

# An Efficient Virtual Resources Allocation Strategy for Cloud Computing

Ning Yu<sup>1,2</sup>, Jinquang Zhang<sup>1,2\*</sup>, Lina Ni<sup>1,2</sup>

<sup>1</sup>College of Information Science and Engineering, Shandong University of Science and Technology, Qingdao Shandong

<sup>2</sup>Shandong Province Key Laboratory of Wisdom Mine Information Technology, Shandong University of Science and Technology, Qingdao Shandong

Email: 920970896@qq.com, \*tjzhangjinquan@126.com, nln2004@163.com

Received: Mar. 5<sup>th</sup>, 2016; accepted: Mar. 26<sup>th</sup>, 2016; published: Mar. 29<sup>th</sup>, 2016

Copyright © 2016 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

In this paper, a virtual resource application-allocation model and an efficient virtual resource allocation strategy are presented aiming at the huge energy consumption problem during the idle time of host server in cloud data center. The strategy can maximize the use of the host server and reduce energy consumption of the infrastructure under the premise of meeting the users' needs. The cloud tasks are submitted to the cloud center after the statistical synthesis at the ends of the cloud, and unified allocated by the resource manager of cloud center in order to achieve optimal utilization of the servers. Simulation results show that our proposed strategy has good effect for different number and type of cloud tasks scheduling, and can improve the utilization of cloud computing center server.

## Keywords

Cloud Computing, Virtual Resource, Allocation Strategy

---

# 云计算中虚拟资源高效分配策略研究

于 宁<sup>1,2</sup>, 张金泉<sup>1,2\*</sup>, 倪丽娜<sup>1,2</sup>

<sup>1</sup>山东科技大学信息科学与工程学院, 山东 青岛

<sup>2</sup>山东科技大学山东省智慧矿山信息技术省级重点实验室, 山东 青岛

\*通讯作者。

Email: 920970896@qq.com, \*tjzhangjinqian@126.com, nln2004@163.com

收稿日期: 2016年3月5日; 录用日期: 2016年3月26日; 发布日期: 2016年3月29日

## 摘要

本文针对目前云数据中心主机服务器在空转时间段对能源消耗巨大的问题, 建立了虚拟资源申请 - 分配模型, 提出了一种虚拟资源分配的策略。该策略在满足用户需求前提下, 对主机服务器最大化利用, 以此来降低基础设施对能源的消耗。通过“端”对云任务进行统计综合, 提交云中心, 云中心的资源管理器对云任务进行统一分配, 以达到对服务器的最佳利用。仿真结果表明, 我们提出的策略对不同任务数目、类型的云任务调度具有较好的效果, 能够提高云计算中心服务器的利用率。

## 关键词

云计算, 虚拟资源, 分配策略

## 1. 引言

云计算通过互联网提供动态、易扩展、虚拟化的资源, 提供这些资源的往往是具有庞大规模计算机集群的云数据中心。通常情况下, 一个云计算中心有数万甚至数百万的服务器, 这些云计算中心的能源消耗是巨大的。可以通过两种途径减少这些大规模基础设施对能源的消耗: 一是降低单一服务器的能耗; 二是进行虚拟资源优化来提高服务集群的整体资源利用率。

目前对于虚拟资源优化的相关研究主要集中于优化资源调度、优化资源分配以及资源有效管理等方面。有学者提出分层调度方法[1]、多目标蚁群算法[2]以及引入经济学方法比如纳什均衡[3]、市场经济机制[4]等提高资源利用率。文献[5]提出多目标遗传算法, 通过设计适应度函数, 解决动态资源分配问题。文献[6]针对资源调度问题中的成本最小化和期限约束模型, 提出了基于遗传算法的动态目标优化策略。文献[7]针对不同的资源类型, 提出了多资源包装算法, 并在 iVCE 中实现了一个原型。文献[8]设计了一个体系架构, 根据应用的特征和当前的工作负载, 在云集群环境中找到最优的 cpu 密集型和 io 密集型工作的分配比例。文献[9]在讨论了多种基本要素的虚拟机迁移情况下, 提出了基于队列的虚拟机迁移模型。文献[10]在考虑时间价格差异和对任务延迟的容忍度情况下, 提出了一个动态的虚拟资源租赁算法, 来使得云服务提供商尽可能达到利润最大化。本文研究虚拟资源高效率整合方法, 提高计算中心的资源利用率, 以达到减少能源消耗、降低运行成本的目的。

## 2. 系统模型

### 2.1. 预备知识

虚拟资源可分为: 计算型、通信型、磁盘型以及三者之间的互相组合[11]。本文所提出的资源分配策略, 在各类资源整体综合的基础上借助云端来进行, 将虚拟资源分成最基本的三类: 计算型、通信型、磁盘型。其中云端代表着云计算中的应用企业。用户通过“端”这个桥梁来获取服务, “端”向服务集群申请资源, 在“端”进行资源统计。假设用  $P$  代表计算型资源,  $N$  代表通信带宽型资源,  $D$  代表磁盘存储型资源, 那么对于用户  $i$  的申请, 端可以建立一个向量来指向用户  $i$  所申请的资源, 并且对用户申请进行统计, 向服务集群进行集中申请。

## 2.2. 问题模型

我们对整个资源申请过程建立模型。其中用户通过端向云服务集群申请资源，端在统计用户的申请后，向资源管理器递交统计结果，资源管理器通过综合分析资源申请与云服务集群中资源的使用情况，计算判断得出虚拟资源分配的最优配置，并将结果提交给云服务集群，最后云集群分配资源。图 1 为我们所提出的虚拟资源申请 - 分配模型。

模型将云资源申请 - 分配整体流程概括化表示。云用户是虚拟资源的申请者，代表云计算模型中的消费者。云服务集群为虚拟资源的提供者，代表云计算模型中的生产者。“端”代表云计算模型中的中间商，通过利用资源管理器，统计云用户的资源申请，在进行综合分析的基础上，向云集群申请到合适的资源并分配给云用户。这样也就完成了一次云资源申请 - 分配的流程。

假设在某一时刻有  $n$  个用户向端提交资源申请，对于这一时刻，端对云用户所申请资源的统计结果用一个  $n \times 3$  的资源申请矩阵  $s1$  表示：

$$s1: \begin{pmatrix} P1 & N1 & D1 \\ P2 & N2 & D2 \\ \vdots & \vdots & \vdots \\ Pn & Nn & Dn \end{pmatrix}$$

假设服务集群中每个主机服务器能承载的最大资源数为  $M$ 。对于整个服务器集群，可分为三类：正在使用且已存储满、正在使用且未用、未使用。第一种可以不予考虑，第二种我们用资源可分配矩阵  $m$  表示，其中  $m_i$  表示第  $i$  台主机剩余可分配的资源。

$$m: \begin{pmatrix} P1' & N1' & D1' \\ P2' & N2' & D2' \\ \vdots & \vdots & \vdots \\ Pn' & Nn' & Dn' \end{pmatrix}$$

在对整个资源申请进行建模分析之后，虚拟资源分配问题可简化为两个分支：

- 一是如何将资源申请首先分配到运行中的未用服务器中；
- 二是如何将未分配的资源进行优化配置，使得新开服务器可以最佳利用。

## 2.3. 解决方案

本文的虚拟资源申请 - 分配模型采用以下步骤：1) 检测运行中的服务器是否可继续分配；2) 可分配则分配并且继续检验；3) 若不可分配则将剩余申请进行综合配置并分配到新开服务器中。

第一个分支：检测运行中的服务器是否可继续分配，可以通过比对建立的两个矩阵。

$$\begin{pmatrix} P1 & N1 & D1 \\ P2 & N2 & D2 \\ \vdots & \vdots & \vdots \\ Pn & Nn & Dn \end{pmatrix} \Leftrightarrow \begin{pmatrix} P1' & N1' & D1' \\ P2' & N2' & D2' \\ \vdots & \vdots & \vdots \\ Pn' & Nn' & Dn' \end{pmatrix}$$

将申请资源矩阵与可分配资源矩阵进行逐一比对。考虑资源分配效率，比对若满足条件  $(P_i, N_i, D_i) \leq (P_i', N_i', D_i')$ ，则直接分配。分配完毕，双方更新，重新比对。这里我们新建一个矩阵  $s2$  作为比对完成后的剩余资源申请矩阵， $s2$  中的矩阵行存储比对完成后剩余矩阵，表示正在运行的服务器不能满足其需求，需要重开服务器。

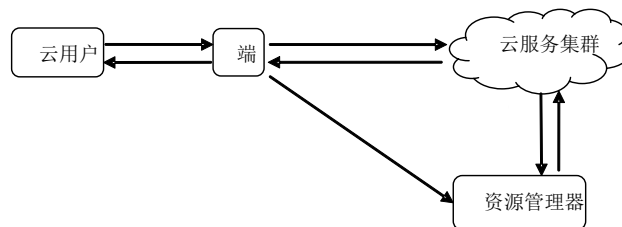


Figure 1. The allocation model of virtual resources  
图 1. 虚拟资源分配模型

若矩阵  $s_2$  非空，则表示有资源申请未满足，并且运行中的服务器已经不能满足资源分配需求，因此需要重开服务器。

这里我们引入线性不等式方程。通过求解线性不等式方程来使得新开服务器资源利用率达到最大。因为  $s_2$  矩阵是从资源申请矩阵抽离出的一部分，我们假设仍有  $t$  个任务未分配，用如下矩阵表示：

$$s_2: \begin{pmatrix} P_1 & N_1 & D_1 \\ P_2 & N_2 & D_2 \\ \vdots & \vdots & \vdots \\ P_t & N_t & D_t \end{pmatrix}$$

对于新开服务器，其能承载的最大资源数为  $(M_1, M_2, M_3)$ ，我们建立一个线性不等式方程。设  $(a_1, a_2, \dots, a_t)$  为一台服务器的最佳配置，可列下列线性方程不等式：

$$\begin{aligned} a_1 P_1 + \dots + a_i P_i + \dots + a_t P_t &\leq M_1 \\ a_1 N_1 + \dots + a_i N_i + \dots + a_t N_t &\leq M_2 \\ a_1 D_1 + \dots + a_i D_i + \dots + a_t D_t &\leq M_3 \end{aligned}$$

将此不等式进行整理得：

$$(a_1 \ a_2 \ \dots \ a_t) \begin{pmatrix} P_1 & N_1 & D_1 \\ \vdots & \vdots & \vdots \\ P_i & N_i & D_i \\ \vdots & \vdots & \vdots \\ P_t & N_t & D_t \end{pmatrix} \leq (M_1, M_2, M_3)$$

对于此不等式，因为缺少目标函数，所以并没有确切的解。而解此不等式的最好办法是枚举所有可能性，即将  $(a_1, a_2, \dots, a_t)$  中，从  $(0, 0, \dots, 0)$ 、 $(0, 0, \dots, 1)\sqrt{2}$  枚举到  $(1, 1, \dots, 1)$ ，然后从其中选出最优解。针对虚拟资源的最优配置理论，我们再添加额外限制条件，令单台主机服务器接受最大数目的申请任务，即：令  $\max = a_1 + a_2 + \dots + a_t$ ，使得  $\max$  最大。

因为云计算面向的客户群是庞大的，同样资源申请数目也是庞大的。所以在建立的  $s_2$  矩阵中，会有多行。因此若对  $s_2$  直接处理，计算量很大。本文选择对  $s_2$  进行分组，其中分组数目综合考虑云任务申请资源大小以及服务器的配置高低。假设分组完成后每组所含云任务数量为 COUNT，此时的  $t = \text{COUNT}$ 。再解上面的线性不等式，根据 COUNT 的数量来决定难易。

### 3. 仿真性能分析

#### 3.1. 仿真环境

仿真平台使用 Java 在 Eclipse 上开发。平台模拟拥有一定数量服务器的数据中心，并随机产生一定

数目的云任务用于实例验证。云任务的数目为 1~7000，已开启的服务器数量定义为 1~50 台之间，未开启的服务器数目为 0~500 台。假定每台主机具有 4 G 内存、10 M 带宽、10 G 存储。

### 3.2. 真性能与分析

1) 固定分组数目，对比先到先得策略与本文策略。

首先固定云任务分组数量为每组 10 个，随机产生 100~7000 个云任务，对比先到先得策略(First Come First Served)与本文策略，如图 2 所示。

系统在接收到虚拟云任务后，通过对云任务进行统计，得到在最优情况下需要开启的主机服务器数目，然后对比先到先得策略与本文策略实际开启的主机数目。图中显示，在云任务只有较少数量时，产生的主机额外开销比较大。但是随着云任务数量的增加，两种策略产生的主机额外开销基本稳定。通过对比，本文的策略优于先到先得策略，主机额外开销稳定在 7% 左右。

2) 分组数目对策略的影响

本文采取策略需要分组来使得新开主机达到最优利用，分组之后每组虚拟云任务数量一般在 5~19 之间，主要根据主机配置以及云任务大小来定。固定云任务数量为 2000，通过对比每组 6~19 个云任务的分组，如图 3 所示，随着每组云任务数量的不断增多，额外主机开销渐趋稳定，但策略运行时间急剧增加。因此通过考虑各方面因素，每组云任务数量定在 10~14，策略会达到最优。

### 4. 总结

本文研究在满足云计算环境用户资源申请的前提下，如何有效降低数据中心服务器的能耗。论文提

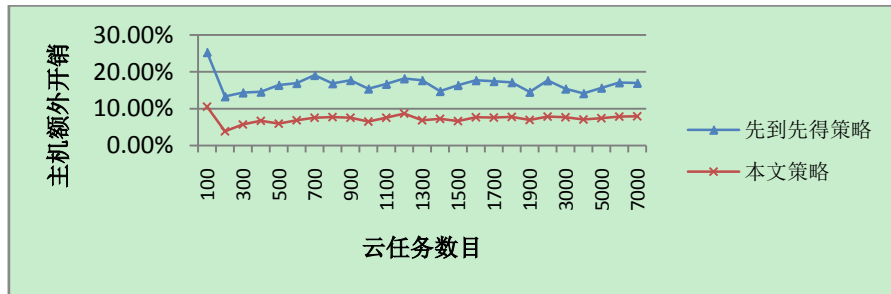


Figure 2. The contrast of additional host-spending between our strategy and FCFS  
图 2. 先到先得策略与本文策略下产生的额外主机开销对比

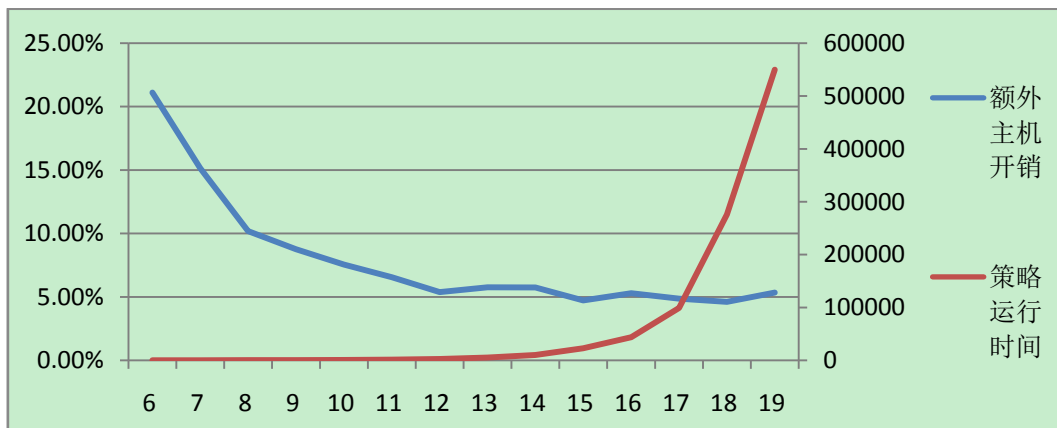


Figure 3. The contrast of additional host-spending and running time among the different number cloud tasks  
图 3. 每组所含云任务数目不同情况下的额外主机开销与运行时间对比

出了一种高效率的虚拟资源分配策略, 通过云中间商“端”对云用户虚拟任务进行统计, 资源管理器对云任务进行分组, 找到每一组中新开启主机的最优任务配比, 以使主机服务器得到最大利用。模拟仿真了不同条件来测试策略的性能。结果表明, 本文策略比先到先分配策略有优势, 并且本文策略在云任务数量增多以及分组数目的变化的情况下均有一定的稳定性, 能成功提高主机服务器的利用率。

## 基金项目

本文得到国家自然科学基金项目(No. 61272093), 山东省高等学校科技计划项目(No. J15LN19, No. J13LN18), 青岛市博士后研究人员应用研究项目(云计算在矿山地质数据存储及分析中的应用)的支持。

## 参考文献 (References)

- [1] Ran, J.H., Li, S.Q., Liu, Y., *et al.* (2011) Study on Management and Dispatching of Virtual Computing Resources Based on Pool Theory. 2011 *7th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM)*, Wuhan, 23-25 September 2011, 1-7.
- [2] Gao, Y., Guan, H., Qi, Z., *et al.* (2013) A Multi-Objective Ant Colony System Algorithm for Virtual Machine Placement in Cloud Computing. *Journal of Computer and System Sciences*, **79**, 1230-1242. <http://dx.doi.org/10.1016/j.jcss.2013.02.004>
- [3] Hassan, M.M. and Alamri, A. (2014) Virtual Machine Resource Allocation for Multimedia Cloud: A Nash Bargaining Approach. *Procedia Computer Science*, **34**, 571-576. <http://dx.doi.org/10.1016/j.procs.2014.07.074>
- [4] You, X., Xu, X., Wan, J., *et al.* (2009) Ras-m: Resource Allocation Strategy Based on Market Mechanism in Cloud Computing. *ChinaGrid Annual Conference, ChinaGrid'09*, Yantai, 21-22 August 2009, 256-263.
- [5] Deng L, Yao L. (2015) Dynamic Allocation of Virtual Resources Based on Genetic Algorithm in the Cloud. *Advances in Services Computing*, **9464**, 153-164.
- [6] Chen, Z.G., Du, K.J., Zhan, Z., *et al.* (2015) Deadline Constrained Cloud Computing Resources Scheduling for Cost Optimization Based on Dynamic Objective Genetic Algorithm. 2015 *IEEE Congress on IEEE Evolutionary Computation (CEC)*, Sendai, 25-28 May 2015, 708-714.
- [7] Huang, D., Du, P., Zhu, C., *et al.* (2015) Multi-Resource Packing for Job Scheduling in Virtual Machine Based Cloud Environment. 2015 *IEEE Symposium on Service-Oriented System Engineering (SOSE)*, San Francisco Bay, 30 March-3 April 2015, 216-221.
- [8] Choi, J., Adufu, T., Kim, Y., *et al.* (2015) A Job Dispatch Optimization Method on Cluster and Cloud for Large-scale High-Throughput Computing Service. 2015 *International Conference on Cloud and Autonomic Computing (ICAC)*, Boston, 21-25 September 2015, 283-290.
- [9] Liaqat, M., Ninoriya, S., Shuja, J., *et al.* (2016) Virtual Machine Migration Enabled Cloud Resource Management: A Challenging Task. arXiv preprint arXiv:1601.03854
- [10] Zhou, A., Sun, Q., Sun, L., *et al.* (2015) Maximizing the Profits of Cloud Service Providers via Dynamic Virtual Resource Renting Approach. *EURASIP Journal on Wireless Communications and Networking*, **2015**, 1-12. <http://dx.doi.org/10.1186/s13638-015-0256-y>
- [11] 周山杰. 云计算环境下面向任务分类的个性虚拟化策略[D]: [硕士学位论文]. 沈阳: 辽宁大学, 2012.