

(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
2018 年 1 月 25 日 (25.01.2018)



(10) 国际公布号
WO 2018/014566 A1

(51) 国际专利分类号:
G06F 9/455 (2006.01)

(21) 国际申请号: PCT/CN2017/076514

(22) 国际申请日: 2017 年 3 月 14 日 (14.03.2017)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201610584168.0 2016 年 7 月 22 日 (22.07.2016) 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八卦岭八卦三路平安大厦, Guangdong 518000 (CN)。

(72) 发明人: 何涛涛 (HE, Taotao); 中国广东省深圳市福田区八卦岭八卦三路平安大厦, Guangdong 518000 (CN)。

(74) 代理人: 深圳市精英专利事务所 (SHENZHEN TALENT PATENT SERVICE); 中国广东省深圳市福田区深南中路 6009 号绿景广场 B 栋 20 层 B, Guangdong 518000 (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, 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,

(54) Title: LOAD BALANCING METHOD AND APPARATUS, COMPUTER-READABLE STORAGE MEDIUM, AND SYSTEM

(54) 发明名称: 一种负载均衡的方法、装置、计算机可读存储介质及系统

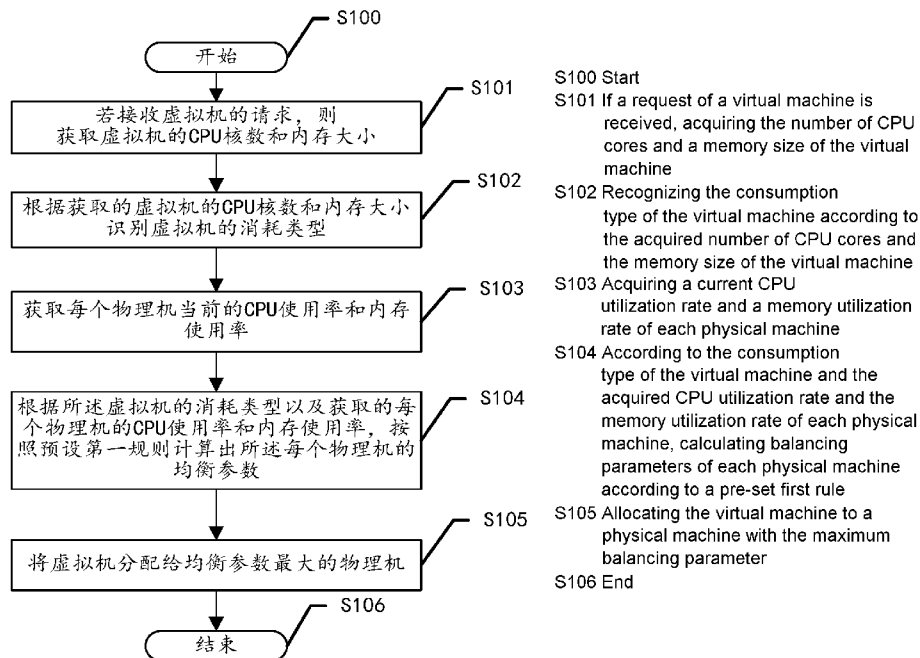


图 1

(57) Abstract: A load balancing method and apparatus, a computer-readable storage medium, and a system. The method comprises: if a request of a virtual machine is received, acquiring the number of CPU cores and a memory size of the virtual machine (S101); recognizing the consumption type of the virtual machine according to the acquired number of CPU cores and the memory size of the virtual machine (S102); acquiring a current CPU utilization rate and a memory utilization rate of each physical machine (S103); according to the consumption type of the virtual machine and the acquired CPU utilization rate and the memory utilization rate of each physical machine, calculating balancing parameters of each physical machine according to a pre-set first rule (S104); and allocating

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))。

the virtual machine to a physical machine with the maximum balancing parameter (S105). According to the method, by calculating balancing parameters, a physical machine is allocated such that a virtual machine operates, thereby realizing load balancing of the physical machine.

(57) 摘要: 一种负载均衡方法、装置、计算机可读存储介质及系统, 方法包括: 若接收虚拟机的请求, 获取虚拟机的CPU核数和内存大小 (S101); 根据获取的虚拟机的CPU核数和内存大小识别虚拟机的消耗类型 (S102); 获取每个物理机当前的CPU使用率和内存使用率 (S103); 根据虚拟机的消耗类型以及获取的每个物理机的CPU使用率和内存使用率, 按照预设第一规则计算出每个物理机的均衡参数 (S104); 将虚拟机分配给均衡参数最大的物理机 (S105)。上述方法通过计算出均衡参数来分配物理机以供虚拟机运行, 实现物理机的负载均衡

说明书

一种负载均衡的方法、装置、计算机可读存储介质及系统

技术领域

本发明涉及计算机技术领域，尤其涉及一种负载均衡的方法、装置、计算机可读存储介质及系统。

背景技术

现有云计算平台上默认采用的虚拟机分配算法是随即算法，容易导致各个物理机的负载差异很大，部分物理机因为负载过大而带来宕机的风险，或者部分物理机因某种负载过小而带来资源的浪费。

发明内容

本发明实施例提供一种负载均衡的方法及其装置，可以实现各个物理机的负载均衡。

第一方面，本发明实施例提供了一种负载均衡的方法，该方法包括：若接收虚拟机的请求，则获取虚拟机的 CPU 核数和内存大小；根据获取的虚拟机的 CPU 核数和内存大小识别虚拟机的消耗类型；获取每个物理机当前的 CPU 使用率和内存使用率；根据虚拟机的消耗类型以及获取的每个物理机的 CPU 使用率和内存使用率，按照预设第一规则计算出每个物理机的均衡参数；将虚拟机分配给均衡参数最大的物理机。

另一方面，本发明实施例提供了一种负载均衡装置，该装置包括获取模块、识别模块、第一计算模块和分配模块，其中，获取模块用于若接收虚拟机的请求，获取虚拟机的 CPU 核数和内存大小以及还用于获取每个物理机当前的 CPU 使用率和内存使用率；识别模块用于根据获取的虚拟机的 CPU 核数和内存大小识别虚拟机的消耗类型；第一计算模块用于根据虚拟机的消耗类型以及获取的每个物理机的 CPU 使用率和内存使用率，按照预设第一规则计算出每个物理机的均衡参数；分配模块用于将虚拟机分配给均衡参数最大的物理机。

另一方面，本发明实施例提供了一种计算机可读存储介质，计算机可读存储介质存储有一个或者一个以上程序，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，以执行以下操作：

若接收虚拟机的请求，则获取虚拟机的 CPU 核数和内存大小；根据获取的虚拟机的 CPU 核数和内存大小识别虚拟机的消耗类型；获取每个物理机当前的 CPU 使用率和内存使用率；根据虚拟机的消耗类型以及获取的每个物理机的 CPU 使用率和内存使用率，按照预设第一规则计算出每个物理机的均衡参数；将虚拟机分配给均衡参数最大的物理机。

另一方面，本发明实施例提供了一种负载均衡系统，系统包括一个或者一个以上的处理器，以及与一个或者一个以上处理器连接的一个或者一个以上存储器，

存储器，用于存储用于实现负载均衡的一个或者一个以上程序；处理器，用于执行存储器中存储的程序，以执行以下操作：

若接收虚拟机的请求，则获取虚拟机的 CPU 核数和内存大小；根据获取的虚拟机的 CPU 核数和内存大小识别虚拟机的消耗类型；获取每个物理机当前的 CPU 使用率和内存使用率；根据虚拟机的消耗类型以及获取的每个物理机的 CPU 使用率和内存使用率，按照预设第一规则计算出每个物理机的均衡参数；将虚拟机分配给均衡参数最大的物理机。

本发明实施例通过计算出每个物理机的均衡参数，再根据均衡参数确定放置虚拟机的物理机，由于均衡参数与物理机的 CPU 使用率、内存使用率和虚拟机的消耗类型相关，其中每个物理机的 CPU 使用率以及内存使用率与物理机集群的均衡相关、虚拟机的消耗类型又与每个物理机内部 CPU 和内存均衡相关，因此该方法即考虑了物理机集群整体的均衡又考虑了单个物理机内部资源的负载均衡，因此更有效的实现了物理机集群的负载均衡。

附图说明

为了更清楚地说明本发明实施例技术方案，下面将对实施例描述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的前提下，还可以根据这些附图获得其他的附图。

图1 是本发明实施例提供一种负载均衡的方法的示意流程图；

图2 是提供的一种物理机集群负载不均衡的折线图；

图3 是图 1 的子流程示意图；

图4是图1的子流程示意图；

图5是图4的子流程示意图；

图6是本发明实施例提供一种物理机集群负载均衡的折线图；

图7是本发明实施例提供一种物理机集群负载均衡的散点图；

图8是本发明实施例提供一种负载均衡装置的组成示意图。

图9是本发明实施例提供第一计算单元的组成示意图。

图10是本发明实施例提供的一种负载均衡系统的硬件结构图。

具体实施方式

下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，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，都属于本发明保护的范围。

应当理解，当在本说明书和所附权利要求书中使用时，术语“包括”和“包含”指示所描述特征、整体、步骤、操作、元素和/或组建的存在，但并不排除一个或多个其它特征、整体、步骤、操作、元素、组建和/或其集合的存在或添加。还应当理解，在此本发明说明书中所使用的术语仅仅是出于描述特定实施例的目的而并不意在限制本发明。如在本发明说明书和所附权利要求书中所使用的那样，除非上下文清楚地指明其它情况，否则单数形式的“一”、“一个”及“该”意在包括复数形式。还应当进一步理解，在本发明说明书和所附权利要求书中使用的术语“和/或”是指相关联列出的项中的一个或多个的任何组合以及所有可能组合，并且包括这些组合。

在云计算平台的数据中心有多主机组成的集群，一个集群中包括若干个物理机，物理机是用于运行虚拟机的主机。由于不断接收来自虚拟机的请求，需要分配物理机来放置虚拟机，因此需要根据物理机的负载从物理机集群中选择负载相对小的物理机来运行虚拟机，进而实现物理机集群的负载均衡。请参看图1，本发明实施例提供的一种负载均衡的方法，如图所示，负载均衡的方法包括S100~S106。

S100，开始。

S101，若接收虚拟机的请求，则获取虚拟机的CPU核数和内存大小。CPU

(central processing unit) 是中央处理器的缩写, CPU 核数用于表示 CPU 核心的数量。

S102, 根据获取的虚拟机的 CPU 核数和内存大小识别虚拟机的消耗类型。具体的, 虚拟机的消耗类型包括常规类型、CPU 消耗型以及内存消耗型, 不同消耗类型的虚拟机的对资源的占用不同, CPU 消耗型所占用的 CPU 资源相比内存消耗型和常规类型占用的 CPU 资源多; 内存消耗型所占用的内存资源相比 CPU 消耗型和常规类型占用的内存资源多。而部分物理机内部 CPU 和内存的资源存在不均衡, 因此需要考虑物理机内部资源的均衡以及虚拟机的消耗类型, 将虚拟机放置于较佳的物理机上来实现物理机内部的资源均衡。具体实现过程为根据获取的 CPU 核数和内存大小的比值来判断虚拟机的消耗类型。

S103, 获取每个物理机当前的 CPU 使用率和内存使用率。本实施例中, 用 U_i 表示第 i 个物理机的 CPU 使用率, M_i 表示第 i 个物理机的内存使用率。

S104, 根据虚拟机的消耗类型以及获取的每个物理机的 CPU 使用率和内存使用率, 按照预设第一规则计算出每个物理机的均衡参数。本实施例中, 用 S_2 表示均衡参数。

均衡参数 S_2 用于表示综合考虑物理机集群整体的负载均衡以及物理机内部 CPU 和内存资源的负载均衡时, 物理机的选择性。物理机的均衡参数 S_2 越大, 则优先考虑该物理机来放置虚拟机。

预设第一规则用于表示虚拟机的消耗类型、获取的每个物理机的 CPU 使用率和内存使用率与均衡参数的关系, 针对不同消耗类型的虚拟机, 物理机的均衡参数与物理机的 CPU 使用率和内存使用率的关系式不同。具体的, 若虚拟机的消耗类型为 CPU 消耗型或者内存消耗型, 物理机的均衡参数还与物理机的 CPU 使用率以及内存使用率的比值相关, 若虚拟机的消耗类型为常规类型, 物理机的均衡参数与物理机的 CPU 使用率以及内存使用率的比值无关。

S105, 将虚拟机分配给均衡参数最大的物理机。由于均衡参数 S_2 用于表示综合考虑物理机集群整体的负载均衡以及物理机内部 CPU 和内存资源的负载均衡时, 物理机的选择性。物理机均衡参数 S_2 越大, 则优先考虑该物理机来放置虚拟机。因此将虚拟机分配给均衡参数 S_2 最大的物理机, 以此来实现物理机集群的负载均衡, 应当理解, 物理机集群的负载均衡包括物理机彼此之间的负载均衡以及单个物理机内部资源的均衡。

S106, 结束。

需要说明的是, 基于物理机内部存在 CPU 和内存资源的不均衡以及不同消耗类型的虚拟机消耗 CPU 或者内存的不同, S102 中识别虚拟机的消耗类型是有必要的。请参看图 2, 如图 2 所示图中所示 L1 为 CPU 使用率, L2 为内存使用率, 其中 16~20 这几个物理机出现了 CPU 使用率高而内存使用率低的情况, 相差较大, 会导致这几个物理机的资源利用率不高, 在满负荷时, 容易造成 CPU 或内存较大的浪费, 且容易导致物理机内 CPU 和内存的均衡问题。因此需要考虑单个物理机内部的内存和 CPU 资源的负载均衡。首先需要识别虚拟机的类型, 进而根据虚拟机的消耗类型和各个物理机内部资源的情况分配适宜的物理机, 例如虚拟机为 CPU 消耗型时, 应该选择内存相对 CPU 消耗高的物理机来放置虚拟机; 虚拟机为内存消耗型时, 应该选择 CPU 相对内存消耗高的物理机来放置虚拟机。请参看图 3, 本发明实施例提供的图 1 的子流程示意图, 如图所示, S102 根据获取的 CPU 核数和内存大小识别虚拟机的消耗类型, 包括:

S301, 计算获取的虚拟机的 CPU 核数与内存大小的比值。

S302, 判断比值是在预设第一范围还是预设第二范围或是预设第三范围。其中预设第二范围大于预设第一范围和预设第三范围, 预设第一范围大于预设第三范围。本实施例中优选预设第一范围为: 大于 0.25 且小于 1, 预设第二范围为: 大于或等于 1, 预设第三范围为小于或等于 0.25。

S303, 若比值在预设第一范围, 识别虚拟机的消耗类型为常规类型。

S304, 若比值在预设第二范围, 识别虚拟机的消耗类型为 CPU 消耗型。

S305, 若比值在预设第三范围, 识别虚拟机的消耗类型为内存消耗型。

上述方法, 根据虚拟机的 CPU 核数以及内存大小识别消耗类型, 所识别的消耗类型用于考虑物理机内部资源的均衡时选择物理机。

还需要说明的是, S104 中计算的均衡参数 S2 用于表示综合考虑物理机集群整体的负载均衡以及物理机内部 CPU 和内存资源的负载均衡时, 物理机的选择性。本实施例中, 用第一均衡度 S1 表示考虑物理机集群整体的负载均衡时, 物理机的选择性; 用第二均衡度 t 表示考虑物理机内部 CPU 和内存资源的负载均衡时, 物理机的选择性。

请参看图 4, 如图所示, S104 根据所述虚拟机的消耗类型以及获取的每个物理机的 CPU 使用率和内存使用率, 按照预设第一规则计算出所述每个物理机

的均衡参数，包括 S401~S405：

S401，将获取的每个物理机的 CPU 使用率和内存使用率按照预设第二规则计算出每个物理机的第一均衡度。

具体的，第一均衡度 S1 用于表示考虑物理机集群整体的负载均衡时，物理机的选择性。应当理解，若只考虑整个集群整体的负载均衡，而不考虑各个物理机内部资源均衡，将选择当前第一均衡度 S1 最大的物理机来放置虚拟机。

应当理解若只考虑内存，而不考虑 CPU 的影响因素时，为了物理机集群的负载均衡，应当将虚拟机分配给当前内存使用率最低的物理机，同理，只考虑 CPU，不考虑内存的影响时，应当将虚拟机的请求给当前 CPU 使用率最低的物理机，本实施例同时考虑 CPU 以及内存的影响，因此第一均衡度 S1 与 CPU 使用率 U_i 和内存使用率 M_i 均相关，第一均衡度 S1 与 CPU 使用率以及内存使用率成函数关系。

预设第二规则用于表示每个物理机的 CPU 使用率和内存使用率与第一均衡度的关系，本实施例中，预设第二规则具体为：若 CPU 和内存对第一均衡度的影响程度相同时或者两者之间的影响程度差可以忽略，一个物理机的 CPU 使用率和内存使用率分别大于另一物理机的 CPU 使用率和内存使用率，则该一个物理机的第一均衡度 S1 小于另一物理机的第一均衡度 S1；若 CPU 和内存对第一均衡度的影响差异不可忽略，第一均衡度与 CPU 使用率、内存使用率以及 CPU 相对内存的选择权重 w_1 有关，选择权重 W_1 用于表示 CPU 和内存对第一均衡度的影响程度差异。

请参照图 5，如图所示，S401 将获取的每个物理机的 CPU 使用率和内存使用率按照预设第二规则计算出每个物理机的第一均衡度，包括 S501~S504：

S501，计算每个物理机当前的 CPU 使用率的倒数和内存使用率的倒数。

S502，对所有物理机的 CPU 使用率的倒数进行求和得到 CPU 总权重，对所有物理机的内存使用率的倒数进行求和得到内存总权重。

S503，计算每个物理机的 CPU 使用率的倒数与 CPU 总权重的比值得到 CPU 占比 c 以及每个物理机的内存使用率的倒数与内存总权重的比值得到内存占比 m 。例如三个物理机的 CPU 使用率为 $1/3$ 、 $1/3$ 、以及 $1/4$ ，则此三个物理机的 CPU 使用率的倒数为 3、3、4，CPU 总权重为 $3+3+4=10$ ，三个物理机的 CPU 占比 c 分别为 $3/10$ 、 $3/10$ 以及 $4/10$ ，或者三个物理机的内存使用率为 $1/3$ 、 $1/3$ 、

以及 1/4, 则此三个物理机的内存使用率的倒数为 3、3、4, 内存总权重为 3+3+4=10, 三个物理机的内存占比 m 分别为 3/10、3/10 以及 4/10。

S504, 根据所计算出的每个物理机的 CPU 占比 c 以及内存占比 m 和预设的 CPU 相对内存的选择权重 $w1$ 计算出第一均衡度 $S1$ 。第一均衡度 $S1$ 与 CPU 占比 c 、内存占比 m 、CPU 相对内存的选择权重 $w1$ 的关系如下:

$$S1=c \times w1+m \times (1-w1)$$

其中预设的 CPU 相对内存的选择权重 $w1$ 表示优先考虑 CPU 影响因素或者内存影响因素, $w1$ 越大表示更优先考虑 CPU 因素对物理机集群整体负载均衡的影响, $w1$ 越小表示更优先考虑内存因素对物理机集群整体负载均衡的影响。本实施例优选忽略 CPU 因素或内存因素的影响区别, 将 $w1$ 设置为 0.5, 其它可行的实施例中, $w1$ 可以根据 CPU 因素或者内存因素的影响程度设置为其他数值。

由于本实施例中优选 $w1$ 等于 0.5, 此时预设第二规则为一个物理机的 CPU 使用率和内存使用率分别大于另一物理机的 CPU 使用率和内存使用率, 该一个物理机的第一均衡度 $S1$ 小于另一物理机的第一均衡度 $S1$ 。

S402, 若虚拟机的消耗类型是常规类型, 赋予每个物理机的均衡参数为对应物理机的第一均衡度。

具体的, 若虚拟机的消耗类型为常规类型, 因为虚拟机的消耗类型为常规类型时, 分配该虚拟机给物理机后, 不会对物理机内部资源造成额外的不均衡。

S403, 若虚拟机的消耗类型是 CPU 消耗型或内存消耗型, 根据获取的每个物理机的 CPU 使用率和内存使用率, 计算出 CPU 平均使用率和内存平均使用率。

具体的, 本实施例中用 U_{avg} 表示物理机集群的 CPU 平均使用率, M_{avg} 表示物理机集群的内存平均使用率。需要说明的是, 若虚拟机的消耗类型是 CPU 消耗型或内存消耗型, 计算物理机的均衡参数还包括 S404 和 S405。

S404, 根据获取的每个物理机的 CPU 使用率和内存使用率以及 CPU 平均使用率和内存平均使用率计算出每个物理机的第二均衡度。

用第二均衡度 t 表示考虑物理机内部 CPU 和内存资源的负载均衡时, 物理机的选择性。第二均衡度 t 等于物理机的 CPU 使用率 U_i 和内存使用率 M_i 的比值与 CPU 平均使用率 U_{avg} 和内存平均使用率 M_{avg} 比值的差, 关系如下所示:

$$t = \frac{U_i}{M_i} - \frac{U_{avg}}{M_{avg}}$$

当 $t > 0$ 是, 表示此物理机的 CPU 使用率相对内存使用率较多即 CPU 消耗相对内存消耗较多, t 越大表示该物理机内 CPU 和内存资源越不均衡; 当 $t < 0$ 时, 表示此物理机的内存使用率相对 CPU 使用率较多, 即内存消耗相对 CPU 消耗较多, t 的绝对值越大, 表示该物理机内 CPU 和内存资源越不均衡。

S405, 根据虚拟机的消耗类型、每个物理机的第一均衡度以及第二均衡度计算出每个物理机的均衡参数。

由于第一均衡度 $S1$ 用于表示考虑物理机集群整体的负载均衡时, 物理机的选择性; 而第二均衡度 t 表示考虑物理机内部 CPU 和内存资源的负载均衡时, 物理机的选择性。而虚拟机的消耗类型为 CPU 消耗型或者内存消耗型时, 虚拟机放置在物理机上后均会对物理机内部资源的均衡造成影响。因此均衡参数与虚拟机的消耗类型、第一均衡度以及第二均衡度相关。

若虚拟机的消耗类型为 CPU 消耗型, 均衡参数 $S2$ 有:

$$S2 = S1 - k \times t^n$$

其中, n 为奇数, k 为正数, k 以及 n 的设置是为了使物理机集群整体均衡以及物理机内部资源均衡的影响相当, 即使 $S1$ 与 $k \times t^n$ 的数值在同一数量级。应当理解, 若虚拟机的消耗类型为 CPU 消耗型, 应该选择内存相对 CPU 消耗高的物理机来运行虚拟机, 以此来均衡物理机的内部资源, 所选择的物理机的第二均衡度 t 值小于零。

若虚拟机的消耗类型为内存消耗型, 均衡参数 $S2$ 有:

$$S2 = S1 + k \times t^n,$$

同理, n 为奇数, k 为正数, k 以及 n 的设置是为了使物理机集群整体均衡以及物理机内部资源均衡的影响相当, 即使 $S1$ 与 $k \times t^n$ 的数值在同一数量级。应当理解, 若虚拟机的消耗类型为内存消耗型, 应该选择 CPU 相对内存消耗高的物理机来运行虚拟机, 以此来均衡物理机内部资源, 所选择的物理机的第二均衡度 t 值大于零。

本实施例中, 将物理机全部分配 8C/8G 和 4C/32G 情况进行计算获知, U_i/M_i 的范围为 (0.32, 2.56), 因此第二均衡度 t 的范围为 (-2.24, 2.24)。

若虚拟机的消耗类型为 CPU 消耗型, 均衡参数 S2:

$$S2 = S1 - \frac{1}{1000} \left(\frac{t}{3} \right)^3 = S1 - \frac{1}{1000} \left(\frac{\frac{U_i}{M_i} - \frac{U_{avg}}{M_{avg}}}{3} \right)^3$$

若虚拟机的消耗类型为内存消耗型, 均衡参数 S2 有:

$$S2 = S1 + \frac{1}{1000} \left(\frac{t}{3} \right)^3 = S1 + \frac{1}{1000} \left(\frac{\frac{U_i}{M_i} - \frac{U_{avg}}{M_{avg}}}{3} \right)^3$$

其中, 对第二均衡度 t 进行处理, 使 t/3 的范围为 (-1,1), 再 3 次方后, 使物理机内 CPU 和内存资源越不均衡时, 对均衡参数 S2 的影响越大, 再除以 1000, 使物理机内部资源的均衡对均衡参数 S2 的影响程度与物理机集群整体均衡的影响程度在同一等级。

从上述关系式可知, 若虚拟机的消耗类型是 CPU 消耗型, 一个物理机的第一均衡度大于另一物理机的第一均衡度, 且该一个物理机的第二均衡度小于另一物理机的第二均衡度, 该一个物理机的均衡参数大于另一物理机的均衡参数; 若虚拟机的消耗类型是内存消耗型, 一个物理机的第一均衡度和第二均衡度分别大于另一物理机的第一均衡度和第二均衡度, 该一个物理机的均衡参数大于另一物理机的均衡参数。

应当理解, 利用上述负载均衡的方法, 即考虑了物理机集群整体的负载均衡, 又考虑了单个物理机内部 CPU 和内存资源的负载均衡, 因此通过该方法使不同物理机的 CPU 使用率和内存使用率接近, 同时单个物理机内部的 CPU 使用率和内存使用率也接近, 进而即保证了物理机彼此之间的负载均衡以及单个物理机内部资源的均衡, 同时还提高了各个物理机内资源的利用率, 使物理机满载时, 运行的虚拟机数量增加了。请参看图 6 和图 7, 图中所示为使用上述负载均衡方法后, 物理机的 CPU 使用率和内存使用率的折线图和散点图, 图中所示 L1 为 CPU 使用率, L2 为内存使用率。

相较于上述方法, 进一步地, 为了衡量物理机集群的均衡情况, 采用综合均衡度以及接收的虚拟机的数量作为衡量标准, 综合均衡度越低, 表示物理机集群的负载越加均衡; 处理的虚拟机越多, 表示物理机内资源利用率越高。负载均衡方法还包括:

根据每个物理机当前的 CPU 使用率和内存使用率,计算出 CPU 平均使用率和内存平均使用率。

根据每个物理机的 CPU 使用率与 CPU 平均使用率的离散程度以及每个物理机的内存使用率与内存平均使用率的离散程度获取综合均衡度 Q, 综合均衡度 Q 用于表示物理机集群的负载均衡度。具体的, 通过对所有物理机的 CPU 使用率以及内存使用率分别进行均方误差计算后进行加权计算。

具体的, 若 U_i 表示第 i 个物理机的 CPU 使用率, U_{avg} 表示物理机集群的 CPU 平均使用率, 则物理机集群的 CPU 均衡度 $Q1$ 为对所有物理机的 CPU 使用率进行均方误差计算:

$$Q1 = \frac{1}{N} \sum_{i=1}^N ((U_i - U_{avg})^2)$$

若 M_i 表示第 i 个物理机的内存使用率, M_{avg} 表示物理机集群的内存平均使用率。物理机集群的内存均衡度 $Q2$ 为对所有物理机的内存使用率进行均方误差计算:

$$Q2 = \frac{1}{N} \sum_{i=1}^N ((M_i - M_{avg})^2)$$

则对所计算出来的 CPU 均衡度以及内存均衡度进行加权计算, 综合均衡度 Q 有:

$$Q = Q1 \times W2 + Q2 \times (1 - W2)$$

其中, $W2$ 表示考虑综合均衡度时, CPU 相对内存的均衡权重, 若 $W2$ 越大, 表示优选考虑 CPU 对综合均衡度的影响。这是因为 CPU 和内存对综合均衡度的影响可能是不同的。本实施例, 暂时设定 CPU 和内存对综合均衡度的影响相近, 忽略之间的差异, 取 $W2$ 为 0.5。

应当理解, 根据综合均衡度 Q 可获知该物理机集群的负载均衡情况, 若综合均衡度越小, 表示越均衡, 因此可以实时根据综合均衡度 Q 监控物理机集群的负载均衡的程度, 以便及时调节物理机集群的负载均衡或者利用综合均衡度 Q 来监控是否出现了异常的分配虚拟机的情况。另一方面, 还可以根据综合均衡度 Q 调整第一均衡度、第二均衡度的计算方式或者根据综合均衡度 Q 调整第一均衡度、第二均衡度与均衡参数之间的关系, 以便物理机集群的负载更加均衡。

还应当理解, 若在实际情况中, 为了扩大规模, 需要增加物理机, 使用该

负载均衡的方法后，增加物理机后，刚增加的物理机的 CPU 使用率和内存使用率为零，远远偏离各自均值，物理机集群的当前的综合均衡度 Q 将会上升，系统相对不均衡，但是根据负载均衡方法，将会优先分配虚拟机给新增的物理机，因此，物理机集群的综合均衡度 Q 将会下降直至平稳，因此新增的物理机可以快速的融入物理机集群中，且物理机集群相对快速的恢复至均衡状态。

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。

本发明还提供了一种计算机可读存储介质，计算机可读存储介质存储有一个或者一个以上程序，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，以执行以下操作：

若接收虚拟机的请求，则获取虚拟机的 CPU 核数和内存大小；

根据获取的虚拟机的 CPU 核数和内存大小识别虚拟机的消耗类型；

获取每个物理机当前的 CPU 使用率和内存使用率；

根据虚拟机的消耗类型以及获取的每个物理机的 CPU 使用率和内存使用率，按照预设第一规则计算出每个物理机的均衡参数；

将虚拟机分配给均衡参数最大的物理机。

优选地，所述一个或者一个以上程序被一个或者一个以上的处理器执行时，在执行根据获取的虚拟机的 CPU 核数和内存大小识别虚拟机的消耗类型时，具体执行以下操作：

计算获取的虚拟机的 CPU 核数与内存大小的比值；

判断比值是在预设第一范围还是预设第二范围或是预设第三范围。其中预设第二范围大于预设第一范围和预设第三范围，预设第一范围大于预设第三范围；

若比值在预设第一范围，识别虚拟机的消耗类型为常规类型；

若比值在预设第二范围，识别虚拟机的消耗类型为 CPU 消耗型；

若比值在预设第三范围，识别虚拟机的消耗类型为内存消耗型，其中所述预设第二范围大于所述预设第一范围和所述预设第三范围，所述预设第一范围大于所述预设第三范围。优选预设第一范围为：大于 0.25 且小于 1，预设第二

范围为：大于或等于 1，预设第三范围为小于或等于 0.25。

优选地，所述一个或者一个以上程序被一个或者一个以上的处理器执行时，在执行根据所述虚拟机的消耗类型以及获取的每个物理机的 CPU 使用率和内存使用率，按照预设第一规则计算出所述每个物理机的均衡参数时，具体执行以下操作：

将获取的每个物理机的 CPU 使用率和内存使用率按照预设第二规则计算出每个物理机的第一均衡度；

若虚拟机的消耗类型是常规类型，赋予每个物理机的均衡参数为对应物理机的第一均衡度；

若虚拟机的消耗类型是 CPU 消耗型或内存消耗型，根据获取的每个物理机的 CPU 使用率和内存使用率，计算出 CPU 平均使用率和内存平均使用率；

根据获取的每个物理机的 CPU 使用率和内存使用率以及 CPU 平均使用率和内存平均使用率计算出每个物理机的第二均衡度；

根据虚拟机的消耗类型、每个物理机的第一均衡度以及第二均衡度计算出每个物理机的均衡参数。

优选地，第二均衡度等于物理机的 CPU 使用率和内存使用率的比值与 CPU 平均使用率和内存平均使用率比值的差。关系如下所示：

$$t = \frac{U_i}{M_i} - \frac{U_{avg}}{M_{avg}}$$

当 $t > 0$ 是，表示此物理机的 CPU 使用率相对内存使用率较多即 CPU 消耗相对内存消耗较多， t 越大表示该物理机内 CPU 和内存资源越不均衡；当 $t < 0$ 时，表示此物理机的内存使用率相对 CPU 使用率较多，即内存消耗相对 CPU 消耗较多， t 的绝对值越大，表示该物理机内 CPU 和内存资源越不均衡。

若虚拟机的消耗类型为 CPU 消耗型，均衡参数 S2 有：

$$S2 = S1 - k \times t^n$$

其中， n 为奇数， k 为正数， k 以及 n 的设置是为了使物理机集群整体均衡以及物理机内部资源均衡的影响相当，即使 $S1$ 与 $k \times t^n$ 的数值在同一数量级。应当理解，若虚拟机的消耗类型为 CPU 消耗型，应该选择内存相对 CPU 消耗高的物理机来运行虚拟机，以此来均衡物理机的内部资源，所选择的物理机的第二均衡度 t 值小于零。

若虚拟机的消耗类型为内存消耗型，均衡参数 S2 有：

$$S2 = S1 + k \times t^n,$$

同理，n 为奇数，k 为正数，k 以及 n 的设置是为了使物理机集群整体均衡以及物理机内部资源均衡的影响相当，即使 S1 与 $k \times t^n$ 的数值在同一数量级。应当理解，若虚拟机的消耗类型为内存消耗型，应该选择 CPU 相对内存消耗高的物理机来运行虚拟机，以此来均衡物理机内部资源，所选择的物理机的第二均衡度 t 值大于零。

本实施例中，将物理机全部分配 8C/8G 和 4C/32G 情况进行计算获知， U_i/M_i 的范围为 (0.32, 2.56)，因此第二均衡度 t 的范围为 (-2.24, 2.24)。

若虚拟机的消耗类型为 CPU 消耗型，均衡参数 S2：

$$S2 = S1 - \frac{1}{1000} \left(\frac{t}{3}\right)^3 = S1 - \frac{1}{1000} \left(\frac{\frac{U_i}{M_i} - \frac{U_{avg}}{M_{avg}}}{3}\right)^3$$

若虚拟机的消耗类型为内存消耗型，均衡参数 S2 有：

$$S2 = S1 + \frac{1}{1000} \left(\frac{t}{3}\right)^3 = S1 + \frac{1}{1000} \left(\frac{\frac{U_i}{M_i} - \frac{U_{avg}}{M_{avg}}}{3}\right)^3$$

其中，对第二均衡度 t 进行处理，使 t/3 的范围为 (-1,1)，再 3 次方后，使物理机内 CPU 和内存资源越不均衡时，对均衡参数 S2 的影响越大，再除以 1000，使物理机内部资源的均衡对均衡参数 S2 的影响程度与物理机集群整体均衡的影响程度在同一等级。

优选地，若所述虚拟机的消耗类型是 CPU 消耗型，一个物理机的第一均衡度大于另一物理机的第一均衡度，且所述一个物理机的第二均衡度小于所述另一物理机的第二均衡度，所述一个物理机的均衡参数大于所述另一物理机的均衡参数；若所述虚拟机的消耗类型是内存消耗型，一个物理机的第一均衡度和第二均衡度分别大于另一物理机的第一均衡度和第二均衡度，所述一个物理机的均衡参数大于所述另一物理机的均衡参数。

优选地，所述一个或者一个以上程序被一个或者一个以上的处理器执行时，在执行将获取的每个物理机的 CPU 使用率和内存使用率按照预设第二规则计算

出每个物理机的第一均衡度时，具体执行以下操作：

计算每个物理机当前的 CPU 使用率的倒数和内存使用率的倒数；

对所有物理机的 CPU 使用率的倒数进行求和得到 CPU 总权重，对所有物理机的内存使用率的倒数进行求和得到内存总权重；

计算每个物理机的 CPU 使用率的倒数与 CPU 总权重的比值得到 CPU 占比 c 以及每个物理机的内存使用率的倒数与内存总权重的比值得到内存占比 m ；

根据所计算出的每个物理机的 CPU 占比 c 以及内存占比 m 和预设的 CPU 相对内存的选择权重 $w1$ 计算出第一均衡度 $S1$ 。

第一均衡度 $S1$ 与 CPU 占比 c 、内存占比 m 、CPU 相对内存的选择权重 $w1$ 的关系如下：

$$S1=c \times w1+m \times (1-w1)$$

其中预设的 CPU 相对内存的选择权重 $w1$ 表示优先考虑 CPU 影响因素或者内存影响因素， $w1$ 越大表示更优先考虑 CPU 因素对物理机集群整体负载均衡的影响， $w1$ 越小表示更优先考虑内存因素对物理机集群整体负载均衡的影响。

优选地，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，以执行以下操作：

根据每个物理机当前的 CPU 使用率和内存使用率，计算出 CPU 平均使用率和内存平均使用率；

根据每个物理机的 CPU 使用率与 CPU 平均使用率的离散程度以及每个物理机的内存使用率与内存平均使用率的离散程度获取综合均衡度 Q ，综合均衡度 Q 用于表示物理机集群的负载均衡。具体的，通过对所有物理机的 CPU 使用率以及内存使用率分别进行均方误差计算后进行加权计算。

请参看图 8，本发明实施例提供的一种负载均衡装置，该装置 100 用于分配物理机给虚拟机放置运行，该装置 100 分别与物理机以及虚拟机通讯，如图所示，负载均衡装置包括接收模块 81、获取模块 82、识别模块 83、第一计算模块 84 和分配模块 85。

其中接收模块 81 用于接收虚拟机的请求。

获取模块 82 用于获取虚拟机的 CPU 核数和内存大小以及还用于获取每个物理机当前的 CPU 使用率和内存使用率。CPU（central processing unit）是中央处理器的缩写，CPU 核数用于表示 CPU 核心的数量。

识别模块 83 用于根据获取的虚拟机的 CPU 核数和内存大小识别虚拟机的消耗类型。

第一计算模块 84 根据所述虚拟机的消耗类型以及获取的每个物理机的 CPU 使用率和内存使用率，按照预设第一规则计算出所述每个物理机的均衡参数。

分配模块 85 用于将虚拟机分配给均衡参数最大的物理机。

具体的，识别模块 83 识别虚拟机的消耗类型时，本实施例中，虚拟机的消耗类型包括常规类型、CPU 消耗型以及内存消耗型，不同消耗类型的虚拟机的对资源的占用不同，CPU 消耗型所占用的 CPU 资源相比内存消耗型和常规类型占用的 CPU 资源多；内存消耗型所占用的内存资源相比 CPU 消耗型和常规类型占用的内存资源多。识别模块 83 还包括第一运算单元 831、判断单元 832 以及处理单元 833。

其中第一运算单元 831 用于计算获取的 CPU 核数与内存大小的比值。

判断单元 832 用于判断比值是在预设第一范围还是预设第二范围或是预设第三范围。其中预设第二范围大于预设第一范围和预设第三范围，预设第一范围大于预设第三范围。本实施例中优选预设第一范围为：大于 0.25 且小于 1，预设第二范围为：大于或等于 1，预设第三范围为小于或等于 0.25。

处理单元 833 用于如判断单元 832 判断出比值在预设第一范围，识别虚拟机的消耗类型为常规类型；判断出比值在预设第二范围，识别虚拟机的消耗类型为 CPU 消耗型以及判断出比值在预设第三范围，识别虚拟机的消耗类型为内存消耗型。

需要说明的是，第一计算模块 84 计算出每个物理机的均衡参数，是综合考虑物理机集群整体的负载均衡以及物理机内部 CPU 和内存资源的负载均衡时，物理机的选择性。本实施例中，用第一均衡度 S1 表示考虑物理机集群整体的负载均衡时，物理机的选择性；用第二均衡度 t 表示考虑物理机内部 CPU 和内存资源的负载均衡时，物理机的选择性。第一计算模块 84 包括第一计算单元 841、设置单元 842、第二计算单元 843 以及第三计算单元 844 和第四计算单元 845。

第一计算单元 841 用于将获取的每个物理机的 CPU 使用率和内存使用率按照预设第二规则计算出每个物理机的第一均衡度。

具体的，第一均衡度 S1 用于表示考虑物理机集群整体的负载均衡时，物理机的选择性。应当理解，若只考虑整个集群整体的负载均衡，而不考虑各个物

理机内部资源均衡,将选择当前第一均衡度 S1 最大的物理机来放置虚拟机。应当理解若只考虑内存,而不考虑 CPU 的影响因素时,为了物理机集群的负载均衡,应当将虚拟机分配给当前内存使用率最低的物理机,同理,只考虑 CPU,不考虑内存的影响时,应当将虚拟机的请求给当前 CPU 使用率最低的物理机,本实施例同时考虑 CPU 以及内存的影响,因此第一均衡度 S1 与 CPU 使用率 U_i 和内存使用率 M_i 均相关,第一均衡度 S1 与 CPU 使用率以及内存使用率成函数关系。

请参看图 9,第一计算单元 841 包括第二运算单元 8411、求和单元 8412、第三运算单元 8413 以及第四运算单元 8414。

其中,第二运算单元 8411 用于计算每个物理机当前的 CPU 使用率的倒数和内存使用率的倒数。

求和单元 8412 用于对所有物理机的 CPU 使用率的倒数进行求和得到 CPU 总权重,对所有物理机的内存使用率的倒数进行求和得到内存总权重。

第三运算单元 8413 用于计算每个物理机的 CPU 使用率的倒数与 CPU 总权重的比值得到 CPU 占比 c 以及每个物理机的内存使用率的倒数与内存总权重的比值得到内存占比 m 。

第四运算单元 8414 用于根据所计算出的每个物理机的 CPU 占比 c 以及内存占比 m 和预设的 CPU 相对内存的选择权重 w_1 计算出第一均衡度 S1。

第一均衡度 S1 与 CPU 占比 c 、内存占比 m 、CPU 相对内存的选择权重 w_1 的关系如下:

$$S1=c \times w_1+m \times (1-w_1)$$

其中预设的 CPU 相对内存的选择权重 w_1 表示优先考虑 CPU 影响因素或者内存影响因素, w_1 越大表示更优先考虑 CPU 因素对物理机集群整体负载均衡的影响, w_1 越小表示更优先考虑内存因素对物理机集群整体负载均衡的影响。本实施例暂时忽略 CPU 因素或内存因素的影响区别,将 w_1 设置为 0.5,其它可行的实施例中, w_1 可以根据 CPU 因素或者内存因素的影响程度设置为其他数值。

应当理解,由于本实施例中 w_1 等于 0.5,因此若一个物理机的 CPU 使用率和内存使用率分别大于另一物理机的 CPU 使用率和内存使用率,该一个物理机的第一均衡度 S1 小于另一物理机的第一均衡度 S1。

设置单元 842 用于若虚拟机的消耗类型是常规类型,赋予每个物理机的均

衡参数为对应物理机的第一均衡度。

第二计算单元 843 用于若虚拟机的消耗类型是 CPU 消耗型或内存消耗型，根据获取的每个物理机的 CPU 使用率和内存使用率，计算出 CPU 平均使用率和内存平均使用率。

具体的，本实施例中用 U_{avg} 表示物理机集群的 CPU 平均使用率， M_{avg} 表示物理机集群的内存平均使用率。

第三计算单元 844 用于根据获取的每个物理机的 CPU 使用率和内存使用率以及 CPU 平均使用率和内存平均使用率计算出每个物理机的第二均衡度。

用第二均衡度 t 表示考虑物理机内部 CPU 和内存资源的负载均衡时，物理机的选择性。第二均衡度 t 等于物理机的 CPU 使用率 U_i 和内存使用率 M_i 的比值与 CPU 平均使用率 U_{avg} 和内存平均使用率 M_{avg} 比值的差，关系如下所示：

$$t = \frac{U_i}{M_i} - \frac{U_{avg}}{M_{avg}}$$

当 $t > 0$ 是，表示此物理机的 CPU 使用率相对内存使用率较多即 CPU 消耗相对内存消耗较多， t 越大表示该物理机内 CPU 和内存资源越不均衡；当 $t < 0$ 时，表示此物理机的内存使用率相对 CPU 使用率较多，即内存消耗相对 CPU 消耗较多， t 的绝对值越大，表示该物理机内 CPU 和内存资源越不均衡。

第四计算单元 845 用于根据虚拟机的消耗类型、每个物理机的第一均衡度以及第二均衡度计算出每个物理机的均衡参数。

由于第一均衡度 $S1$ 用于表示考虑物理机集群整体的负载均衡时，物理机的选择性；而第二均衡度 t 表示考虑物理机内部 CPU 和内存资源的负载均衡时，物理机的选择性。而虚拟机的消耗类型为 CPU 消耗型或者内存消耗型时，虚拟机放置在物理机上后均会对物理机内部资源的均衡造成影响。因此均衡参数与虚拟机的消耗类型、第一均衡度以及第二均衡度相关。

若虚拟机的消耗类型为 CPU 消耗型，均衡参数 $S2$ 有：

$S2 = S1 - k \times t^n$ 其中， n 为奇数， k 为正数， k 以及 n 的设置是为了使物理机集群整体均衡以及物理机内部资源均衡的影响相当，即使 $S1$ 与 $k \times t^n$ 的数值在同一数量级。应当理解，若虚拟机的消耗类型为 CPU 消耗型，应该选择内存相对 CPU 消耗高的物理机来运行虚拟机，以此来均衡物理机的内部资源，所选择的物理机的第二均衡度 t 值小于零。

若虚拟机的消耗类型为内存消耗型，均衡参数 S2 有：

$S2 = S1 + k \times t^n$ ，同理，n 为奇数，k 为正数，k 以及 n 的设置是为了使物理机集群整体均衡以及物理机内部资源均衡的影响相当，即使 S1 与 $k \times t^n$ 的数值在同一数量级。应当理解，若虚拟机的消耗类型为内存消耗型，应该选择 CPU 相对内存消耗高的物理机来运行虚拟机，以此来均衡物理机内部资源，所选择的物理机的第二均衡度 t 值大于零。

本实施例中，将物理机全部分配 8C/8G 和 4C/32G 情况进行计算获知， U_i/M_i 的范围为 (0.32, 2.56)，因此 t 的范围为 (-2.24, 2.24)。

若虚拟机的消耗类型为 CPU 消耗型，均衡参数 S2：

$$S2 = S1 - \frac{1}{1000} \left(\frac{t}{3} \right)^3 = S1 - \frac{1}{1000} \left(\frac{\frac{U_i}{M_i} - \frac{U_{avg}}{M_{avg}}}{3} \right)^3$$

若虚拟机的消耗类型为内存消耗型，均衡参数 S2 有：

$$S2 = S1 + \frac{1}{1000} \left(\frac{t}{3} \right)^3 = S1 + \frac{1}{1000} \left(\frac{\frac{U_i}{M_i} - \frac{U_{avg}}{M_{avg}}}{3} \right)^3$$

其中，对第二均衡度 t 进行处理，使 t/3 的范围为 (-1,1)，再 3 次方后，使物理机内 CPU 和内存资源越不均衡时，对均衡参数 S2 的影响越大，再除以 1000，使物理机内部资源的均衡对均衡参数 S2 的影响程度与物理机集群整体均衡的影响程度在同一等级。

从上述关系式可知，若虚拟机的消耗类型是 CPU 消耗型，一个物理机的第一均衡度大于另一物理机的第一均衡度，且该一个物理机的第二均衡度小于另一物理机的第二均衡度，该一个物理机的均衡参数大于另一物理机的均衡参数；若虚拟机的消耗类型是内存消耗型，一个物理机的第一均衡度和第二均衡度分别大于另一物理机的第一均衡度和第二均衡度，该一个物理机的均衡参数大于另一物理机的均衡参数。

还需要说明的是，为了衡量物理机集群的均衡情况，采用综合均衡度以及接收的虚拟机的数量作为衡量标准，综合均衡度越低，表示物理机集群的负载越加均衡；处理的虚拟机越多，表示物理机内资源利用率越高。该负载均衡装置 100 还包括第二计算模块 86 以及第三计算模块 87。

其中第二计算模块 86 用于根据每个物理机当前的 CPU 使用率和内存使用率, 计算出 CPU 平均使用率和内存平均使用率。

第三计算模块 87 用于根据每个物理机的 CPU 使用率与 CPU 平均使用率的离散程度以及每个物理机的内存使用率与内存平均使用率的离散程度获取综合均衡度 Q, 综合均衡度 Q 用于表示物理机集群的负载均衡。具体的, 通过对所有物理机的 CPU 使用率以及内存使用率分别进行均方误差计算后进行加权计算。

具体的, 若 U_i 表示第 i 个物理机的 CPU 使用率, U_{avg} 表示物理机集群的 CPU 平均使用率, 则物理机集群的 CPU 均衡度 $Q1$ 为对所有物理机的 CPU 使用率进行均方误差计算:

$$Q1 = \frac{1}{N} \sum_{i=1}^N ((U_i - U_{avg})^2)$$

若 M_i 表示第 i 个物理机的内存使用率, M_{avg} 表示物理机集群的内存平均使用率。物理机集群的内存均衡度 $Q2$ 为对所有物理机的内存使用率进行均方误差计算:

$$Q2 = \frac{1}{N} \sum_{i=1}^N ((M_i - M_{avg})^2)$$

则对所计算出来的 CPU 均衡度以及内存均衡度进行加权计算, 综合均衡度 Q 有:

$$Q = Q1 \times W2 + Q2 \times (1 - W2)$$

其中, $W2$ 表示考虑综合均衡度时, CPU 相对内存的均衡权重, 若 $W2$ 越大, 表示优选考虑 CPU 对综合均衡度的影响。这是因为 CPU 和内存对综合均衡度的影响可能是不同的。应当理解, 根据综合均衡度 Q 可获知该物理机集群的负载均衡情况, 若综合均衡度越小, 表示越均衡。

上述负载均衡的方法和装置, 即考虑了物理机集群整体的负载均衡, 又考虑了单个物理机内部 CPU 和内存资源的负载均衡, 因此通过该方法使不同物理机的 CPU 使用率和内存使用率接近, 同时单个物理机内部的 CPU 使用率和内存使用率也接近, 进而即保证了负载均衡, 同时还提高了各个物理机内资源的利用率, 使物理机满载时, 运行的虚拟机数量增加了, 且若扩大规模增加物理机时, 通过该负载均衡的方法可以使新增的物理机快速融入物理机集群中, 使其 CPU 使用率和内存使用率与均值快速接近, 使物理机集群恢复至均衡状态。

需要说明的是，本发明实施例方法中的步骤可以根据实际需要进行顺序调整、合并和删减。

本发明实施例装置中的单元可以根据实际需要进行合并、划分和删减。

另一方面，在硬件实现上，以上接收模块 81、获取模块 82 以及识别模块 83 等可以以硬件形式内嵌于或独立于一种负载均衡系统中，也可以以软件形式存储于一种负载均衡系统的存储器中，以便于处理器执行以上各个单元对应的操作。该处理器可以为中央处理单元（CPU）、微处理器、单片机等。

如图 10 所示，图 10 是本发明实施例提供的一种负载均衡系统的硬件结构图，在本实施例中，该一种负载均衡系统包括一个或者一个以上处理器 1001、以及存储器 1002。上述处理器 1001 以及存储器 1002 通过总线连接。

存储器 1002，用于存储用于实现负载均衡的一个或者一个以上程序。具体实现中，本发明实施例的存储器 1002 可以是系统存储器，比如，挥发性的（诸如 RAM），非易失性的（诸如 ROM，闪存等），或者两者的结合。具体实现中，本发明实施例的存储器 1002 还可以是系统之外的外部存储器，比如，磁盘、光盘、磁带等。

处理器 1001，用于执行存储器 1002 中存储的程序，以执行以下操作：

若接收虚拟机的请求，则获取虚拟机的 CPU 核数和内存大小；

根据获取的虚拟机的 CPU 核数和内存大小识别虚拟机的消耗类型；

获取每个物理机当前的 CPU 使用率和内存使用率；

根据虚拟机的消耗类型以及获取的每个物理机的 CPU 使用率和内存使用率，按照预设第一规则计算出每个物理机的均衡参数；

将虚拟机分配给均衡参数最大的物理机。

优选地，处理器 1001 执行存储器 1002 中存储的程序时，以执行根据获取的虚拟机的 CPU 核数和内存大小识别虚拟机的消耗类型时，具体执行以下操作：

计算获取的虚拟机的 CPU 核数与内存大小的比值；

判断比值是在预设第一范围还是预设第二范围或是预设第三范围。其中预设第二范围大于预设第一范围和预设第三范围，预设第一范围大于预设第三范围；

若比值在预设第一范围，识别虚拟机的消耗类型为常规类型；

若比值在预设第二范围，识别虚拟机的消耗类型为 CPU 消耗型；

若比值在预设第三范围，识别虚拟机的消耗类型为内存消耗型，其中所述预设第二范围大于所述预设第一范围和所述预设第三范围，所述预设第一范围大于所述预设第三范围。优选预设第一范围为：大于 0.25 且小于 1，预设第二范围为：大于或等于 1，预设第三范围为小于或等于 0.25。

优选地，处理器 1001 执行存储器 1002 中存储的程序时，以执行根据所述虚拟机的消耗类型以及获取的每个物理机的 CPU 使用率和内存使用率，按照预设第一规则计算出所述每个物理机的均衡参数时，具体执行以下操作：

将获取的每个物理机的 CPU 使用率和内存使用率按照预设第二规则计算出每个物理机的第一均衡度；

若虚拟机的消耗类型是常规类型，赋予每个物理机的均衡参数为对应物理机的第一均衡度；

若虚拟机的消耗类型是 CPU 消耗型或内存消耗型，根据获取的每个物理机的 CPU 使用率和内存使用率，计算出 CPU 平均使用率和内存平均使用率；

根据获取的每个物理机的 CPU 使用率和内存使用率以及 CPU 平均使用率和内存平均使用率计算出每个物理机的第二均衡度；

根据虚拟机的消耗类型、每个物理机的第一均衡度以及第二均衡度计算出每个物理机的均衡参数。

优选地，第二均衡度等于物理机的 CPU 使用率和内存使用率的比值与 CPU 平均使用率和内存平均使用率比值的差。关系如下所示：

$$t = \frac{U_i}{M_i} - \frac{U_{avg}}{M_{avg}}$$

当 $t > 0$ 是，表示此物理机的 CPU 使用率相对内存使用率较多即 CPU 消耗相对内存消耗较多， t 越大表示该物理机内 CPU 和内存资源越不均衡；当 $t < 0$ 时，表示此物理机的内存使用率相对 CPU 使用率较多，即内存消耗相对 CPU 消耗较多， t 的绝对值越大，表示该物理机内 CPU 和内存资源越不均衡。

若虚拟机的消耗类型为 CPU 消耗型，均衡参数 $S2$ 有：

$S2 = S1 - k \times t^n$ 其中， n 为奇数， k 为正数， k 以及 n 的设置是为了使物理机集群整体均衡以及物理机内部资源均衡的影响相当，即使 $S1$ 与 $k \times t^n$ 的数值在同一数量级。应当理解，若虚拟机的消耗类型为 CPU 消耗型，应该选择内存相对 CPU 消耗高的物理机来运行虚拟机，以此来均衡物理机的内部资源，所选

择的物理机的第二均衡度 t 值小于零。

若虚拟机的消耗类型为内存消耗型，均衡参数 $S2$ 有：

$S2 = S1 + k \times t^n$ ，同理， n 为奇数， k 为正数， k 以及 n 的设置是为了使物理机集群整体均衡以及物理机内部资源均衡的影响相当，即使 $S1$ 与 $k \times t^n$ 的数值在同一数量级。应当理解，若虚拟机的消耗类型为内存消耗型，应该选择 CPU 相对内存消耗高的物理机来运行虚拟机，以此来均衡物理机内部资源，所选择的物理机的第二均衡度 t 值大于零。

本实施例中，将物理机全部分配 8C/8G 和 4C/32G 情况进行计算获知， U_i/M_i 的范围为 (0.32, 2.56)，因此第二均衡度 t 的范围为 (-2.24, 2.24)。

若虚拟机的消耗类型为 CPU 消耗型，均衡参数 $S2$ ：

$$S2 = S1 - \frac{1}{1000} \left(\frac{t}{3} \right)^3 = S1 - \frac{1}{1000} \left(\frac{\frac{U_i}{M_i} - \frac{U_{avg}}{M_{avg}}}{3} \right)^3$$

若虚拟机的消耗类型为内存消耗型，均衡参数 $S2$ 有：

$$S2 = S1 + \frac{1}{1000} \left(\frac{t}{3} \right)^3 = S1 + \frac{1}{1000} \left(\frac{\frac{U_i}{M_i} - \frac{U_{avg}}{M_{avg}}}{3} \right)^3$$

其中，对第二均衡度 t 进行处理，使 $t/3$ 的范围为 (-1,1)，再 3 次方后，使物理机内 CPU 和内存资源越不均衡时，对均衡参数 $S2$ 的影响越大，再除以 1000，使物理机内部资源的均衡对均衡参数 $S2$ 的影响程度与物理机集群整体均衡的影响程度在同一等级。

优选地，若所述虚拟机的消耗类型是 CPU 消耗型，一个物理机的第一均衡度大于另一物理机的第一均衡度，且所述一个物理机的第二均衡度小于所述另一物理机的第二均衡度，所述一个物理机的均衡参数大于所述另一物理机的均衡参数；若所述虚拟机的消耗类型是内存消耗型，一个物理机的第一均衡度和第二均衡度分别大于另一物理机的第一均衡度和第二均衡度，所述一个物理机的均衡参数大于所述另一物理机的均衡参数。

优选地，处理器 1001 执行存储器 1002 中存储的程序时，以执行将获取的每个物理机的 CPU 使用率和内存使用率按照预设第二规则计算出每个物理机的第一均衡度时，具体执行以下操作：

计算每个物理机当前的 CPU 使用率的倒数和内存使用率的倒数；

对所有物理机的 CPU 使用率的倒数进行求和得到 CPU 总权重,对所有物理机的内存使用率的倒数进行求和得到内存总权重；

计算每个物理机的 CPU 使用率的倒数与 CPU 总权重的比值得到 CPU 占比 c 以及每个物理机的内存使用率的倒数与内存总权重的比值得到内存占比 m ；

根据所计算出的每个物理机的 CPU 占比 c 以及内存占比 m 和预设的 CPU 相对内存的选择权重 $w1$ 计算出第一均衡度 $S1$ 。

第一均衡度 $S1$ 与 CPU 占比 c 、内存占比 m 、CPU 相对内存的选择权重 $w1$ 的关系如下：

$$S1=c \times w1+m \times (1-w1)$$

其中预设的 CPU 相对内存的选择权重 $w1$ 表示优先考虑 CPU 影响因素或者内存影响因素, $w1$ 越大表示更优先考虑 CPU 因素对物理机集群整体负载均衡的影响, $w1$ 越小表示更优先考虑内存因素对物理机集群整体负载均衡的影响。

优选地, 处理器 1001 执行存储器 1002 中存储的程序时还执行以下操作：

根据每个物理机当前的 CPU 使用率和内存使用率, 计算出 CPU 平均使用率和内存平均使用率；

根据每个物理机的 CPU 使用率与 CPU 平均使用率的离散程度以及每个物理机的内存使用率与内存平均使用率的离散程度获取综合均衡度 Q , 综合均衡度 Q 用于表示物理机集群的负载均衡。具体的, 通过对所有物理机的 CPU 使用率以及内存使用率分别进行均方误差计算后进行加权计算。

本发明实施例中所述处理器可以是中央处理单元(Central Processing Unit, CPU), 该处理器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。

所属领域的技术人员可以清楚地了解到, 为了描述的方便和简洁, 上述描述的工具和单元的具体工作过程, 可以参考前述方法实施例中的对应过程, 在此不再赘述。

在本申请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，所述单元的划分，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，例如多个单元或组建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。另外，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、装置或单元的间接耦合或通信连接，也可以是的，机械的或其它的形式连接。

所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本发明实施例方案的目的。

另外，在本发明各个实施例中的各功能单元可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以是两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现，也可以采用软件功能单元的形式实现。

所述集成的单元如果以软件功能单元的形式实现并作为独立的产品销售或使用，可以存储在一个计算机可读取存储介质中。基于这样的理解，本发明的技术方案本质上或者说对现有技术做出贡献的部分，或者该技术方案的全部或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一个存储介质中，包括若干指令用以使得一台计算机设备（可以是个人计算机，服务器，或者网络设备等）执行本发明各个实施例所述方法的全部或部分步骤。而前述的存储介质包括：U 盘、移动硬盘、只读存储器（ROM, Read-Only Memory）、随机存取存储器（RAM, Random Access Memory）、磁碟或者光盘等各种可以存储程序代码的介质。

以上所述，仅为本发明的具体实施方式，但本发明的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本发明揭露的技术范围内，可轻易想到各种等效的修改或替换，这些修改或替换都应涵盖在本发明的保护范围之内。因此，本发明的保护范围应以权利要求的保护范围为准。

权 利 要 求 书

1.一种负载均衡的方法，其特征在于，包括：

若接收虚拟机的请求，则获取所述虚拟机的 CPU 核数和内存大小；

根据获取的虚拟机的 CPU 核数和内存大小识别所述虚拟机的消耗类型；

获取每个物理机当前的 CPU 使用率和内存使用率；

根据所述虚拟机的消耗类型以及获取的每个物理机的 CPU 使用率和内存使用率，按照预设第一规则计算出所述每个物理机的均衡参数；

将所述虚拟机分配给所述均衡参数最大的物理机。

2.根据权利要求 1 所述的方法，其特征在于，所述根据获取的 CPU 核数和内存大小识别所述虚拟机的消耗类型，包括：

计算获取的所述 CPU 核数与所述内存大小的比值；

若所述比值在预设第一范围，识别所述虚拟机的消耗类型为常规类型；

若所述比值在预设第二范围，识别所述虚拟机的消耗类型为 CPU 消耗型；

若所述比值在预设第三范围，识别所述虚拟机的消耗类型为内存消耗型，其中所述预设第二范围大于所述预设第一范围和所述预设第三范围，所述预设第一范围大于所述预设第三范围。

3.根据权利要求 1 所述的方法，其特征在于，所述虚拟机的消耗类型包括常规类型、CPU 消耗型以及内存消耗型，所述根据所述虚拟机的消耗类型以及获取的每个物理机的 CPU 使用率和内存使用率，按照预设第一规则计算出所述每个物理机的均衡参数，包括：

将获取的每个物理机的 CPU 使用率和内存使用率按照预设第二规则计算出每个物理机的第一均衡度；

若所述虚拟机的消耗类型是常规类型，赋予所述每个物理机的均衡参数为对应物理机的第一均衡度；

若所述虚拟机的消耗类型是 CPU 消耗型或内存消耗型，根据获取的每个物理机的 CPU 使用率和内存使用率，计算出 CPU 平均使用率和内存平均使用率；

根据获取的每个物理机的 CPU 使用率和内存使用率以及所述 CPU 平均使用率和所述内存平均使用率计算出每个物理机的第二均衡度；

根据所述虚拟机的消耗类型、所述每个物理机的第一均衡度以及第二均衡度计算出所述每个物理机的均衡参数。

4.根据权利要求 3 所述的方法,其特征在于,所述将获取的每个物理机的 CPU 使用率和内存使用率按照预设第二规则计算出每个物理机的第一均衡度,包括:

计算每个物理机当前的 CPU 使用率的倒数和内存使用率的倒数;

对所有物理机的 CPU 使用率的倒数进行求和得到 CPU 总权重,对所有物理机的内存使用率的倒数进行求和得到内存总权重;

计算每个物理机的所述 CPU 使用率的倒数与所述 CPU 总权重的比值得到 CPU 占比以及每个物理机的所述内存使用率的倒数与所述内存总权重的比值得到内存占比;

根据所计算出的每个物理机的所述 CPU 占比以及所述内存占比和预设的 CPU 相对内存的选择权重计算出第一均衡度。

5.根据权利要求 1 所述的方法,其特征在于,所述方法还包括:

根据每个物理机当前的所述 CPU 使用率和所述内存使用率,计算出 CPU 平均使用率和内存平均使用率;

根据每个物理机的所述 CPU 使用率与所述 CPU 平均使用率的离散程度以及每个物理机的所述内存使用率与所述内存平均使用率的离散程度获取综合均衡度,所述综合均衡度用于表示物理机集群的负载均衡度,以便根据综合均衡度来监控物理机集群的负载均衡的情况。

6.一种负载均衡装置,其特征在于,包括:

获取模块,用于若接收虚拟机的请求,获取所述虚拟机的 CPU 核数和内存大小以及还用于获取每个物理机当前的 CPU 使用率和内存使用率;

识别模块,用于根据获取的虚拟机的 CPU 核数和内存大小识别所述虚拟机的消耗类型;

第一计算模块,用于根据所述虚拟机的消耗类型以及获取的每个物理机的 CPU 使用率和内存使用率,按照预设第一规则计算出所述每个物理机的均衡参数;

分配模块,用于将所述虚拟机分配给所述均衡参数最大的物理机。

7.根据权利要求 6 所述的装置,其特征在于,所述识别模块包括:

第一运算单元,用于计算获取的所述 CPU 核数与所述内存大小的比值;

处理单元,用于若所述比值在预设第一范围,识别所述虚拟机的消耗类型为常规类型;以及还用于若所述比值在预设第二范围,识别所述虚拟机的消耗

类型为 CPU 消耗型；以及还用于若所述比值在预设第三范围，识别所述虚拟机的消耗类型为内存消耗型，其中所述预设第二范围大于所述预设第一范围和所述预设第三范围，所述预设第一范围大于所述预设第三范围。

8.根据权利要求 6 所述的装置，其特征在于，所述虚拟机的消耗类型包括常规类型、CPU 消耗型以及内存消耗型，所述第一计算模块包括：

第一计算单元，用于将获取的每个物理机的 CPU 使用率和内存使用率按照预设第二规则计算出每个物理机的第一均衡度；

设置单元，用于若所述虚拟机的消耗类型是常规类型，赋予所述每个物理机的均衡参数为对应物理机的第一均衡度；

第二计算单元，用于若所述虚拟机的消耗类型是 CPU 消耗型或内存消耗型，根据获取的每个物理机的 CPU 使用率和内存使用率，计算出 CPU 平均使用率和内存平均使用率；

第三计算单元，用于根据获取的每个物理机的 CPU 使用率和内存使用率以及所述 CPU 平均使用率和所述内存平均使用率计算出每个物理机的第二均衡度；

第四计算单元，用于根据所述虚拟机的消耗类型、所述每个物理机的第一均衡度以及第二均衡度计算出所述每个物理机的均衡参数。

9.根据权利要求 8 所述的装置，其特征在于，所述第一计算单元包括：

第二运算单元，用于计算每个物理机当前的 CPU 使用率的倒数和内存使用率的倒数；

求和单元，用于对所有物理机的 CPU 使用率的倒数进行求和得到 CPU 总权重，对所有物理机的内存使用率的倒数进行求和得到内存总权重；

第三运算单元，用于计算每个物理机的所述 CPU 使用率的倒数与所述 CPU 总权重的比值得到 CPU 占比以及每个物理机的所述内存使用率的倒数与所述内存总权重的比值得到内存占比；

第四运算单元，用于根据所计算出的每个物理机的 CPU 占比以及内存占比和预设的 CPU 相对内存的选择权重计算出第一均衡度。

10.根据权利要求 6 所述的装置，其特征在于，所述装置还包括：

第二计算模块，用于根据每个物理机当前的 CPU 使用率和内存使用率，计算出 CPU 平均使用率和内存平均使用率；

第三计算模块，用于根据每个物理机的 CPU 使用率与 CPU 平均使用率的离

散程度以及每个物理机的内存使用率与内存平均使用率的离散程度获取综合均衡度，所述综合均衡度用于表示物理机集群的负载均衡度，以便根据综合均衡度来监控物理机集群的负载均衡的情况。

11.一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，以执行以下操作：

若接收虚拟机的请求，则获取所述虚拟机的 CPU 核数和内存大小；

根据获取的虚拟机的 CPU 核数和内存大小识别所述虚拟机的消耗类型；

获取每个物理机当前的 CPU 使用率和内存使用率；

根据所述虚拟机的消耗类型以及获取的每个物理机的 CPU 使用率和内存使用率，按照预设第一规则计算出所述每个物理机的均衡参数；

将所述虚拟机分配给所述均衡参数最大的物理机。

12.根据权利要求 11 所述的计算机可读存储介质，其特征在于，所述一个或者一个以上程序被一个或者一个以上的处理器执行时，在执行所述根据获取的 CPU 核数和内存大小识别所述虚拟机的消耗类型时，具体执行以下操作：

计算获取的所述 CPU 核数与所述内存大小的比值；

若所述比值在预设第一范围，识别所述虚拟机的消耗类型为常规类型；

若所述比值在预设第二范围，识别所述虚拟机的消耗类型为 CPU 消耗型；

若所述比值在预设第三范围，识别所述虚拟机的消耗类型为内存消耗型，其中所述预设第二范围大于所述预设第一范围和所述预设第三范围，所述预设第一范围大于所述预设第三范围。

13.根据权利要求 11 所述的计算机可读存储介质，其特征在于，所述虚拟机的消耗类型包括常规类型、CPU 消耗型以及内存消耗型，所述一个或者一个以上程序被一个或者一个以上的处理器执行时，在执行所述根据所述虚拟机的消耗类型以及获取的每个物理机的 CPU 使用率和内存使用率，按照预设第一规则计算出所述每个物理机的均衡参数时，具体执行以下操作：

将获取的每个物理机的 CPU 使用率和内存使用率按照预设第二规则计算出每个物理机的第一均衡度；

若所述虚拟机的消耗类型是常规类型，赋予所述每个物理机的均衡参数为对应物理机的第一均衡度；

若所述虚拟机的消耗类型是 CPU 消耗型或内存消耗型,根据获取的每个物理机的 CPU 使用率和内存使用率,计算出 CPU 平均使用率和内存平均使用率;

根据获取的每个物理机的 CPU 使用率和内存使用率以及所述 CPU 平均使用率和所述内存平均使用率计算出每个物理机的第二均衡度;

根据所述虚拟机的消耗类型、所述每个物理机的第一均衡度以及第二均衡度计算出所述每个物理机的均衡参数。

14.根据权利要求 13 所述的计算机可读存储介质,其特征在于,所述一个或者一个以上程序被一个或者一个以上的处理器执行时,在执行将获取的每个物理机的 CPU 使用率和内存使用率按照预设第二规则计算出每个物理机的第一均衡度时,具体执行以下操作:

计算每个物理机当前的 CPU 使用率的倒数和内存使用率的倒数;

对所有物理机的 CPU 使用率的倒数进行求和得到 CPU 总权重,对所有物理机的内存使用率的倒数进行求和得到内存总权重;

计算每个物理机的所述 CPU 使用率的倒数与所述 CPU 总权重的比值得到 CPU 占比以及每个物理机的所述内存使用率的倒数与所述内存总权重的比值得到内存占比;

根据所计算出的每个物理机的所述 CPU 占比以及所述内存占比和预设的 CPU 相对内存的选择权重计算出第一均衡度。

15.根据权利要求 11 所述的计算机可读存储介质,其特征在于,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,以执行以下操作:

根据每个物理机当前的所述 CPU 使用率和所述内存使用率,计算出 CPU 平均使用率和内存平均使用率;

根据每个物理机的所述 CPU 使用率与所述 CPU 平均使用率的离散程度以及每个物理机的所述内存使用率与所述内存平均使用率的离散程度获取综合均衡度,所述综合均衡度用于表示物理机集群的负载均衡度,以便根据综合均衡度来监控物理机集群的负载均衡的情况。

16.一种负载均衡系统,其特征在于,包括:所述系统包括一个或者一个以上的处理器,以及与所述一个或者一个以上处理器连接的一个或者一个以上存储器,

所述存储器,用于存储用于实现负载均衡的一个或者一个以上程序;

所述处理器,用于执行所述存储器中存储的程序,以执行以下操作:

若接收虚拟机的请求，则获取所述虚拟机的 CPU 核数和内存大小；

根据获取的虚拟机的 CPU 核数和内存大小识别所述虚拟机的消耗类型；

获取每个物理机当前的 CPU 使用率和内存使用率；

根据所述虚拟机的消耗类型以及获取的每个物理机的 CPU 使用率和内存使用率，按照预设第一规则计算出所述每个物理机的均衡参数；

将所述虚拟机分配给所述均衡参数最大的物理机。

17.根据权利要求 16 所述的系统，其特征在于，所述处理器执行所述存储器中存储的程序时，在执行所述根据获取的 CPU 核数和内存大小识别所述虚拟机的消耗类型时，具体执行以下操作：

计算获取的所述 CPU 核数与所述内存大小的比值；

若所述比值在预设第一范围，识别所述虚拟机的消耗类型为常规类型；

若所述比值在预设第二范围，识别所述虚拟机的消耗类型为 CPU 消耗型；

若所述比值在预设第三范围，识别所述虚拟机的消耗类型为内存消耗型，其中所述预设第二范围大于所述预设第一范围和所述预设第三范围，所述预设第一范围大于所述预设第三范围。

18.根据权利要求 16 所述的系统，其特征在于，所述虚拟机的消耗类型包括常规类型、CPU 消耗型以及内存消耗型，所述处理器执行所述存储器中存储的程序时，在执行所述根据所述虚拟机的消耗类型以及获取的每个物理机的 CPU 使用率和内存使用率，按照预设第一规则计算出所述每个物理机的均衡参数时，具体执行以下操作：

将获取的每个物理机的 CPU 使用率和内存使用率按照预设第二规则计算出每个物理机的第一均衡度；

若所述虚拟机的消耗类型是常规类型，赋予所述每个物理机的均衡参数为对应物理机的第一均衡度；

若所述虚拟机的消耗类型是 CPU 消耗型或内存消耗型，根据获取的每个物理机的 CPU 使用率和内存使用率，计算出 CPU 平均使用率和内存平均使用率；

根据获取的每个物理机的 CPU 使用率和内存使用率以及所述 CPU 平均使用率和所述内存平均使用率计算出每个物理机的第二均衡度；

根据所述虚拟机的消耗类型、所述每个物理机的第一均衡度以及第二均衡度计算出所述每个物理机的均衡参数。

19.根据权利要求 18 所述的系统,其特征在于,所述处理器执行所述存储器中存储的程序时,在执行所述将获取的每个物理机的 CPU 使用率和内存使用率按照预设第二规则计算出每个物理机的第一均衡度时,具体执行以下操作:

计算每个物理机当前的 CPU 使用率的倒数和内存使用率的倒数;

对所有物理机的 CPU 使用率的倒数进行求和得到 CPU 总权重,对所有物理机的内存使用率的倒数进行求和得到内存总权重;

计算每个物理机的所述 CPU 使用率的倒数与所述 CPU 总权重的比值得到 CPU 占比以及每个物理机的所述内存使用率的倒数与所述内存总权重的比值得到内存占比;

根据所计算出的每个物理机的所述 CPU 占比以及所述内存占比和预设的 CPU 相对内存的选择权重计算出第一均衡度。

20.根据权利要求 16 所述的系统,其特征在于,所述处理器执行所述存储器中存储的程序时,还执行以下操作:

根据每个物理机当前的所述 CPU 使用率和所述内存使用率,计算出 CPU 平均使用率和内存平均使用率;

根据每个物理机的所述 CPU 使用率与所述 CPU 平均使用率的离散程度以及每个物理机的所述内存使用率与所述内存平均使用率的离散程度获取综合均衡度,所述综合均衡度用于表示物理机集群的负载均衡度,以便根据综合均衡度来监控物理机集群的负载均衡的情况。

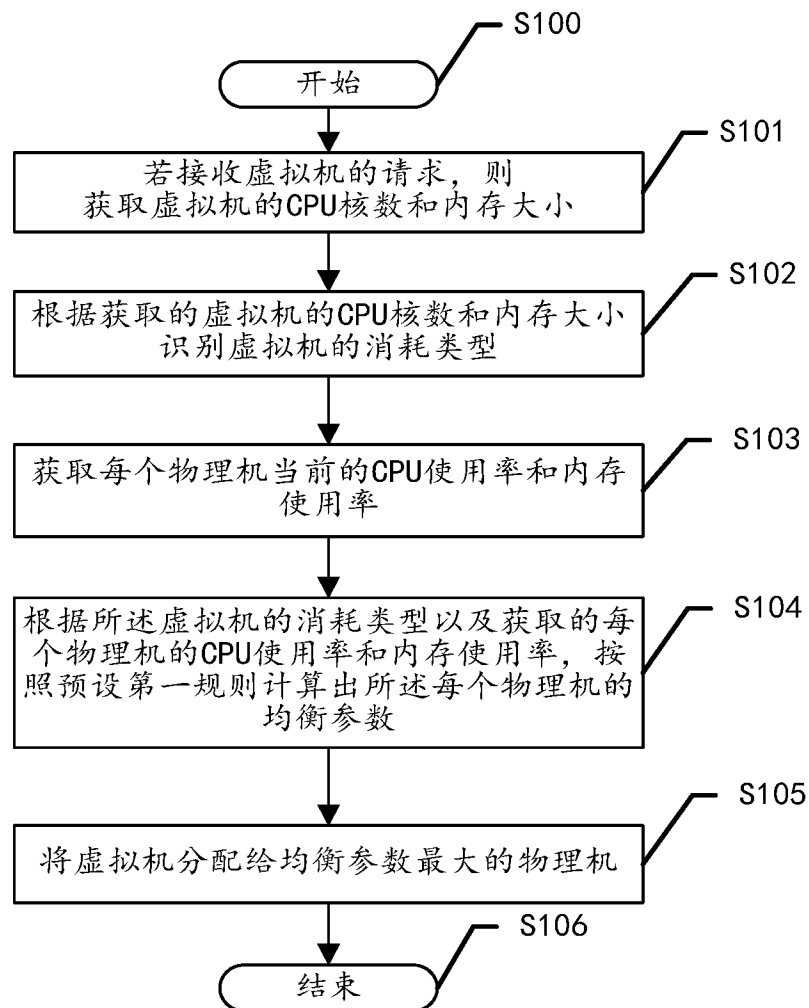


图 1

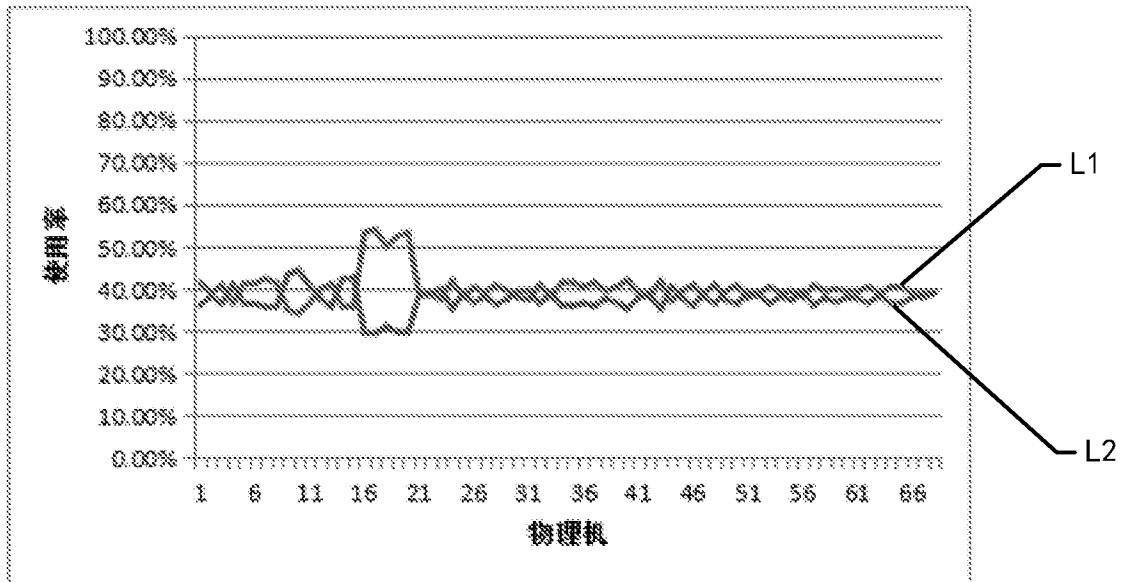


图 2

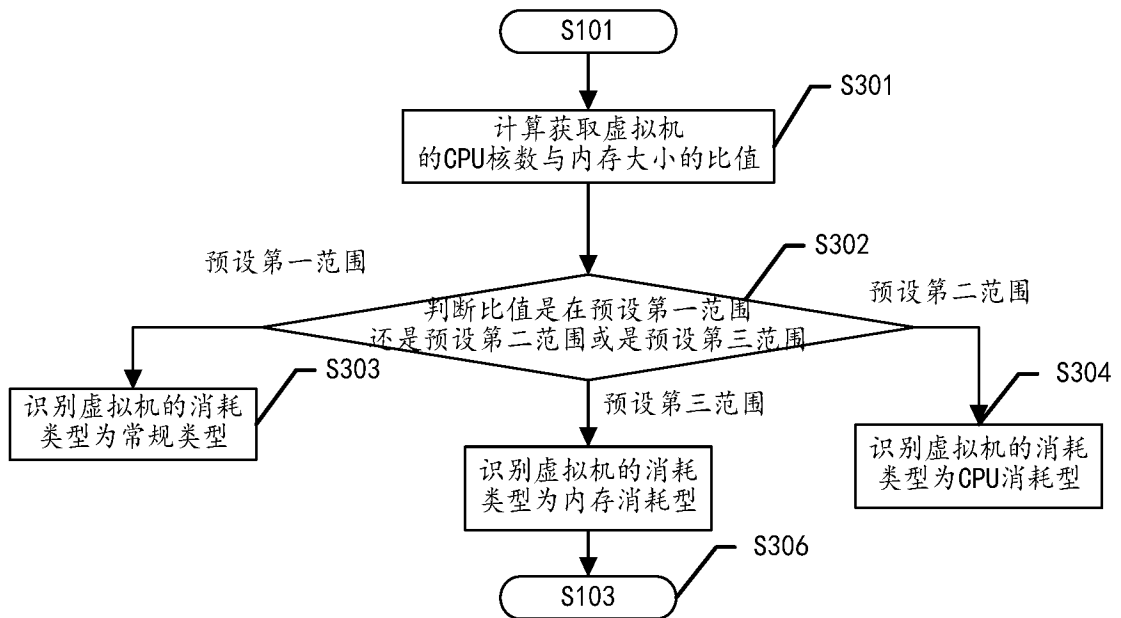


图 3

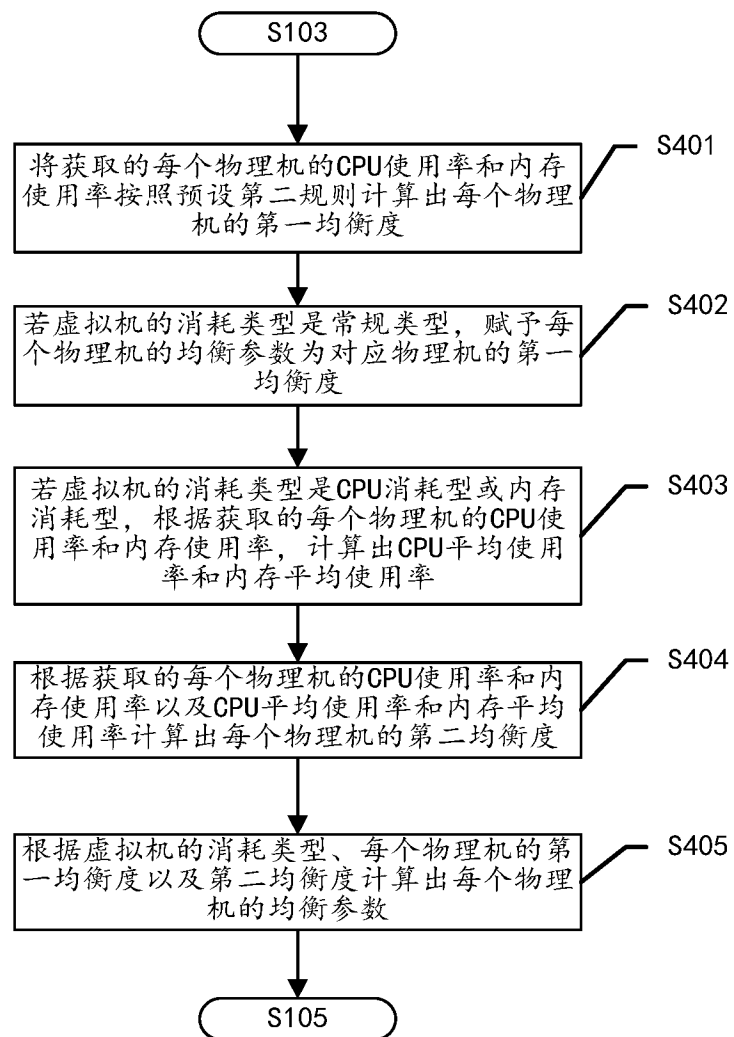


图 4

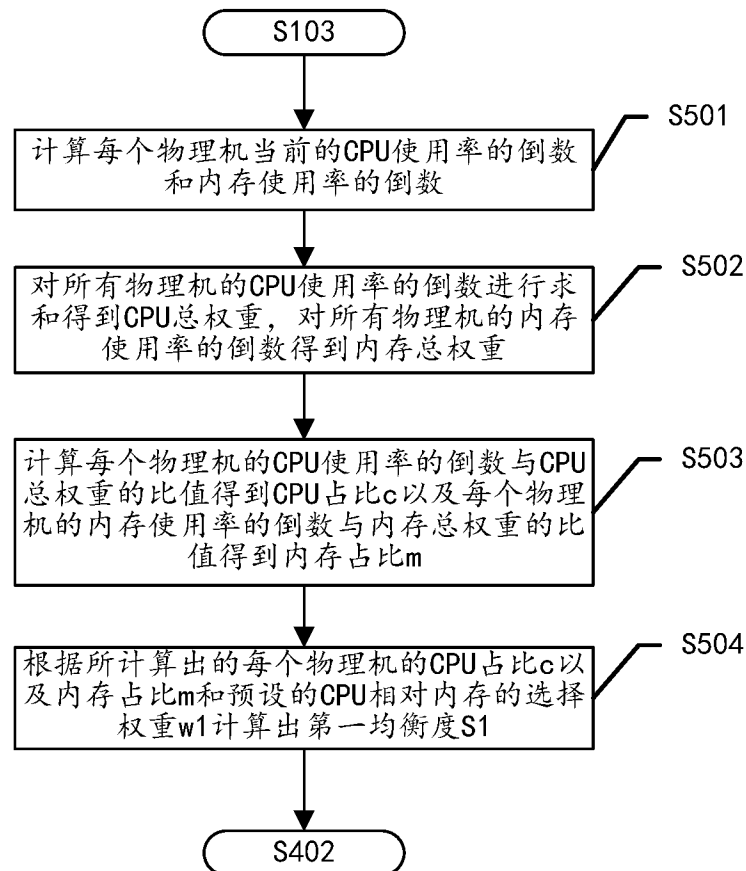


图 5

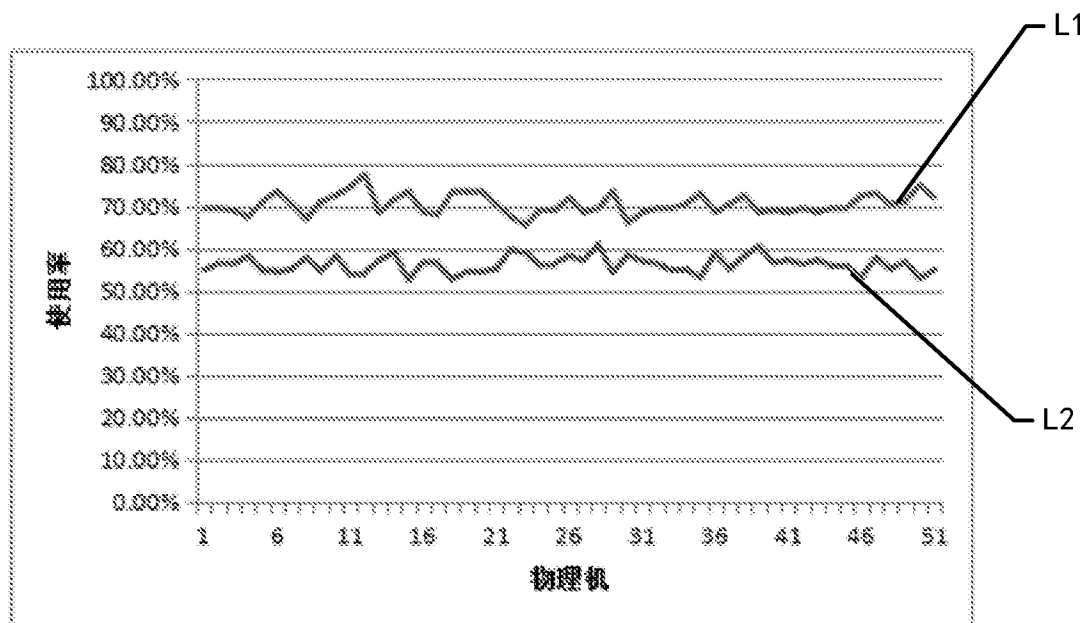


图 6

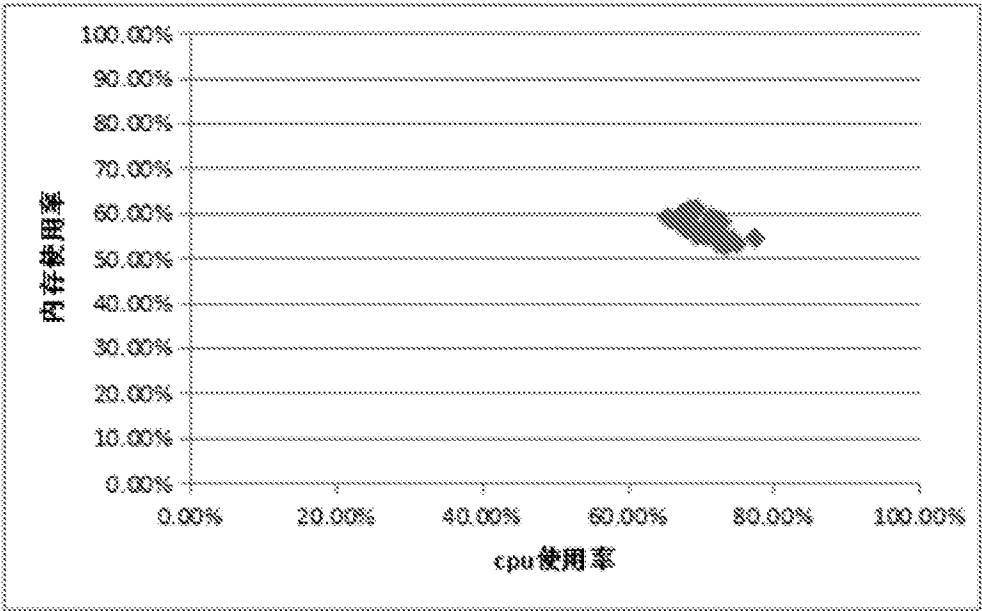


图 7

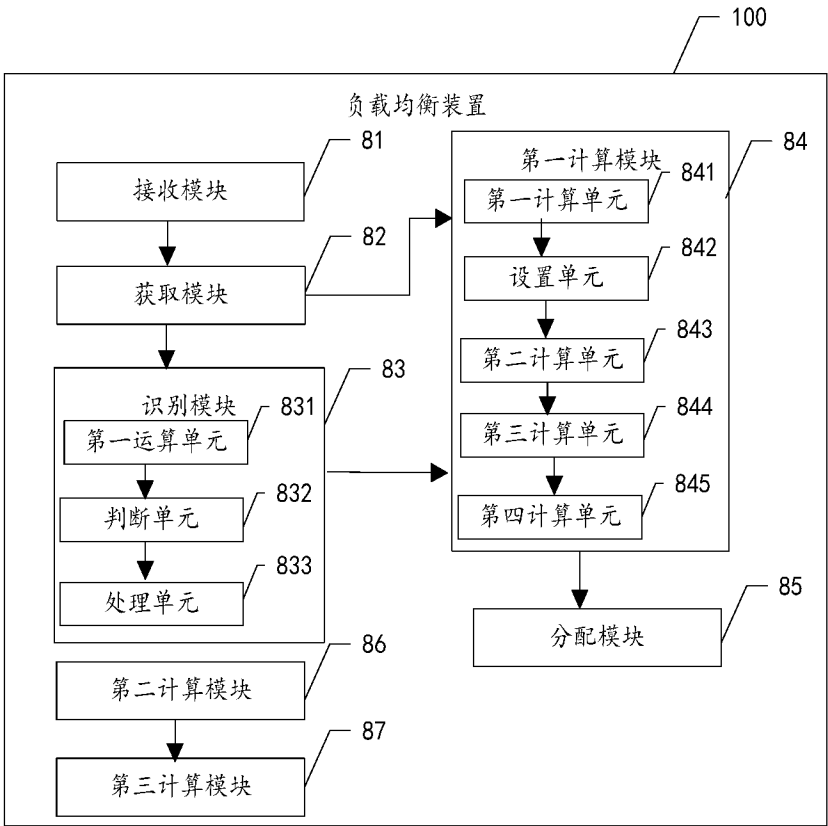


图 8

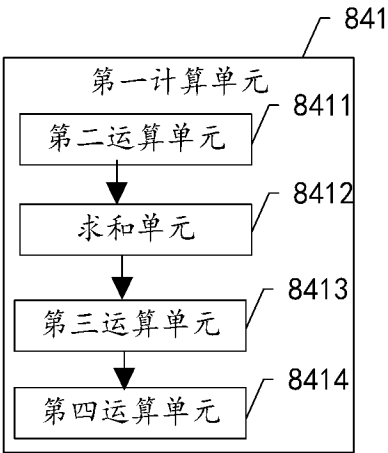


图 9

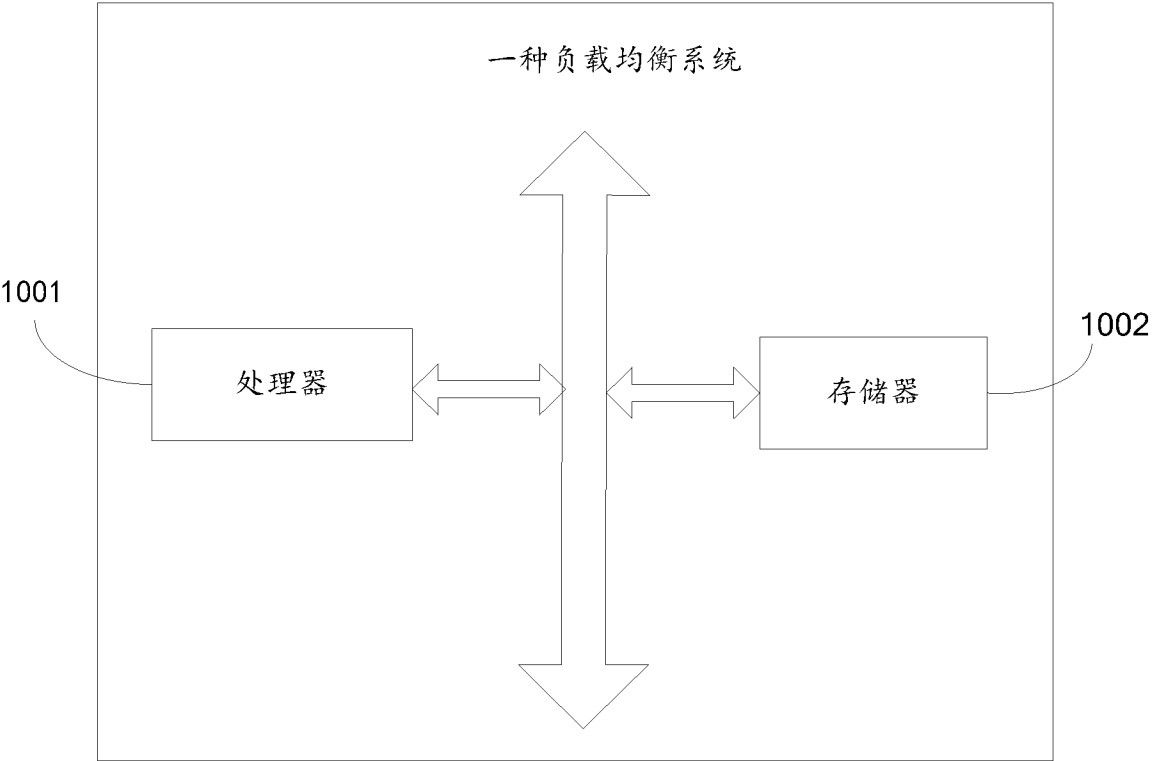


图 10

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2017/076514

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)

WPI, EPODOC, CNPAT, CNKI: CPU, load, balance, equalization, virtual machine, physical machine, physical host, memory, usage rate, utilization rate

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PY	CN 106445629 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.), 22 February 2017 (22.02.2017), claims 1-10, and description, paragraphs [0019]-[0029] and [0139]-[0143]	1-20
Y	CN 103514046 A (HUAWEI TECHNOLOGIES CO., LTD.), 15 January 2014 (15.01.2014), description, paragraphs [0114]-[0129] and [0167]-[0191]	1-20
Y	CN 101504620 A (HUAWEI TECHNOLOGIES CO., LTD. et al.), 12 August 2009 (12.08.2009), description, page 2, lines 5-11	1-20
A	CN 102232282 A (HUAWEI TECHNOLOGIES CO., LTD.), 02 November 2011 (02.11.2011), the whole document	1-20
A	CN 102185779 A (TIAN, Wenhong), 14 September 2011 (14.09.2011), the whole document	1-20

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	“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
“A” document defining the general state of the art which is not considered to be of particular relevance	“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
“E” earlier application or patent but published on or after the international filing date	“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
“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)	“&” document member of the same patent family
“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	

Date of the actual completion of the international search 02 June 2017 (02.06.2017)	Date of mailing of the international search report 13 June 2017 (13.06.2017)
Name and mailing address of the ISA/CN: State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No.: (86-10) 62019451	Authorized officer LI, Jia Telephone No.: (86-10) 53311134

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2017/076514

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 106445629 A	22 February 2017	None	
CN 103514046 A	15 January 2014	None	
CN 101504620 A	12 August 2009	None	
CN 102232282 A	02 November 2011	WO 2011110026 A1	15 September 2011
		US 2012204176 A1	09 August 2012
CN 102185779 A	14 September 2011	None	

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)) WPI, EPODOC, CNPAT, CNKI: 负载, 均衡, 虚拟机, 物理机, CPU, 内存, 使用率, load, balance, equalization, virtual machine, physical machine, physical host, memory, usage rate, utilization rate																				
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>PX</td> <td>CN 106445629 A (平安科技深圳有限公司) 2017年 2月 22日 (2017 - 02 - 22) 权利要求1-10, 说明书[0019]-[0029]、[139]-[143]段</td> <td>1-20</td> </tr> <tr> <td>Y</td> <td>CN 103514046 A (华为技术有限公司) 2014年 1月 15日 (2014 - 01 - 15) 说明书[0114]-[0129]、[0167]-[0191]段</td> <td>1-20</td> </tr> <tr> <td>Y</td> <td>CN 101504620 A (华为技术有限公司 等) 2009年 8月 12日 (2009 - 08 - 12) 说明书第2页第5-11行</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 102232282 A (华为技术有限公司) 2011年 11月 2日 (2011 - 11 - 02) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 102185779 A (田文洪) 2011年 9月 14日 (2011 - 09 - 14) 全文</td> <td>1-20</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	PX	CN 106445629 A (平安科技深圳有限公司) 2017年 2月 22日 (2017 - 02 - 22) 权利要求1-10, 说明书[0019]-[0029]、[139]-[143]段	1-20	Y	CN 103514046 A (华为技术有限公司) 2014年 1月 15日 (2014 - 01 - 15) 说明书[0114]-[0129]、[0167]-[0191]段	1-20	Y	CN 101504620 A (华为技术有限公司 等) 2009年 8月 12日 (2009 - 08 - 12) 说明书第2页第5-11行	1-20	A	CN 102232282 A (华为技术有限公司) 2011年 11月 2日 (2011 - 11 - 02) 全文	1-20	A	CN 102185779 A (田文洪) 2011年 9月 14日 (2011 - 09 - 14) 全文	1-20
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																		
PX	CN 106445629 A (平安科技深圳有限公司) 2017年 2月 22日 (2017 - 02 - 22) 权利要求1-10, 说明书[0019]-[0029]、[139]-[143]段	1-20																		
Y	CN 103514046 A (华为技术有限公司) 2014年 1月 15日 (2014 - 01 - 15) 说明书[0114]-[0129]、[0167]-[0191]段	1-20																		
Y	CN 101504620 A (华为技术有限公司 等) 2009年 8月 12日 (2009 - 08 - 12) 说明书第2页第5-11行	1-20																		
A	CN 102232282 A (华为技术有限公司) 2011年 11月 2日 (2011 - 11 - 02) 全文	1-20																		
A	CN 102185779 A (田文洪) 2011年 9月 14日 (2011 - 09 - 14) 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 2017年 6月 2日		国际检索报告邮寄日期 2017年 6月 13日																		
ISA/CN的名称和邮寄地址 中华人民共和国国家知识产权局 (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 李佳 电话号码 (86-10) 53311134																		

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2017/076514

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	106445629	A	2017年 2月 22日	无	
CN	103514046	A	2014年 1月 15日	无	
CN	101504620	A	2009年 8月 12日	无	
CN	102232282	A	2011年 11月 2日	WO 2011110026 A1	2011年 9月 15日
				US 2012204176 A1	2012年 8月 9日
CN	102185779	A	2011年 9月 14日	无	