以使各个所述资源调节器采集对应的所述服务器集群的资源占用率
S502
若任一所述资源占用率均大于预设的扩容阈值，则向管理员的终端发送扩容提示信息

图 5

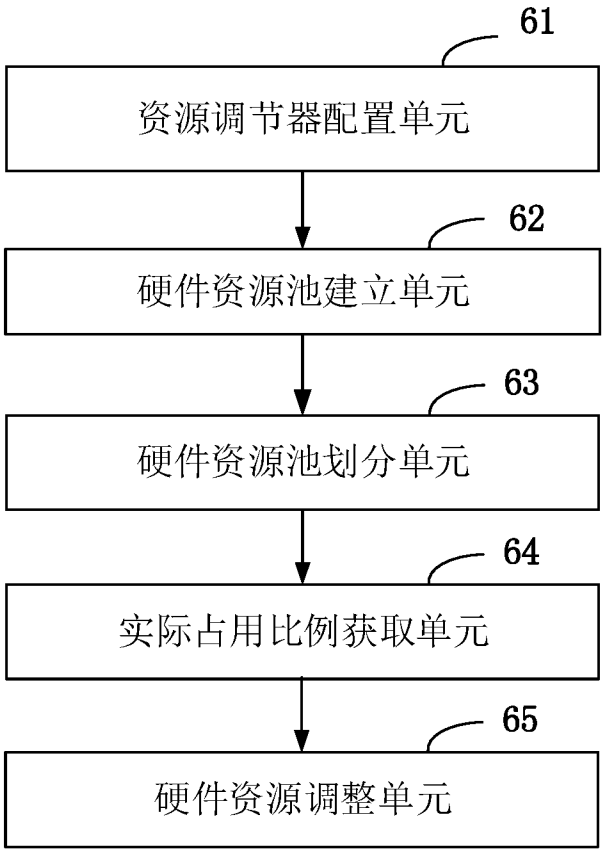

61
资源调节器配置单元
62
硬件资源池建立单元
63
硬件资源池划分单元
64
实际占用比例获取单元
65
硬件资源调整单元

图 6

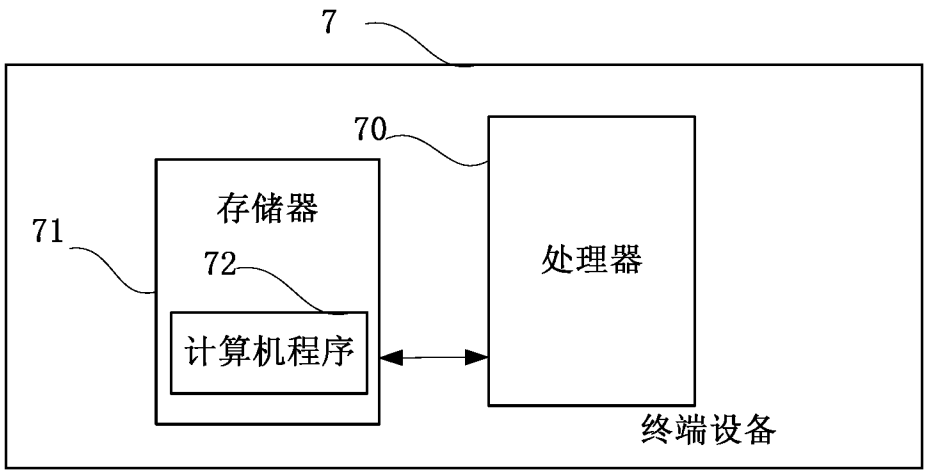

7
70
存储器
71
72
计算机程序
处理器
终端设备

图 7

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2019/091499

A. CLASSIFICATION OF SUBJECT MATTER

G06F 9/455(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, CNKI, WPI, EPODOC: 服务器, 集群, 资源池, 虚拟, 租赁, 阈值, 使用率, 占用率, 分配, 划分, 组, 实时, 云, 计算, 存储, 数据, 节点, 区域, 扩容, 转移, 模型, 初始, 稽核, 标准差, 位置, 地图, 硬件, 内存, 比例, CPU, memory, server, trunked, group, resource, pool, virtual, rental, threshold, utilization, rate, occupy, hardware, distribute, divide, real, time, cloud, calculate, storage, data, node, region, zone, scale, transfer, model, initial, audit, deviation, location, map, proportion, assign

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 109857518 A (PING'AN TECHNOLOGY (SHENZHEN) CO., LTD.) 07 June 2019 (2019-06-07) claims 1-10	1-20
Y	CN 107205030 A (CHENGDU PUSH MEDIA CULTURE PROPAGATION CO., LTD.) 26 September 2017 (2017-09-26) description, paragraphs [0025]-[0067]	1-20
Y	CN 105912397 A (LE HOLDINGS (BEIJING) CO., LTD.) 31 August 2016 (2016-08-31) description, paragraphs [0028]-[0079]	1-20
A	CN 103812789 A (FOSHAN POWER SUPPLY BUREAU GUANGDONG POWER GRID CORP.) 21 May 2014 (2014-05-21) entire document	1-20
A	US 2011138019 A1 (ELECTRONICS AND TELECOMMUNICATIONS RESEARCH INSTITUTE) 09 June 2011 (2011-06-09) entire document	1-20

☐ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
17 September 2019	**09 October 2019**

Name and mailing address of the ISA/CN	Authorized officer
China National Intellectual Property Administration (ISA/CN) **No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088** **China**	
Facsimile No. **(86-10)62019451**	Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2019/091499

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 109857518 A	07 June 2019	None	
CN 107205030 A	26 September 2017	None	
CN 105912397 A	31 August 2016	WO 2017166654 A1	05 October 2017
CN 103812789 A	21 May 2014	None	
US 2011138019 A1	09 June 2011	KR 20110065159 A	15 June 2011

国际检索报告

国际申请号
PCT/CN2019/091499

A. 主题的分类

G06F 9/455(2006.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06F

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称，和使用的检索词（如使用））

CNPAT, CNKI, WPI, EPODOC: 服务器, 集群, 资源池, 虚拟, 租赁, 阈值, 使用率, 占用率, 分配, 划分, 组, 实时, 云, 计算, 存储, 数据, 节点, 区域, 扩容, 转移, 模型, 初始, 稽核, 标准差, 位置, 地图, 硬件, 内存, 比例, CPU, memory, server, trunked, group, resource, pool, virtual, rental, threshold, utilization, rate, occupy, hardware, distribute, divide, real, time, cloud, calculate, storage, data, node, region, zone, scale, transfer, model, initial, audit, deviation, location, map, proportion, assign

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
PX	CN 109857518 A (平安科技深圳有限公司) 2019年 6月 7日 (2019-06-07) 权利要求1-10	1-20
Y	CN 107205030 A (成都博视美达文化传播有限公司) 2017年 9月 26日 (2017-09-26) 说明书第[0025]-[0067]段	1-20
Y	CN 105912397 A (乐视控股北京有限公司) 2016年 8月 31日 (2016-08-31) 说明书第[0028]-[0079]段	1-20
A	CN 103812789 A (广东电网公司佛山供电局) 2014年 5月 21日 (2014-05-21) 全文	1-20
A	US 2011138019 A1 (ELECTRONICS AND TELECOMMUNICATIONS RESEARCH INSTITUTE) 2011年 6月 9日 (2011-06-09) 全文	1-20

☐ 其余文件在C栏的续页中列出。

☑ 见同族专利附件。

* 引用文件的具体类型：

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件，或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件（如具体说明的）

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布，与申请不相抵触，但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件，单独考虑该文件，认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件，当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时，要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期	国际检索报告邮寄日期
2019年 9月 17日	2019年 10月 9日
ISA/CN的名称和邮寄地址 中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451	受权官员 张雪 电话号码 86-(10)-53961613

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2019/091499

检索报告引用的专利文件	公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN 109857518 A	2019年 6月 7日	无	
CN 107205030 A	2017年 9月 26日	无	
CN 105912397 A	2016年 8月 31日	WO 2017166654 A1	2017年 10月 5日
CN 103812789 A	2014年 5月 21日	无	
US 2011138019 A1	2011年 6月 9日	KR 20110065159 A	2011年 6月 15日