

(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 3 月 19 日 (19.03.2020)



(10) 国际公布号
WO 2020/052151 A1

(51) 国际专利分类号:
H04L 29/08 (2006.01)

(21) 国际申请号: PCT/CN2018/122637

(22) 国际申请日: 2018 年 12 月 21 日 (21.12.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201811071652.9 2018 年 9 月 13 日 (13.09.2018) CN

(71) 申请人: 深圳壹账通智能科技有限公司(ONE CONNECT SMART TECHNOLOGY CO., LTD. (SHENZHEN)) [CN/CN]; 中国广东省深圳市前海深港合作区前湾一路 1 号 A 栋 201 室, Guangdong 518000 (CN)。

(72) 发明人: 徐欣(XU, Xin); 中国广东省深圳市前海深港合作区前湾一路 1 号 A 栋 201 室, Guangdong 518000 (CN)。

(74) 代理人: 深圳市世纪恒程知识产权代理事务所(CENFO INTELLECTUAL PROPERTY AGENCY); 中国广东省深圳市南山区粤海街道高新技术产业园北区松坪山路 3 号奥特讯电力大厦 201, Guangdong 518057 (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, 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,

(54) Title: SERVICE NODE MANAGEMENT METHOD, APPARATUS AND DEVICE, AND COMPUTER READABLE STORAGE MEDIUM

(54) 发明名称: 服务节点的管理方法、装置、设备及计算机可读存储介质

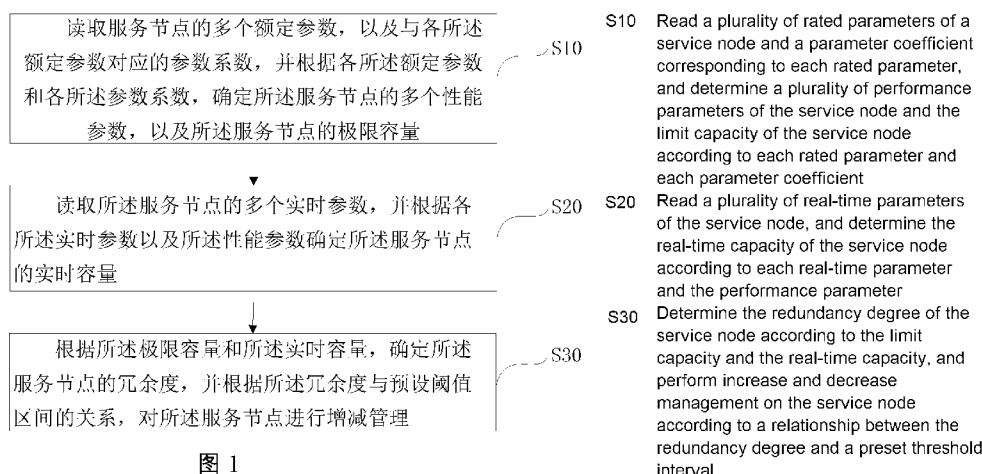


图 1

(57) Abstract: A service node management method, apparatus and device, and a computer readable storage medium. The method comprises: reading a plurality of rated parameters of a service node and a parameter coefficient corresponding to each rated parameter, and determining a plurality of performance parameters of the service node and the limit capacity of the service node according to each rated parameter and each parameter coefficient (S10); reading a plurality of real-time parameters of the service node, and determining the real-time capacity of the service node according to each real-time parameter and the performance parameter (S20); and determining the redundancy degree of the service node according to the limit capacity and the real-time capacity, and performing increase and decrease management on the service node according to a relationship between the redundancy degree and a preset threshold interval (S30). According to the method, the increase and decrease management of the service node can be automatically and accurately performed on the basis of the relationship between the redundancy degree that reflects the access traffic currently provided by the service node and the preset threshold interval, so that the process of performing increase and decrease management by manual repeated operations can be avoided, the operation and maintenance costs can be reduced, and the high availability of each service node in a hybrid cloud

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))。

architecture can be improved.

(57) 摘要: 一种服务节点的管理方法、装置、设备及计算机可读存储介质, 所述方法包括: 读取服务节点的多个额定参数, 以及与各额定参数对应的参数系数, 并根据各额定参数和各参数系数, 确定服务节点的多个性能参数, 以及服务节点的极限容量(S10); 读取服务节点的多个实时参数, 并根据各实时参数以及性能参数确定服务节点的实时容量(S20); 根据极限容量和实时容量, 确定服务节点的冗余度, 并根据冗余度与预设阈值区间的关系, 对服务节点进行增减管理(S30)。所述方法通过反映服务节点当前可提供访问量的冗余度和预设阈值区间的关系, 对服务节点进行自动准确的增减管理, 避免了人工重复操作进行增减管理, 降低了运维成本, 提高了混合云架构中各服务节点的高可用性。

服务节点的管理方法、装置、设备及计算机可读存储介质

本申请要求于2018年09月13日提交中国专利局、申请号为201811071652.9、发明名称为“服务节点的管理方法、装置、设备及计算机可读存储介质”的中国专利申请的优先权，其全部内容通过引用结合在申请中。

技术领域

本申请主要涉及网络通信技术领域，具体地说，涉及一种服务节点的管理方法、装置、设备及计算机可读存储介质。

背景技术

随着云技术的发展，很多互联网公司在使用传统IDC(Internet Data Center , 互联网数据中心) 网络环境的同时增加了云平台网络环境；而传统IDC网络环境的服务器节点运行着公司至关重要的服务，如果迁入到云平台网络环境存在巨大的迁移风险；从而需要将传统IDC网络环境和云平台网络环境以混合云架构的方式使用。

但是目前对于传统IDC网络环境和云平台网络环境共同向外提供服务的混合云架构，无法准确获取各服务节点的负载和带宽占用状态信息；而使得无法根据实时流量对服务节点实现自动的弹性伸缩，即无法自动增加或删减服务节点；导致在访问请求剧增的情况下，无法准确自动部署服务节点来应对访问请求，而需要人工多次重复操作进行管理，增加了运维成本，且不利于混合云架构的高可用性。

发明内容

本申请的主要目的是提供一种服务节点的管理方法、装置、设备及计算机可读存储介质，旨在解决现有技术混合云架构无法针对访问请求量的变化准确自动

部署服务节点，运维成本高，可用性低的问题。

为实现上述目的，本申请提供一种服务节点的管理方法，所述服务节点的管理方法包括以下步骤：

读取服务节点的多个额定参数，以及与各所述额定参数对应的参数系数，并根据各所述额定参数和各所述参数系数，确定所述服务节点的多个性能参数，以及所述服务节点的极限容量；

读取所述服务节点的多个实时参数，并根据各所述实时参数以及所述性能参数确定所述服务节点的实时容量；

根据所述极限容量和所述实时容量，确定所述服务节点的冗余度，并根据所述冗余度与预设阈值区间的关系，对所述服务节点进行增减管理。

此外，为实现上述目的，本申请还提出一种服务节点的管理装置，所述服务节点的管理装置包括：

读取模块，用于读取服务节点的多个额定参数，以及与各所述额定参数对应的参数系数，并根据各所述额定参数和各所述参数系数，确定所述服务节点的多个性能参数，以及所述服务节点的极限容量；

确定模块，用于读取所述服务节点的多个实时参数，并根据各所述实时参数以及所述性能参数确定所述服务节点的实时容量；

管理模块，用于根据所述极限容量和所述实时容量，确定所述服务节点的冗余度，并根据所述冗余度与预设阈值区间的关系，对所述服务节点进行增减管理。

此外，为实现上述目的，本申请还提出一种服务节点的管理设备，所述服务节点的管理设备包括：存储器、处理器、通信总线以及存储在所述存储器上的

服务节点的管理程序；

所述通信总线用于实现处理器和存储器之间的连接通信；

所述处理器用于执行所述服务节点的管理程序，以实现以下步骤：

读取服务节点的多个额定参数，以及与各所述额定参数对应的参数系数，并根据各所述额定参数和各所述参数系数，确定所述服务节点的多个性能参数，以及所述服务节点的极限容量；

读取所述服务节点的多个实时参数，并根据各所述实时参数以及所述性能参数确定所述服务节点的实时容量；

根据所述极限容量和所述实时容量，确定所述服务节点的冗余度，并根据所述冗余度与预设阈值区间的关系，对所述服务节点进行增减管理。

此外，为实现上述目的，本申请还提供一种计算机可读存储介质，所述计算机可读存储介质存储有一个或者一个以上程序，所述一个或者一个以上程序可被一个或者一个以上的处理器执行以用于：

读取服务节点的多个额定参数，以及与各所述额定参数对应的参数系数，并根据各所述额定参数和各所述参数系数，确定所述服务节点的多个性能参数，以及所述服务节点的极限容量；

读取所述服务节点的多个实时参数，并根据各所述实时参数以及所述性能参数确定所述服务节点的实时容量；

根据所述极限容量和所述实时容量，确定所述服务节点的冗余度，并根据所述冗余度与预设阈值区间的关系，对所述服务节点进行增减管理。

本实施例的服务节点的管理方法，根据所读取的服务节点的多个额定参数以及与各额定参数对应的各参数系数，确定服务节点的多个性能参数以及极限容

量 ;同时还根据所读取的服务器的多个实时参数以及所确定的多个性能参数进一步确定服务器的实时容量 ,进而根据极限容量和实时容量确定服务节点的冗余度 ;因实时容量表征服务节点中各资源的实时使用程度 ,从而通过冗余度可反映服务节点实时所能提供的访问服务量 ;而预设阈值区间表征服务节点所提供访问服务量的合理范围 ,用冗余度和预设阈值区间的关系 ,判断服务节点实时所提供服务量是否合理 ,进而根据合理性对服务节点进行增减管理 ;避免了人工重复操作进行增减管理 ,实现对服务节点的自动增减管理部署 ,降低了运维成本 ,提高了混合云架构中各服务节点的高可用性。

附图说明

图1是本申请的服务节点的管理方法第一实施例的流程示意图 ;

图2是本申请的服务节点的管理装置第一实施例的功能模块示意图 ;

图3是本申请实施例方法涉及的硬件运行环境的设备结构示意图。

本申请目的的实现、功能特点及优点将结合实施例 ,参照附图做进一步说明。

具体实施方式

应当理解 ,此处所描述的具体实施例仅仅用以解释本申请 ,并不用于限定本申请。

本申请提供一种服务节点的管理方法。

请参照图1 ,图1为本申请服务节点的管理方法第一实施例的流程示意图。在本实施例中 ,所述服务节点的管理方法包括 :

步骤S10 ,读取服务节点的多个额定参数 ,以及与各所述额定参数对应的参数系数 ,并根据各所述额定参数和各所述参数系数 ,确定所述服务节点的多个性能参数 ,以及所述服务节点的极限容量 ;

本申请的服务节点的管理方法应用于网络环境平台的控制中心 ,适用于通过控制中心实现对网络环境平台中各服务节点的管理。本实施例的网络环境为由传统IDC网络环境和云平台网络环境共同向外提供服务的混合云架构 ,如金融机

构中反欺诈平台的混合云架构；服务节点则包括传统IDC网络环境和云平台网络环境中所有向外提供服务的节点，外部的访问通过向服务节点发起访问请求，通过服务节点接入到网络环境，获得网络环境的资源。网络环境中包括大量服务节点，在提供不同类型服务的同时，分担各服务节点的负载量；不同服务节点之间所具有的硬件资源存在差异性，使得各服务节点提供服务的能力不同，如8核8G的服务节点与4核8G服务节点之间的差异性。即不同硬件资源表征服务节点具有不同的性能，将服务节点所具有硬件资源的参数作为服务节点额定参数，包括核心数、频率、内存类型值、内存大小值、磁盘类型值、IO带宽值、该服务节点所分配的网络带宽值。此外，考虑到不同的额定参数对服务节点的性能影响程度存在差异，从而针对各个额定参数预先设定不同的参数系数，通过参数系数的不同来体现各个额定参数对服务节点性能的影响；其中参数系数经试验确定，设定多个不同参数系数进行试验，拟合确定最佳的参数系数。

在对服务节点管理过程中，读取服务节点的多个额定参数，以及与各个额定参数对应的参数系数，以确定服务节点所具有的性能。而因不同额定参数表征服务节点在不同方面的性能，从而根据不同的额定参数及对应参数系数，可确定服务节点的多个性能参数。如由核心数、频率及其对应参数系数，所确定的服务节点中由CPU（Central Processing Unit，中央处理器）的核数和频率决定的性能参数；而由内存类型值、内存大小值及其对应参数系数，所确定的服务节点中由内存类型和内存大小决定的性能参数；由磁盘类型值、IO带宽值及其对应参数系数，所确定的服务节点中由磁盘类型及IO带宽决定的性能参数；由服务节点分配的网络带宽值及其对应参数系数，所确定的网络带宽的性能值参数。在确定各个性能参数后，通过各个性能参数可反映服务节点整体所能提供服务的能力；此整

体所能提供服务的能力用极限容量确定,即根据各性能参数确定服务节点的极限容量,以表征服务节点在所有硬件资源均充分使用的情况下,其所能提供的最大访问量。

步骤S20,读取所述服务节点的多个实时参数,并根据各所述实时参数以及所述性能参数确定所述服务节点的实时容量;

可理解地,因不同用户对服务节点的访问时间不一样,使得服务节点在不同时间所接收到的访问量不相同,本实施例用实时容量表征服务节点实时所提供的访问能力。读取服务节点的多个实时参数,实时参数可包括CPU实时使用率、内存的实时占有率、磁盘的实时IO占用率和网络带宽实际占用率等;通过服务节点实际使用过程中的实时参数对各性能参数的影响,确定服务节点的实时容量。

步骤S30,根据所述极限容量和所述实时容量,确定所述服务节点的冗余度,并根据所述冗余度与预设阈值区间的关系,对所述服务节点进行增减管理。

进一步地,在确定服务节点最大所能提供的访问量以及实际提供的访问量之后,可根据两者确定服务节点的冗余度,即服务节点剩余可接入的访问量;其中冗余度最大极限值接近于1,而最小极限值接近于0;最大极限值表征服务节点可接入的最大访问量,最小极限值表征服务节点可接入的最小访问量。因服务节点用于对外提供服务,需要考虑到服务节点的带宽、响应速度等;在带宽一定的情况下,响应速度与访问量相关;若访问量过大,可能会拉低响应速度。从而需要控制访问量在一定范围内,具体地,预先设定预设阈值区间,根据冗余度与此预设阈值区间的大小关系,确定服务节点的访问量大小;进而对服务节点进行增减管理,以确保服务节点的访问量在一定范围内。

预设阈值区间为通过多次试验所确定,包括上边界阈值和下边界阈值的数

值区间；上边界阈值表征冗余度较大，服务节点的访问量较小，资源消耗较小；而下边界阈值表征冗余度较小，服务节点的访问量较大，资源消耗较大。对于访问量较大的下边界阈值，考虑到服务节点的极限容量为其理论上所能接入的最大访问量，在实际使用过程中，因外界环境因素的影响可能并不能达到此理论值，从而所设定的下边界阈值为比极限容量数值小的数值。而为了准确反映服务节点的资源即将被消耗殆尽，通过压力测试的实验确定此下边界阈值。具体地，服务节点对外暴露有接口，通过接口调用的方式实现对服务节点的访问测试；对接口的调用频度可反映服务节点的访问请求次数；使用不同的调用频度对接口进行调用，以对服务节点进行压力测试，确定压力测试达到拐点的耗时，其中拐点为服务节点资源急剧消耗点；因耗时也可体现服务节点资源的充足情况，耗时越短说明服务节点资源较丰富；从而试验调用频度和耗时平衡的拐点，在此平衡的拐点，调用频度较高且耗时也较高，表示服务节点访问量大，资源即将消耗殆尽，而将此拐点的数值作为下边界阈值。其中根据冗余度与预设阈值区间的关系，对服务节点进行增减管理的步骤包括：

步骤S31，将所述冗余度和预设阈值区间对比，判断所述冗余度是否在预设阈值区间内；

更进一步地，在确定服务节点的冗余度后，将冗余度和预设阈值区间对比，判断冗余度是否在预设阈值区间内，即服务节点的访问量是否在合适的范围内。将冗余度分别和预设阈值区间的上边界阈值、下边界阈值对比，判断冗余度是否大于下边界阈值，并小于上边界阈值。

步骤S32，当所述冗余度小于预设阈值区间下边界数值时，增加与所述服务节点对应的新服务节点；

当判断出服务节点的冗余度小于预设阈值区间的下边界数值时，说明服务节点剩余可接入的访问请求数量很少，当前的访问量很大，可能影响其响应速度。从而自动增加与服务节点对应的新服务节点，两者的对应关系在于新增的服务节点与原服务节点所提供的服务类型一致。为新增的服务节点分配IP地址、CPU核数、内存大小值、带宽值等资源，以使用新增服务节点提供服务，分担原服务节点的访问压力。

步骤S33，当所述冗余度大于预设阈值区间上边界数值时，对所述服务节点进行删除操作。

当判断出服务节点的冗余度大于预设阈值区间的上边界数值时，说明服务节点剩余可接入的访问请求数量很大，当前的访问量很小，资源没有得到充分利用。从而对此服务节点进行自动删除操作，以释放并回收系统资源；同时将对此服务节点的访问请求接入到其他与此服务节点所提供类似服务的服务节点上，以在对外提供服务的同时，充分利用系统资源。

本实施例的服务节点的管理方法，根据所读取的服务节点的多个额定参数以及与各额定参数对应的各参数系数，确定服务节点的多个性能参数以及极限容量；同时还根据所读取的服务器的多个实时参数以及所确定的多个性能参数进一步确定服务器的实时容量，进而根据极限容量和实时容量确定服务节点的冗余度；因实时容量表征服务节点中各资源的实时使用程度，从而通过冗余度可反映服务节点实时所能提供的访问服务量；而预设阈值区间表征服务节点所提供访问服务量的合理范围，用冗余度和预设阈值区间的关系，判断服务节点实时所提供访问量是否合理，进而根据合理性对服务节点进行增减管理；避免了人工重复操作进行增减管理，实现对服务节点的自动增减管理部署，降低了运维成本，提高了混

合云架构中各服务节点的高可用性。

进一步地，在本申请服务器节点的管理方法另一实施例中，所述根据各所述额定参数和各所述参数系数，确定所述服务节点的多个性能参数，以及所述服务节点的极限容量的步骤包括：

步骤S11，将各所述额定参数和各所述参数系数基于参数属性进行分组，并根据各组中的所述额定参数和所述参数系数，生成所述服务节点的多个性能参数；

更进一步地，不同额定参数具有不同的参数属性，表征服务节点在不同方面的性能；从而可按照参数属性对额定参数进行分组，将具有相同参数属性的额定参数划分为同一组；通过同一组内额定参数表征服务节点在同一方面的性能。因性能还与参数系数相关，从而在将额定参数进行分组的同时，还将与额定参数对应的参数系数划分到与额定参数对应的组内，以通过各组内的额定参数和参数系数生成反映服务节点性能的多个性能参数。具体地，因服务节点在CPU方面的性能参数由核心数和频率决定，将核心数和频率设定为具有同样的参数属性，而将两者及其对应的参数系数划分到同一组。因服务节点在内存方面的性能参数由内存类型值和内存大小决定，将内存类型值和内存大小设定为具有同样的参数属性，而将两者及其对应的参数系数划分到同一组。服务节点在磁盘类型方面的性能参数由磁盘类型值和IO带宽值决定，将磁盘类型值和IO带宽值设定为具有同样的参数属性，而将两者及其对应的参数系数划分到同一组。服务节点在网络带宽方面的性能参数由服务节点分配的网络带宽值决定，将服务节点分配的网络带宽值及其对应的参数系数划分为一组。在将各额定参数及其各参数系数进行分组后，根据各组中所具有的额定参数及其参数系数，可生成各组中反映服务节点性能的性能参数。

具体地，为了根据各组中额定参数及其对应的各参数系数，确定服务节点的多个性能参数，预先通过多次试验设定有预设公式；将各组中的额定参数及其对应的各参数系数输入到预设公式中，代替其中的自变量进行计算，计算所得到的结果即为性能参数。因各组中所具有的额定参数不相同，表征服务器节点在不同方面的性能，计算时将各组中的额定参数及其对应参数系数输入到预设公式中，得到不同的性能参数。具体地预设公式为：

$$y = m_1 * x_1 + m_2 * x_2$$

其中y为性能参数，x1、x2为各额定参数，m1、m2为与额定参数对应的参数系数。

性能参数因x1、x2以及m1、m2取值的不同而不同。具体地，当在确定服务节点中由CPU核数和频率决定的性能参数y，即服务节点在CPU方面的性能参数时，x1和x2分别为核心数和频率（GHZ），m1、m2则分别为与核心数和频率对应的参数系数。当在确定服务节点中由内存类型和内存大小决定的性能参数y，即服务节点在内存方面的性能参数时，x1和x2分别为内存类型值和内存大小值（GB），m1、m2分别为与内存类型值和内存大小值对应的参数系数；其中内存类型值为依据不同的内存类型而设定的参数值，如ddr2（Double-Data-RateTwoSynchronousDynamicRandomAccessMemory，第二代双倍数据率同步动态随机存取存储器）型的内存类型值比ddr3型的内存类型值低。当在确定服务节点中由磁盘类型及IO带宽决定的性能参数y，即服务节点在磁盘类型方面的性能参数时，x1和x2分别为磁盘类型值和IO带宽值（MB），m1、m2则分别为与磁盘类型值和IO带宽值对应的参数系数；其中磁盘类型值为依据不同的磁盘类型而设定的参数值，如ssd（Solid State Drives，固态硬盘）盘的磁盘类

型值比机械盘的磁盘类型值高。当在确定服务节点中由网络带宽决定的性能值参数 y 时，因网络带宽仅与服务节点所分配的网络带宽值确定，从而将 x_1 作为由服务节点分配的网络带宽值， k_1 则为与由服务节点分配的网络带宽值对应的参数系数；而 x_2 的取值为0，使得与仅与 x_1 、 k_1 相关。将各组中的额定参数及其对应的参数系数，按照上述组合分别输入到预设公式中，得到表征服务节点各个方面性能的多个性能参数。

步骤S12，读取与各所述性能参数对应的预设权重值，并将各所述预设权重值和各所述性能参数输入到第一预设公式中，计算所述服务节点的极限容量；

其中第一预设公式： $ValX = k_1 \lg A_1 + k_2 \lg A_2 + k_3 \lg A_3 + k_4 \lg A_4$ ， $ValX$ 为极限容量， A_1 、 A_2 、 A_3 和 A_4 为各所述性能参数， k_1 、 k_2 、 k_3 和 k_4 为各所述预设权重值， $\lg$ 为取对数值。

进一步地，在确定表征服务节点各个方面性能的多个性能参数后，虽然通过各个性能参数可反映服务节点整体所能提供服务的能力，但各个方面的性能对服务节点整体性能的影响程度存在差异；某些性能参数对服务节点整体提供服务的能力影响较大，而另一些性能参数对服务节点整体提供服务的能力影响较小。为了表征各性能参数对服务节点整体性能的影响，预先针对反映各个方面性能的性能参数设定不同的预设权重值，预设权重值通过多次试验或历史数据统计确定，以表征各个性能参数对整体性能的影响。在确定各个性能参数后，结合与各性能参数对应的预设权重值，确定反映服务节点整体性能的极限容量。具体地，预先通过多次试验设定有第一预设公式，将各性能参数及其对应的预设权重值输入到第一预设公式中，代替其中的自变量进行计算，计算所得到的结果即为服务节点的极限容量，其中，第一预设公式为： $ValX = k_1 \lg A_1 + k_2 \lg A_2 + k_3 \lg A_3 + k_4 \lg A_4$

其中，ValX为极限容量，A1为由CPU核数和频率决定的性能参数、A2由内存类型和内存大小决定的性能参数、A3为由磁盘类型及IO带宽决定的性能参数、A4为由网络带宽决定的性能值参数，k1、k2、k3和k4分别为与A1、A2、A3和A4对应的预设权重值，lg为取对数值。

将由预设公式计算得到的由CPU核数和频率决定的性能参数及其对应的预设权重值，分别替换此第一预设公式中的A1和k1，将由预设公式计算得到的由内存类型和内存大小决定的性能参数及其对应的预设权重值，分别替换此第一预设公式中的A2和k2，将由预设公式计算得到的由磁盘类型及IO带宽决定的性能参数及其对应的预设权重值，分别替换此第一预设公式中的A3和k3，将由预设公式计算得到的由网络带宽决定的性能值参数及其对应的预设权重值，分别替换此第一预设公式中的A4和k4，对此经替换操作的第一预设公式进行计算，所得到的计算结果即为极限容量，表征服务节点所能承受的最大访问量。

进一步地，在本申请服务器节点的管理方法另一实施例中，所述根据各所述实时参数以及所述性能参数确定所述服务节点的实时容量的步骤包括：

将各所述实时参数、各所述性能参数以及各所述预设权重值输入到第二预设公式中，计算所述服务节点的实时容量；

其中所述第二预设公式：

$$\text{ValY} = k_1 \lg A_1 * Q_1 + k_2 \lg A_2 * Q_2 + k_3 \lg A_3 * Q_3 + k_4 \lg A_4 * Q_4$$
，ValY为实时容量，A1、A2、A3和A4为各所述性能参数，k1、k2、k3和k4为各所述预设权重值，Q1、Q2、Q3和Q4为各实时参数，lg为取对数值。

本实施例为了对服务节点实时所接入访问请求的实时容量进行计数，预先设置有第二预设公式；将读取的各实时参数，各性能参数以及与各性能参数对应

的各预设权重值输入到第二预设公式中，代替其中的自变量进行计算，计算所得到的结果即为服务节点的实时容量，具体地，第二预设公式为：

$$\text{ValY} = k_1 \lg A_1 * Q_1 + k_2 \lg A_2 * Q_2 + k_3 \lg A_3 * Q_3 + k_4 \lg A_4 * Q_4$$

其中，ValY为实时容量，Q1为CPU实时使用率、Q2为内存的实时占有率、Q3为磁盘的实时IO占用率、Q4为网络带宽实际占用率，A1为由CPU核数和频率决定的性能参数、A2由内存类型和内存大小决定的性能参数、A3为由磁盘类型及IO带宽决定的性能参数、A4为由网络带宽决定的性能值参数，k1、k2、k3和k4分别为与A1、A2、A3和A4对应的预设权重值，lg为取对数值。

将读取的CPU实时使用率，以及由预设公式计算得到的由CPU核数和频率决定的性能参数及其对应的预设权重值，分别替换此第二预设公式中的Q1、A1和k1；将读取的内存的实时占有率，以及由预设公式计算得到的由内存类型和内存大小决定的性能参数及其对应的预设权重值，分别替换此第二预设公式中的Q2、A2和k2；将读取的磁盘的实时IO占用率，以及由预设公式计算得到的由磁盘类型及IO带宽决定的性能参数及其对应的预设权重值，分别替换此第二预设公式中的Q3、A3和k3；将读取的网络带宽实际占用率，以及由预设公式计算得到的由网络带宽决定的性能值参数及其对应的预设权重值，分别替换此第二预设公式中的Q4、A4和k4；对此经替换操作的第二预设公式进行计算，所得到的计算结果即为实时容量，表征服务节点当前实时所接入的访问量。

进一步地，在本申请服务器节点的管理方法另一实施例中，所述根据所述极限容量和所述实时容量，确定所述服务节点的冗余度的步骤包括：

将所述极限容量和所述实时容量输入到第三预设公式中，计算所述服务节点的冗余度；

其中所述第三预设公式： $R = 1 - \text{ValY}/\text{ValX}$ ， R 为冗余度， ValY 为实时容量， ValX 为极限容量。

更进一步地，在确定服务节点的极限容量和实时容量后，为了根据极限容量和实时容量确定服务节点的冗余度，预先通过实验设置有第三预设公式；将确定的极限容量和实时容量传输到第三预设公式中，代替其中的自变量进行计算，所得到的结果即为服务节点的冗余度，具体地，第三预设公式为： $R = 1 - \text{ValY}/\text{ValX}$

其中， R 为冗余度、 ValY 为实时容量、 ValX 为极限容量。

将经第一预设公式计算得到极限容量，以及经第二预设公式计算得到的实时容量分别替换此第三预设公式中的 ValX 、 ValY ；对此经替换操作的第三预设公式进行计算，所得到的计算结果即为服务节点的冗余度；表征服务节点当前可接入的访问量，以通过此冗余度对服务节点管理，进行增加或删除操作。

进一步地，在本申请服务器节点的管理方法另一实施例中，所述增加与所述服务节点对应的新服务节点的步骤之后包括：

步骤S34，读取与所述新服务节点对应的新增额定参数，并将所述新增额定参数与其他服务节点的额定参数对比，确定所述其他服务节点的额定参数中与所述新增额定参数匹配的目标额定参数；

可理解地，因各服务节点所具有的资源不同，所提供服务能力不同，如8核8G的服务节点的服务能力优于4核8G服务节点的服务能力。为了更好的对访问请求进行响应，应将访问请求优先接入到服务能力好的服务节点。而为了表征各服务节点的服务能力的优劣，本实施例设置有根据各服务节点所具有的资源，对各服务节点分配访问权重的机制，且其中资源配置越高所分配的访问权重越大。从而在新增服务节点后，相应的需要对此新服务节点分配访问权重。具体地，在

新增服务节点后,读取此新服务节点所对应具有的额定参数,并将其具有的额定参数作为新增额定参数,以和其他服务节点的额定参数进行区分;再将此新增的额定参数和其他服务节点所具有的额定参数对比,确定其他服务节点的额定参数中与此新增额定参数匹配的目标额定参数。其中在匹配的过程中优先确定与新增额定参数完全一致的额定参数,将此完全一致的额定参数作为目标额定参数;若其他服务节点的额定参数中不存在与此新增额定参数完全一致的额定参数,则确定与此新增额定参数差异最少的额定参数,而将此差异最少的额定参数作为目标额定参数。

步骤S35,确定所述目标额定参数归属的服务节点,并根据所述归属的服务节点所具有的访问权重,为所述新服务节点分配访问权重。

进一步地,因服务节点的访问权重与服务节点的资源配置相关,而资源配置由额定参数表征;在确定目标额定参数后,因目标额定参数与新增额定参数完全一致,或差异很小,使得目标额定参数所表征的资源配置与新增额定参数所表征的资源配置相同或类似;从而可由目标额定参数对应服务节点所具有的访问权重,来确定新服务节点的访问权重。因目标额定参数与新增额定参数与其他服务节点的额定参数对比确定,从而可根据对比过程确定目标额定参数所来源的服务节点,即目标额定参数所归属的服务节点;进而读取此所归属的服务节点所具有的访问权重,以此访问权重为依据,为新服务节点分配访问权重。当新增额定参数与目标额定参数完全一致时,可将新服务节点的访问权重设置为与归属的服务节点所具有的访问权重相同;而当新增额定参数所表征的服务能力高于目标额定参数,则将新服务节点的访问权重设置为高于归属的服务节点所具有的访问权重;当新增额定参数所表征的服务能力低于目标额定参数,则将新服务节点的访问权

重设置为低于归属的服务节点所具有的访问权重。根据具有类似额定参数的服务节点的访问权重为新服务节点分配访问权重,可使新服务节点的访问权重更为准确;便于后续根据访问权重将访问请求接入到各服务节点,以提供更为优质的服务。

进一步地,在本申请服务器节点的管理方法另一实施例,所述为所述新服务节点分配访问权重的步骤之后包括:

步骤S36,当接收到服务请求时,根据所述服务请求的类型确定与所述服务请求对应的目标服务节点,并判断所述目标服务节点是否存在多个;

更进一步地,当用户有访问网络环境,获取网络资源的需求时,可通过接入到网络环境的终端发送服务请求;且对于不同的网络资源,所发送的服务请求中所携带的标识符不一样。当接收到此服务请求时,读取其中所携带的标识符,并将此标识符与预先设置的标识类型表中各标识符对比,确定标识类型表中与携带的标识符对应的标识符,进而根据标识类型表中与此对应标识符对应的类型,即可确定此服务请求的类型。因不同服务节点所对应的服务类型不相同,在确定服务请求的类型后,可进一步确定各服务节点中提供此类型服务的服务节点,并将此服务节点作为与服务请求对应的目标服务节点。因对于某些请求量很大的服务请求,为了提供更好的服务,通常设置有多个服务节点;从而在确定目标服务节点后,需要进一步判断此目标服务节点是否存在多个,以从多个目标服务节点中确定为服务请求提供服务的节点。

步骤S37,若存在多个所述目标服务节点,则判断各所述目标服务节点所具有目标访问权重的大小关系,并调用具有最大所述目标访问权重的目标访问节点对所述服务请求进行响应。

本实施中对于同一类型的服务节点设置有相同的标识，在确定目标服务节点后，读取其中所携带的标识，并判断是否还存在其他服务节点也携带有此标识。若不存在其他服务节点也携带有此标识，则说明不存在多个目标服务节点，将此目标服务节点确定为服务请求提供服务的节点，由其对服务请求进行响应。若存在其他服务节点也携带有此标识，则说明存在多个目标服务节点，则读取各个目标服务节点所具有的目标访问权重，并将各个目标访问权重进行对比，确定各自的大小关系。因目标访问权重越大，则目标服务节点所能提供的服务越好；从而从大小关系中确定目标访问权重的最大值，将具有此最大目标访问权重的目标访问节点确定为服务请求提供服务的节点，并调用其对服务请求进行响应，以使服务请求的服务最优。

此外，请参照图2，本申请提供一种服务节点的管理装置，在本申请服务节点的管理装置第一实施例中，所述服务节点的管理装置包括：

读取模块10，用于读取服务节点的多个额定参数，以及与各所述额定参数对应的参数系数，并根据各所述额定参数和各所述参数系数，确定所述服务节点的多个性能参数，以及所述服务节点的极限容量；

确定模块20，用于读取所述服务节点的多个实时参数，并根据各所述实时参数以及所述性能参数确定所述服务节点的实时容量；

管理模块30，用于根据所述极限容量和所述实时容量，确定所述服务节点的冗余度，并根据所述冗余度与预设阈值区间的关系，对所述服务节点进行增减管理。

本实施例的服务节点的管理装置，读取模块10根据所读取的服务节点的多个额定参数以及与各额定参数对应的各参数系数，确定服务节点的多个性能参数

以及极限容量 ;同时确定模块20还根据所读取的服务器的多个实时参数以及所确定的多个性能参数进一步确定服务器的实时容量 ,进而管理模块30根据极限容量和实时容量确定服务节点的冗余度 ;因实时容量表征服务节点中各资源的实时使用程度 ,从而通过冗余度可反映服务节点实时所能提供的访问服务量 ;而预设阈值区间表征服务节点所提供访问服务量的合理范围 ,用冗余度和预设阈值区间的关系 ,判断服务节点实时所提供服务量是否合理 ,进而根据合理性对服务节点进行增减管理 ;避免了人工重复操作进行增减管理 ,实现对服务节点的自动增减管理部署 ,降低了运维成本 ,提高了混合云架构中各服务节点的高可用性。

其中 ,上述服务节点的管理装置的各虚拟功能模块存储于图3所示服务节点的管理设备的存储器1005中 ,处理器1001执行服务节点的管理程序时 ,实现图2所示实施例中各个模块的功能。

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

参照图3 ,图3是本申请实施例方法涉及的硬件运行环境的设备结构示意图。

本申请实施例服务节点的管理设备可以是 PC(personal computer ,个人计算机) ,也可以是智能手机、平板电脑、电子书阅读器、便携计算机等终端设备。

如图 3 所示 ,该服务节点的管理设备可以包括 :处理器 1001 ,例如 CPU (Central Processing Unit ,中央处理器) ,存储器 1005 ,通信总线 1002。其中 ,通信总线 1002 用于实现处理器 1001 和存储器 1005 之间的连接通信。存储器 1005 可以是高速 RAM (random access memory ,随机存取存储器) ,也可以是稳定的存储器 (non-volatile memory) ,例如磁盘存储器。存储器 1005 可选的还可以是独立于前述处理器 1001 的存储装置。

可选地 ,该服务节点的管理设备还可以包括用户接口、网络接口、摄像头、

RF (Radio Frequency , 射频) 电路 , 传感器、音频电路、WiFi (Wireless Fidelity , 无线宽带) 模块等等。用户接口可以包括显示屏 (Display)、输入单元比如键盘 (Keyboard) , 可选用户接口还可以包括标准的有线接口、无线接口。网络接口可选的可以包括标准的有线接口、无线接口 (如WI-FI接口)。

本领域技术人员可以理解,图3中示出的服务节点的管理设备结构并不构成对服务节点的管理设备的限定,可以包括比图示更多或更少的部件,或者组合某些部件,或者不同的部件布置。

如图3所示,作为一种计算机可读存储介质的存储器1005中可以包括操作系统、网络通信模块以及服务节点的管理程序。操作系统是管理和控制服务节点的管理设备硬件和软件资源的程序,支持服务节点的管理程序以及其它软件和/或程序的运行。网络通信模块用于实现存储器1005内部各组件之间的通信,以及与服务节点的管理设备中其它硬件和软件之间通信。

在图3所示的服务节点的管理设备中,处理器1001用于执行存储器1005中存储的服务节点的管理程序,实现上述服务节点的管理方法各实施例中的步骤。

本申请提供了一种计算机可读存储介质,所述计算机可读存储介质存储有一个或者一个以上程序,所述一个或者一个以上程序还可被一个或者一个以上的处理器执行以用于实现上述服务节点的管理方法各实施例中的步骤。

还需要说明的是,在本文中,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、物品或者装置不仅包括那些要素,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、物品或者装置所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括该要素的过程、方法、物品或者装置中还存在另外的相同要素。

上述本申请实施例序号仅仅为了描述,不代表实施例的优劣。

通过以上的实施方式的描述,本领域的技术人员可以清楚地了解到上述实施例方法可借助软件加必需的通用硬件平台的方式来实现,当然也可以通过硬件,但很多情况下前者是更佳的实施方式。基于这样的理解,本申请的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来,该计算机软件产品存储在如上所述的一个计算机可读存储介质(如ROM/RAM、磁碟、光盘)中,包括若干指令用以使得一台终端设备(可以是手机,计算机,服务器,或者网络设备等)执行本申请各个实施例所述的方法。

以上所述仅为本申请的优选实施例,并非因此限制本申请的专利范围,凡是在本申请的构思下,利用本申请说明书及附图内容所作的等效结构变换,或直接/间接运用在其他相关的技术领域均包括在本申请的专利保护范围内。

权 利 要 求 书

1、一种服务节点的管理方法，其特征在于，所述服务节点的管理方法包括以下步骤：

读取服务节点的多个额定参数，以及与各所述额定参数对应的参数系数，并根据各所述额定参数和各所述参数系数，确定所述服务节点的多个性能参数，以及所述服务节点的极限容量；

读取所述服务节点的多个实时参数，并根据各所述实时参数以及所述性能参数确定所述服务节点的实时容量；

根据所述极限容量和所述实时容量，确定所述服务节点的冗余度，并根据所述冗余度与预设阈值区间的关系，对所述服务节点进行增减管理。

2、如权利要求1所述的服务节点的管理方法，其特征在于，所述根据各所述额定参数和各所述参数系数，确定所述服务节点的多个性能参数，以及所述服务节点的极限容量的步骤包括：

将各所述额定参数和各所述参数系数基于参数属性进行分组，并根据各组中的所述额定参数和所述参数系数，生成所述服务节点的多个性能参数；

读取与各所述性能参数对应的预设权重值，并将各所述预设权重值和各所述性能参数输入到第一预设公式中，计算所述服务节点的极限容量；

其中所述第一预设公式： $ValX = k_1lgA_1 + k_2lgA_2 + k_3lgA_3 + k_4lgA_4$ ， $ValX$ 为极限容量， A_1 、 A_2 、 A_3 和 A_4 为各所述性能参数， k_1 、 k_2 、 k_3 和 k_4 为各所述预设权重值， lg 为取对数值。

3、如权利要求2所述的服务节点的管理方法，其特征在于，所述根据各所

述实时参数以及所述性能参数确定所述服务节点的实时容量的步骤包括：

将各所述实时参数、各所述性能参数以及各所述预设权重值输入到第二预设公式中，计算所述服务节点的实时容量；

其中所述第二预设公式：

$$\text{ValY} = k_1 \lg A_1 * Q_1 + k_2 \lg A_2 * Q_2 + k_3 \lg A_3 * Q_3 + k_4 \lg A_4 * Q_4$$
，ValY为实时容量，A1、A2、A3和A4为各所述性能参数，k1、k2、k3和k4为各所述预设权重值，Q1、Q2、Q3和Q4为各实时参数，lg为取对数值。

4、如权利要求3所述的服务节点的管理方法，其特征在于，所述根据所述极限容量和所述实时容量，确定所述服务节点的冗余度的步骤包括：

将所述极限容量和所述实时容量输入到第三预设公式中，计算所述服务节点的冗余度；

其中所述第三预设公式： $R = 1 - \text{ValY} / \text{ValX}$ ，R为冗余度，ValY为实时容量，ValX为极限容量。

5、如权利要求1所述的服务节点的管理方法，其特征在于，所述根据所述冗余度与预设阈值区间的关系，对所述服务节点进行增减管理的步骤包括：

将所述冗余度和预设阈值区间对比，判断所述冗余度是否在预设阈值区间内；

当所述冗余度小于预设阈值区间下边界数值时，增加与所述服务节点对应的新服务节点；

当所述冗余度大于预设阈值区间上边界数值时，对所述服务节点进行删除操作。

6、如权利要求5所述的服务节点的管理方法，其特征在于，所述增加与所

述服务节点对应的新服务节点的步骤之后包括：

读取与所述新服务节点对应的新增额定参数，并将所述新增额定参数与其他服务节点的额定参数对比，确定所述其他服务节点的额定参数中与所述新增额定参数匹配的目标额定参数；

确定所述目标额定参数归属的服务节点，并根据所述归属的服务节点所具有的访问权重，为所述新服务节点分配访问权重。

7、如权利要求6所述的服务节点的管理方法，其特征在于，所述为所述新服务节点分配访问权重的步骤之后包括：

当接收到服务请求时，根据所述服务请求的类型确定与所述服务请求对应的目标服务节点，并判断所述目标服务节点是否存在多个；

若存在多个所述目标服务节点，则判断各所述目标服务节点所具有目标访问权重的大小关系，并调用具有最大所述目标访问权重的目标访问节点对所述服务请求进行响应。

8、一种服务节点的管理装置，其特征在于，所述服务节点的管理装置包括：

读取模块，用于读取服务节点的多个额定参数，以及与各所述额定参数对应的参数系数，并根据各所述额定参数和各所述参数系数，确定所述服务节点的多个性能参数，以及所述服务节点的极限容量；

确定模块，用于读取所述服务节点的多个实时参数，并根据各所述实时参数以及所述性能参数确定所述服务节点的实时容量；

管理模块，用于根据所述极限容量和所述实时容量，确定所述服务节点的冗余度，并根据所述冗余度与预设阈值区间的关系，对所述服务节点进行增减管理。

9、一种服务节点的管理设备 ,其特征在于 ,所述服务节点的管理设备包括 :
存储器、处理器、通信总线以及存储在所述存储器上的服务节点的管理程序 ;
所述通信总线用于实现处理器和存储器之间的连接通信 ;
所述处理器用于执行所述服务节点的管理程序 ,以实现以下步骤 :
读取服务节点的多个额定参数 ,以及与各所述额定参数对应的参数系数 ,
并根据各所述额定参数和各所述参数系数 ,确定所述服务节点的多个性能参数 ,
以及所述服务节点的极限容量 ;
读取所述服务节点的多个实时参数 ,并根据各所述实时参数以及所述性能
参数确定所述服务节点的实时容量 ;
根据所述极限容量和所述实时容量 ,确定所述服务节点的冗余度 ,并根据
所述冗余度与预设阈值区间的关系 ,对所述服务节点进行增减管理。

10、如权利要求9所述的服务节点的管理设备 ,其特征在于 ,所述根据各所
述额定参数和各所述参数系数 ,确定所述服务节点的多个性能参数 ,以及所述服
务节点的极限容量的步骤包括 :

将各所述额定参数和各所述参数系数基于参数属性进行分组 ,并根据各组
中的所述额定参数和所述参数系数 ,生成所述服务节点的多个性能参数 ;

读取与各所述性能参数对应的预设权重值 ,并将各所述预设权重值和各所
述性能参数输入到第一预设公式中 ,计算所述服务节点的极限容量 ;

其中所述第一预设公式 : $ValX = k_1lgA_1 + k_2lgA_2 + k_3lgA_3 + k_4lgA_4$, $ValX$ 为
极限容量 , A_1 、 A_2 、 A_3 和 A_4 为各所述性能参数 , k_1 、 k_2 、 k_3 和 k_4 为各所述预设
权重值 , lg 为取对数值。

11、如权利要求10所述的服务节点的管理设备 ,其特征在于 ,所述根据各

所述实时参数以及所述性能参数确定所述服务节点的实时容量的步骤包括：

将各所述实时参数、各所述性能参数以及各所述预设权重值输入到第二预设公式中，计算所述服务节点的实时容量；

其中所述第二预设公式：

$ValY = k_1 \lg A_1 * Q_1 + k_2 \lg A_2 * Q_2 + k_3 \lg A_3 * Q_3 + k_4 \lg A_4 * Q_4$ ， $ValY$ 为实时容量， A_1 、 A_2 、 A_3 和 A_4 为各所述性能参数， k_1 、 k_2 、 k_3 和 k_4 为各所述预设权重值， Q_1 、 Q_2 、 Q_3 和 Q_4 为各实时参数， $\lg$ 为取对数值。

12、如权利要求11所述的服务节点的管理设备，其特征在于，所述根据所述极限容量和所述实时容量，确定所述服务节点的冗余度的步骤包括：

将所述极限容量和所述实时容量输入到第三预设公式中，计算所述服务节点的冗余度；

其中所述第三预设公式： $R = 1 - ValY/ValX$ ， R 为冗余度， $ValY$ 为实时容量， $ValX$ 为极限容量。

13、如权利要求9所述的服务节点的管理设备，其特征在于，所述根据所述冗余度与预设阈值区间的关系，对所述服务节点进行增减管理的步骤包括：

将所述冗余度和预设阈值区间对比，判断所述冗余度是否在预设阈值区间内；

当所述冗余度小于预设阈值区间下边界数值时，增加与所述服务节点对应的新服务节点；

当所述冗余度大于预设阈值区间上边界数值时，对所述服务节点进行删除操作。

14、如权利要求13所述的服务节点的管理设备，其特征在于，所述增加与

所述服务节点对应的新服务节点的步骤之后,所述处理器用于执行所述服务节点的管理程序,以实现以下步骤:

读取与所述新服务节点对应的新增额定参数,并将所述新增额定参数与其他服务节点的额定参数对比,确定所述其他服务节点的额定参数中与所述新增额定参数匹配的目标额定参数;

确定所述目标额定参数归属的服务节点,并根据所述归属的服务节点所具有的访问权重,为所述新服务节点分配访问权重。

15、一种计算机可读存储介质,其特征在于,所述计算机可读存储介质上存储有服务节点的管理程序,所述服务节点的管理程序被处理器执行,实现以下步骤:

读取服务节点的多个额定参数,以及与各所述额定参数对应的参数系数,并根据各所述额定参数和各所述参数系数,确定所述服务节点的多个性能参数,以及所述服务节点的极限容量;

读取所述服务节点的多个实时参数,并根据各所述实时参数以及所述性能参数确定所述服务节点的实时容量;

根据所述极限容量和所述实时容量,确定所述服务节点的冗余度,并根据所述冗余度与预设阈值区间的关系,对所述服务节点进行增减管理。

16、如权利要求15所述的计算机可读存储介质,其特征在于,所述根据各所述额定参数和各所述参数系数,确定所述服务节点的多个性能参数,以及所述服务节点的极限容量的步骤包括:

将各所述额定参数和各所述参数系数基于参数属性进行分组,并根据各组中的所述额定参数和所述参数系数,生成所述服务节点的多个性能参数;

读取与各所述性能参数对应的预设权重值，并将各所述预设权重值和各所述性能参数输入到第一预设公式中，计算所述服务节点的极限容量；

其中所述第一预设公式： $ValX = k_1lgA_1 + k_2lgA_2 + k_3lgA_3 + k_4lgA_4$ ， $ValX$ 为极限容量， A_1 、 A_2 、 A_3 和 A_4 为各所述性能参数， k_1 、 k_2 、 k_3 和 k_4 为各所述预设权重值， lg 为取对数值。

17、如权利要求16所述的计算机可读存储介质，其特征在于，所述根据各所述实时参数以及所述性能参数确定所述服务节点的实时容量的步骤包括：

将各所述实时参数、各所述性能参数以及各所述预设权重值输入到第二预设公式中，计算所述服务节点的实时容量；

其中所述第二预设公式： $ValY = k_1lgA_1 * Q_1 + k_2lgA_2 * Q_2 + k_3lgA_3 * Q_3 + k_4lgA_4 * Q_4$ ， $ValY$ 为实时容量， A_1 、 A_2 、 A_3 和 A_4 为各所述性能参数， k_1 、 k_2 、 k_3 和 k_4 为各所述预设权重值， Q_1 、 Q_2 、 Q_3 和 Q_4 为各实时参数， lg 为取对数值。

18、如权利要求17所述的计算机可读存储介质，其特征在于，所述根据所述极限容量和所述实时容量，确定所述服务节点的冗余度的步骤包括：

将所述极限容量和所述实时容量输入到第三预设公式中，计算所述服务节点的冗余度；

其中所述第三预设公式： $R = 1 - ValY/ValX$ ， R 为冗余度， $ValY$ 为实时容量， $ValX$ 为极限容量。

19、如权利要求15所述的计算机可读存储介质，其特征在于，所述根据所述冗余度与预设阈值区间的关系，对所述服务节点进行增减管理的步骤包括：

将所述冗余度和预设阈值区间对比，判断所述冗余度是否在预设阈值区间

内；

当所述冗余度小于预设阈值区间下边界数值时，增加与所述服务节点对应的新服务节点；

当所述冗余度大于预设阈值区间上边界数值时，对所述服务节点进行删除操作。

20、如权利要求15所述的计算机可读存储介质，其特征在于，所述增加与
所述服务节点对应的新服务节点的步骤之后，所述服务节点的管理程序被处理器
执行，实现以下步骤：

读取与所述新服务节点对应的新增额定参数，并将所述新增额定参数与其他
服务节点的额定参数对比，确定所述其他服务节点的额定参数中与所述新增额
定参数匹配的目标额定参数；

确定所述目标额定参数归属的服务节点，并根据所述归属的服务节点所具
有的访问权重，为所述新服务节点分配访问权重。

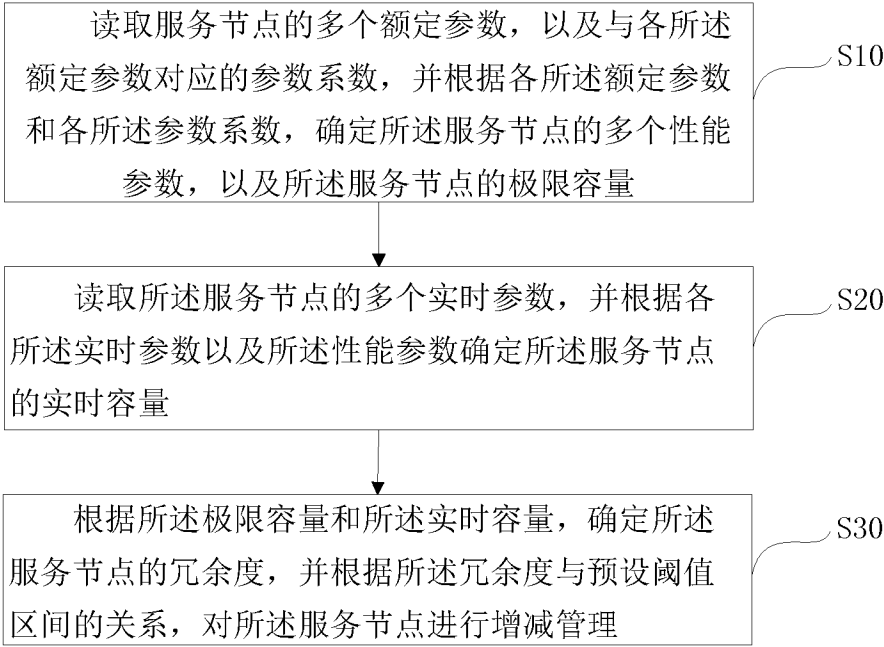


图 1

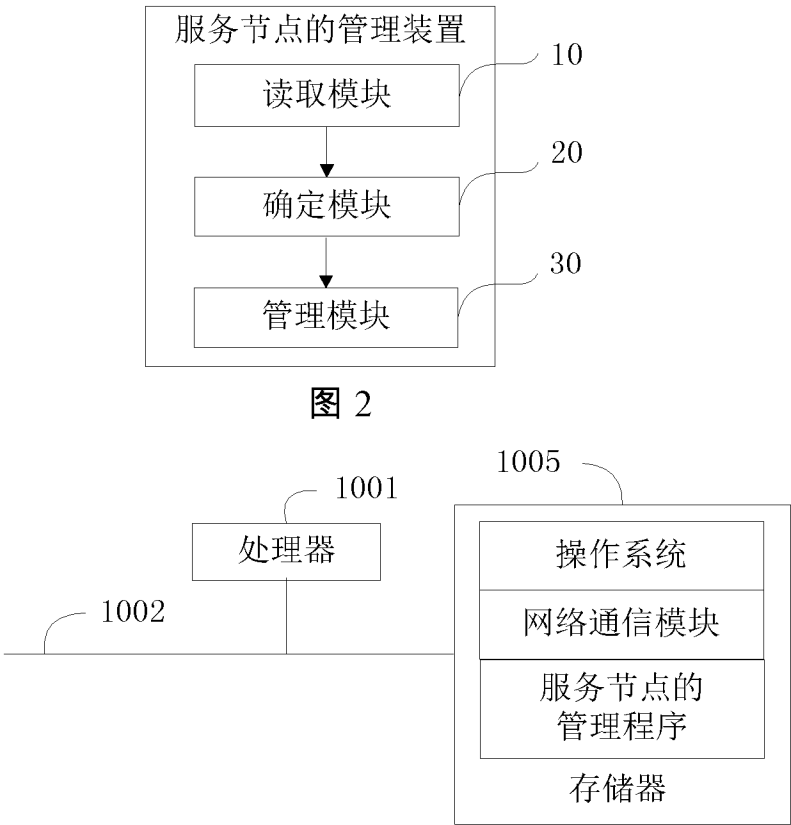


图 3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/122637

A. CLASSIFICATION OF SUBJECT MATTER

H04L 29/08(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)

H04L

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)

WPI, EPODOC, CNPAT, CNKI, IEEE: 服务, 节点, 参数, 性能, 管理, 云平台, 网络, 容量, 增加, 增减, 删除, 收缩, 扩展,
service, node, parameter, management, cloud, platform, network, capacity, add to, reduce, delete, shrink, expand**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 106533839 A (ZHENGZHOU YUNHAI INFORMATION TECHNOLOGY CO., LTD.) 22 March 2017 (2017-03-22) description, paragraphs [0055]-[0087]	1-20
A	CN 104010028 A (SOUTH CHINA UNIVERSITY OF TECHNOLOGY) 27 August 2014 (2014-08-27) entire document	1-20
A	CN 105393518 B (HUAWEI TECHNOLOGIES CO., LTD.) 19 March 2019 (2019-03-19) entire document	1-20
A	US 2013046885 A1 (CA, INCORPORATED) 21 February 2013 (2013-02-21) 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

03 June 2019

Date of mailing of the international search report

17 June 2019

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2018/122637

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	106533839	A	22 March 2017	None			
CN	104010028	A	27 August 2014	None			
CN	105393518	B	19 March 2019	WO	2015196369	A1	30 December 2015
US	2013046885	A1	12 June 2012	US	2011213880	A1	01 September 2011
				US	2006013134	A1	19 January 2006
				US	2014207951	A1	24 July 2014

国际检索报告

国际申请号

PCT/CN2018/122637

A. 主题的分类

H04L 29/08 (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)

H04L

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))

WPI, EPDOC, CNPAT, CNKI, IEEE: 服务, 节点, 参数, 性能, 管理, 云平台, 网络, 容量, 增加, 增减, 删除, 收缩, 扩展, service, node, parameter, management, cloud, platform, network, capacity, add to, reduce, delete, shrink, expand

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
X	CN 106533839 A (郑州云海信息技术有限公司) 2017年 3月 22日 (2017 - 03 - 22) 说明书第0055-0087段	1-20
A	CN 104010028 A (华南理工大学) 2014年 8月 27日 (2014 - 08 - 27) 全文	1-20
A	CN 105393518 B (华为技术有限公司) 2019年 3月 19日 (2019 - 03 - 19) 全文	1-20
A	US 2013046885 A1 (CA, INC.) 2013年 2月 21日 (2013 - 02 - 21) 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

2019年 6月 3日

国际检索报告邮寄日期

2019年 6月 17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

授权官员

杨庆丽

电话号码 86-(10)-53961433

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/122637

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	106533839	A	2017年 3月 22日	无			
CN	104010028	A	2014年 8月 27日	无			
CN	105393518	B	2019年 3月 19日	W0	2015196369	A1	2015年 12月 30日
US	2013046885	A1	2012年 6月 12日	US	2011213880	A1	2011年 9月 1日
				US	2006013134	A1	2006年 1月 19日
				US	2014207951	A1	2014年 7月 24日