**一种“为大留大”异构资源的虚拟机在线调度方法与装置**

# 1．发明背景与现有技术

随着虚拟化云计算技术的不断推进，各种云计算平台已经广泛应用于产业界。云计算带给IT一种“计算即服务”的新应用模式，原有分散的计算资源在逻辑上聚集为资源池，基于服务模式按需分配资源，即云平台的计算服务是可以被租用的。用户可以随时租用虚拟机，即向云平台请求云计算服务，然后云平台通过分配计算资源执行应用程序来满足用户需求。在虚拟化技术支持下，这种服务供应通过虚拟机部署、调度物理机计算资源执行应用程序来实现。同时，虚拟机作为一种应用执行的环境，又被部署在特定的物理机中来执行应用程序。于是，在云平台的资源调度中，以虚拟机为单位，虚拟机给出了逻辑上用户对计算资源的需求，云平台的调度系统基于整体系统资源分配的目标考虑，进行决策为其选择合适的物理机进行部署虚拟机。现有云平台虚拟机分配是一种基于整体的分配，即不能跨越多个物理机进行分配。在云平台的资源调度中，广泛存在的问题是在线资源调度。随着用户计算服务请求不断到来，高效的云平台应能实时快速地做出处理，使用户能够获得满意的服务，避免损失。本发明将以在线资源调度为背景，旨在针对异构资源场景给出一种有效的在线资源调度方案。

然而，云平台的在线资源调度与物理机资源、虚拟机资源请求向量的多维度、大小、物理资源容量、调度策略、调度优化目标等诸多要素直接相关。虚拟机资源请求的不同，物理机资源自身的异构，以及云平台基于成本、性能、节能等各种优化目标下的理论求解最优资源调度决策算法的”NP”（非多项式时间）之难，都为寻找有效的大规模云平台资源调度方法带来诸多困难。现有云平台的虚拟机在线资源调度方法中，为了初始放置，有基于节能、最大化资源利用率而最小化同构物理机数量为基础的虚拟机调度部署决策方法；为了负载均衡，有基于系统资源负载状态确保各物理机之间的负载均衡的资源调度方法；为了系统可靠性和高可用性，基于某个或各个维度资源负载的热点预测与消除的调度方法。然而这些方法大多关注的是资源调度本身,为简化资源调度决策做出了基于同构物理资源的假设，因此，对于异构资源场景缺乏有效的在线调度方法。基于此，本专利旨在为云平台的在线资源调度给出一种基于“为大留大”的异构资源调度装置，本装置“为大留大”的目的是使在线调度方法能够为用户请求中的大虚拟机预留大物理机资源，有效实现异构物理资源调度分配，使随时到来的各种大小不同的虚拟机请求，在及时进行分配时保证有合适的物理机资源分配。

事实上，现有在线调度方法大多为了简化问题求解，只考虑解决同构场景的物理机资源选择问题。在同构场景下的在线调度方法，当需要为一个虚拟机资源请求分配物理机资源时，在线调度时往往对物理机的选择决策通常采用随机任意选取一台基于容量能够放置的最合适的物理机的方法。然而，在异构场景中，由于物理机资源异构的复杂性，意味着初始物理机资源容量有大小，物理机的选择随时受到虚拟机请求大小的约束，现有调度方法难以确保大物理机为大虚拟机使用。比如，在线调度过程中，对于给定的虚拟机资源请求，在进行物理机的选择决策并且同时有多个物理机满足容量大小的放置要求时，此时存在决策“先分配小容量物理机，还是大容量物理机”的问题。由于目前的虚拟机技术中单个虚拟机的资源分配不能跨多个物理机，因此虚拟机只能放置在基于容量大小放得下的一个物理机中。所以，当虚拟机请求到来时，对当前虚拟机的在线资源分配，应考虑即将到来的未来不可预知的大容量虚拟机放置的需要。所以，基于同构假设的对物理机一视同仁的随机选取物理机的方法，已经不适合于异构场景的在线调度。

# 2．本发明技术方案

鉴于同构场景下的在线调度方法不能有效解决异构场景的在线调度问题，原因是没有考虑为即将到来的更大资源需求的虚拟机调度预留足够的物理机资源造成未来虚拟机甚至找不到合适的物理机而使调度失败，或开启新额外物理机占用额外资源开销，降低系统资源利用率。为有效解决此问题，本发明结合了大虚拟机请求到来的不可预知性，在云平台的在线调度中引入一种基于“为大留大”的异构物理机资源的选取策略，通过对资源多维折算、标准化归一处理得到物理机资源的综合大小，然后给出一种异构程度系数来刻画多个物理机之间的异构程度，通过优先分配具有小资源容量的物理机，以保证预留一定比例的大容量物理机，为在线过程中即将到来的不可预知的具有大容量资源请求虚拟机放置给出足够容量，确保在异构场景下在线调度资源分配的有效性。

1. **异构资源在线调度装置**

如图1，本发明技术方案给出一种基于“为大留大”异构资源的虚拟机在线调度装置，本装置可作为云平台的一个组件，直接为云平台调度使用。云平台首先接收来自用户的虚拟机在线调度请求，然后将请求发送给在线调度装置，同时提供放置优化目标，调度装置进行处理，返回在线虚拟机具体的放置物理机位置给云平台，最后由云平台实施具体的虚拟机放置，如在给定的物理机上开启虚拟机，并满足特定的系统优化目标。优化目标可以是负载均衡，也可以是少物理机节能。



图1 异构资源在线调度装置

1. **基于两阶段的处理流程**

本装置集成了“为大留大”异构资源的虚拟机在线调度方法，如图2，其基本实现过程分为两个阶段:物理机预处理与虚拟机在线放置处理阶段。

由于物理机资源是一个多维向量，并具有不同的单位量纲，如CPU按照核的个数，内存按照MB来计，为区分不同资源大小的物理机，需要对多维度资源进行降维处理，得到一个综合值，然后基于这个值来区分物理机的大小。因此，本发明在物理机资源预处理阶段，完成物理资源大小的计算，然后对给定的异构物理机按资源容量大小进行排序，接着对物理机进行分组，最后再基于Rounding方法取各维度资源值最小值得到一个资源向量表示该组的具体资源大小。在虚拟机请求处理阶段，为虚拟机选择合适的物理机进行放置，包括选择可放置该虚拟机的物理机集合、选可放置的物理的子集合、从子集合中选放置该虚拟机的物理机。



图2 两阶段流程图

**（1）问题假设。**假设一个云平台的物理机台数为M，待放置的虚拟机台数为N，虚拟机和物理机的资源维度为D，物理机i维度d的资源容量为，各维度资源的权重（资源越重要，权重值越大），虚拟机的资源需求为。

**（2）阶段一：异构物理机资源预处理。**

如图3，其基本流程如下：

**第1步：输入物理机的资源容量。**

**第2步：基于消除量纲不一致的归一化处理。**将各维度资源值进行标准化处理，转换为0-1之间的值。对所有物理机，对每个维度d, 取该资源维度的最大值，利用公式： ,对每个i,计算出物理机i各维度的资源容量的标准值，得到标准化物理机资源值向量。

**第3步：基于历史虚拟机请求资源信息获取物理机各维度资源权重。**云平台给定一组历史虚拟机的请求资源，则。

**第4步：基于降维的多维资源折算处理。**以标准值为基础，基于加权求和的方法计算各物理资源的综合值，表示实际物理机的大小。计算方法：对每个物理机，使用加权和公式：作为物理机的综合资源。

**第5步：基于综合资源值升序排序处理。**按照的大小将全部物理机进行从小到大升序排序。

**第6步：基于常数k确定异构程度阀值Threshold。**由云平台输入常数k和物理机台数TN，则Threshold=k/TN。

**第7步：基于异构程度对物理机进行分组处理。**为更好区分不同物理机的异构性，依据资源值大小构成和物理机资源构成的不同将物理机进行分类，引入一个指标：物理机异构程度（Heterogeneous degree），定义为物理机的种类数（Type）和物理机总台数 (Total Number)的比值，即。给定一个表示异构程度强弱的阀值为Threshold，按照以下两种情况进行分类：

1. 如果（说明物理机的异构程度比较低），则按照物理机的类别，将物理机化归为几类；
2. 如果（说明物理机的异构程度比较高），设定某个常数，对已经排好序的物理机从最小的开始，每连续的个物理机划归为一组，最后剩下的不满个为一组。

**第8步：基于Rounding方法确定每个组的资源向量，作为物理机组的基本向量。**

Rounding的思想是向下取值，我们为每个异构物理机资源组，给出一个组资源值指标。指标值的获得方式是通过选取该组中的各物理机的各维度资源非零最小值组成的资源向量作为该组的资源向量。

基本流程如图3所示。



**图3 异构物理资源预处理流程**

**阶段二：虚拟机在线放置处理。**基本流程如图4所示。



图4 虚拟机在线放置流程

**第1步：输入虚拟机资源请求序列，物理机列表，台数，物理机资源容量，分组，以及放置优化目标信息。**优化目标是负载均衡或节能，被事先设置为一个标记值，如1表示负载均衡，2表示做节能。

**第2步，对每个虚拟机，遍历一阶段预处理完成后的物理机组，判断该组资源向量是否能够满足虚拟机的资源请求，如果可以，将该组中的全部物理机加入到集合中，**构成集合，并且要求中的物理机将保持物理机预处理后的升序排序，即；

**第3步： 基于放置优化目标确定最佳放置物理机。为实现“为大留大”策略，**即考虑为将来大虚拟机的放置预留大容量物理机。本发明引入基于预留系数的选取最佳物理机方法，即给出预留系数为，满足，选取物理机时只要使选出的子集合：满足和，求出最合适的。本发明又取为黄金分割点位置0.618，即黄金分割资源预留系数法，黄金分割法是随机不确定性理论上往往达到最优值的一个值，符合人们实际。

**第4步：结合系统优化目标，基于贪心的虚拟机放置策略进行放置。**如果是负载均衡，则从可选的物理机中选取一个剩余资源最大的物理机放入；如果是节能，则从物理机中选取一个能使装入该虚拟机后剩余资源最少的物理机，选取FF（first fit）、BF（best fit）、WF（worst fit）和随机算法中的一种方法，从子集合中选出一台物理机，将虚拟机放入。最后更新该物理机所在的组的资源向量。

**第5步：输出所有虚拟机的放置结果，完成放置，结束在线调度过程。**

# 3．本发明的技术保护点

本发明有以下几个技术保护点：

**保护点1：**给出了一种可作为云平台调度系统插件的“为大留大”的异构资源虚拟机在线调度装置。

**保护点2：**异构物理机预处理过程中，对于多维度物理机资源采用了去除量纲不一致的归一化处理方法的应用，基于归一化、加权求和降维处理的多维异构物理机资源的资源折算方法应用。

**保护点3：**物理机处理过程中，第一次引入异构程度系数，来刻画物理机之间的异构属性，给出了基于物理机的异构程度的分组、Rounding组向量确定方法。

**保护点4：**在线调度处理过程中，第一次引入预留系数，提出了基于“为大留大”的虚拟机异构资源调度方法，对所有满足物理机进行排序，利用预留系数，确保能够留出一部分大物理机，为将来的大虚拟机所用。第一次引入基于0.618黄金分割法确定资源预留系数，取的是随机不确定性选择理论上的最优值，比较符合实际，为虚拟机从可放置的物理机集合中选取部分物理机，从而预留出综合资源较大的物理机，使得后面出现资源请求较大的虚拟机时，也可以有大的物理机能够满足资源要求，从而使得云平台可以尽量服务发出请求的用户，然后最终目标是使得云平台总体的资源的利用率最高，即收益最大。

# 4．本发明有益效果

本发明的有益效果归纳为如下四点：

**保护点1**给出了一个虚拟机异构资源在线调度装置，提出了基于“为大留大”的虚拟机在线调度方法，解决了云计算平台基于异构物理资源的虚拟机在线放置调度问题，并且此方法可扩展用于虚拟化集群系统。

**保护点2**给出的消除量纲不一致的异构多维资源的折算、资源降维归一处理方法的应用，为异构资源的处理提供了一种有效方法，为异构物理资源的排序、大物理资源的预留提供了依据。

**保护点3**首次引入异构程度指标，基于此的分组Rouding确定组向量的方法解决了云平台虚拟机在线调度过程中异构物理机资源无法刻画的问题。这样的做法降低了搜索物理机的规模，提高了在线资源调度的效率。

**保护点4**首次引入预留系数，量化了大物理机预留数量，提出以黄金分割系数的方法获得最优值，切合实际，体现了具有实用性。基于0.618黄金分割位的资源预留方法的引入，黄金分割位比较符合人们的习惯，有效避免了资源预留系数常数选取的不确定性。

# 5．检索情况

|  |  |  |  |
| --- | --- | --- | --- |
| 国家/地区 | 检索网站 | 检索公式 | 文献数量 |
| US/EP | https://www.google.com.hk/?tbm=pts | physical machine， heterogeneous resource， sort | 0 |
| US/EP | https://www.google.com.hk/?tbm=pts | physical machine， heterogeneous degree， group | 0 |
| US/EP | https://www.google.com.hk/?tbm=pts | virtual machine， placement | 5 |
| US/EP | https://www.google.com.hk/?tbm=pts | physical machine， rounding | 0 |
| CN | <http://www.sipo.gov.cn/zljs/> | 异构资源 虚拟机 在线调度 | 0 |
| CN | <http://www.sipo.gov.cn/zljs/> | 物理机（可有可无）， 异构资源， 排序 | 1 |
| CN | <http://www.sipo.gov.cn/zljs/> | 异构资源， 排序， 分组 | 0 |

**关键字**：虚拟机在线调度， online scheduling, 资源预留（resource reservation）, 异构资源，heterogeneous resource, 虚拟机部署（Virtual machine placement）

# 6．典型专利分析

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **专利名称** | **专利号** | **机构** | **所涉及 领域** | **内容** | **该专利保护的内容** |
| Scheduler of virtual machine module, scheduling method therefor, and device containing computer software | US  2012  8241835  B2 | Institute for information industry, Taipei | Virtual machine，scheduling method | 该专利定义了拥有分类模块和调度模块的调度器。分类模块接受VM模块的信息，并分析通过分析资源比例参数得到分类结果信息；调度模块根据分类模块得到的信息做出调度决策。 | 该专利保护了基于分类的虚拟机调度算法。 |
| Scheduling system | EP  2011  2323035  A1 | VirtualLogix SA | Scheduling | 该专利包括，第一个调度步骤中，选择虚拟机物理CPU上运行；第二个调度步骤中，选择vCPU在VM上运行；再选取一个物理CPU，并将上述VM放在该物理CPU上运行。 | 该专利保护了关于数据处理系统的调度方法。 |