

基于 Qos 综合满意度的云计算任务调度策略

葛君伟,王清玲,方义秋

(重庆邮电大学 计算机科学与技术学院,重庆 400065)

摘 要: 如何提高调度效率成为云计算商业化发展必须要考虑的因素. 作者提出一种综合考虑用户和服务商满意度的任务调度策略,能够有效的提高用户满意度,降低任务成本,提高系统资源利用率. 通过 Cloudsim 实验表明,提出的算法总体上优于传统的 Min-Min 以及顺序执行算法.

关键词: 云计算;任务调度;服务质量;满意度;成本;负载均衡

中图分类号: TP31

文献标识码: A

文章编号: 1000—7180(2016)10—0111—04

A Satisfaction-Based Task Scheduling in Cloud Computing

GE Jun-wei, WANG Qing-ling, FANG Yi-qiu

(School of Computer Science and Technology, Chongqing University of Posts and
Telecommunications, Chongqing 400065, China)

Abstract: How to complete task scheduling efficiently, improve the satisfaction of customers and service providers, has become a major factor for the development of cloud computing commercialization. The author has put forward a task scheduling strategy which is considering the satisfaction of customers and service providers. By taking this strategy, we can increase the satisfaction of customers, decrease the coast of tasks, and improve the efficiency of system sources. The result of the simulation platform shows that the proposed strategy is generally superior to traditional min-min algorithm and sequential execution algorithm.

Key words: cloud computing; task scheduling; service quality; satisfaction; cost; Load balance

1 引言

近年来,云计算任务调度问题吸引了各行各业的人对其进行研究^[1,2]. 针对云计算环境下的任务调度通常是采用一定的调度策略为任务绑定合适的 vm 资源. 文献[3]从服务商角度出发,追求任务成本最小化,但未充分考虑系统负载,造成资源的浪费,从而影响服务商的收益. 文献[4]提出一种基于负载均衡的任务调度算法,通过增设服务器最佳期望序列及负载均衡指数,采用任务连接数约束服务器选择,使系统达到较好的负载均衡度,同时缩短任务完成时间,从而达到了优化的目的,但是未考虑任务的成本问题,造成任务的执行成本过高. 文献[5]提出一种新型的云环境下 Qos-aware 服务质量感知

的任务调度算法,该算法结合贪心算法的思想,充分考虑用户的满意度,但未考虑服务提供商的成本,不符合云计算持久发展的要求.

云计算的商业特性使得其经济因素成为要重点考虑的调度指标. 只有满足市场经济的需求,云计算才能长久发展^[6,7]. 因此,必须综合考虑用户 Qos 要求以及服务商的服务收益,其任务调度策略必须确保用户的 Qos 尽量得到满足,服务提供商满意,这样才符合商业发展的基本要求. 基于现有算法的局限性,本文从云计算持久发展的角度出发,综合考虑用户和服务商的需求,提出了一种综合考虑用户满意度和服务商满意度的任务调度策略. 通过 Cloudsim 平台验证,该算法能较好地提高用户满意度,降低任务执行成本,提高了资源的利用率.

2 基于云计算的任务调度模型

基于 Qos 的任务调度就是根据云计算系统中已有的计算机资源集合、用户提交的任务集合、某些调度规则以及用户的 Qos 要求,将任务分配在合适的虚拟机资源上的任务执行等待队列中,等待被执行的过程.云计算的调度目标是对用户提交的任务实现最优调度.大多数任务调度算法是通过虚拟机级别上内置的调度策略进行,使用最优跨度、服务质量、负载均衡和经济原则等来衡量任务调度的效果.

为了简化调度过程的复杂性,建立一个更加有效的任务调度模型,本文作如下假设:

- (1)任务之间彼此独立运行,不存在相互依赖关系;
- (2)任务在虚拟机中的执行时间以 CPU 执行效率度量;
- (3)任务一旦执行,不能被别的任务中断;
- (4)任务处于等待状态也要耗费成本,可理解为系统维护费用.

整个调度过程模型如图 1 所示.

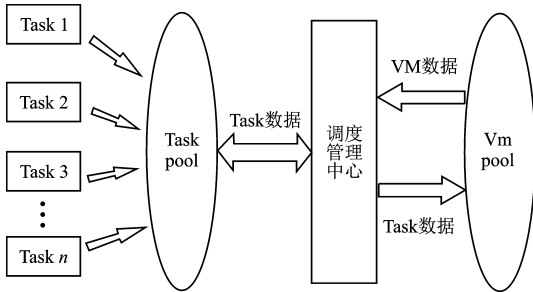


图 1 任务调度基本框架图

2.1 任务及虚拟机的基本定义如下:

定义 1 m 个任务集 $\text{Task} = \{T_1, T_2, T_3, \dots, T_m\}$ 合中, T_i 表示集合中第 i 个任务,它的子属性用一个四元组 $(\text{Id}, \text{Length}, E_{\text{price}}, E_{\text{time}})$,其中 Id 是任务的唯一的标识号; Length 表示任务的长度; E_{price} 表示用户期待任务完成的费用; E_{time} 表示任务的期待完成的时间.

定义 2 在云系统中, n 个虚拟机的集合表示为 $\text{VM} = \{\text{VM}_1, \text{VM}_2, \text{VM}_3, \dots, \text{VM}_n\}$, VM_j 表示集合中第 j 个虚拟机. 它的性能参数集可以用一个五元组 $(\text{Id}, \text{Cpu}, \text{Bw}, \text{Mm}, \text{Hd})$ 表示,其中 Id 是任务的唯一的标识号; Cpu 表示虚拟机的任务执行能力; Bw 表示虚拟机的带宽; Mm 表示虚拟机的内存大小; Hd 表示虚拟机的硬盘大小.

2.2 任务调度过程中需要的参数描述

2.2.1 任务的 Qos 参数设置

云计算的目的是为海量的用户提供不同的服务,用户申请的任务的类型各种各样,如有的用户偏好费用低的服务,有的用户偏好质量好的服务等. Qos 可用作衡量服务满意度的一个重要标准. 本文按照 Qos 对任务进行分类,目前仅考虑这两个参数,也可扩展为多个参数.

(1)完成时间:主要对时间要求度高,追求较短的完成时间的用户常会选择此参数.

(2)费用:云计算作为一种商业计算模式,按需付费是它的主要特点,因此对费用的考虑也是必要的.

2.2.2 任务与虚拟机匹配中需要考虑的参数

定义 3 任务 i 在虚拟机 j 上的总完成时间定义如下:

$$\text{Time}_{ij} = \text{Time}_{\text{wait}_{ij}} + \text{Time}_{\text{run}_{ij}} \quad (1)$$

由于通信时间相比等待和执行时间过小,本文中暂不做考虑.

定义 4 任务 i 在虚拟机 j 上的成本定义如下:

$$\text{Cost}_{ij} = \text{Cost}_{\text{run}_{ij}} + \text{Cost}_{\text{wait}_{ij}} \quad (2)$$

$$\text{Cost}_{\text{run}_{ij}} = (P_{\text{Cpu}_j} + P_{\text{Bw}_j} + P_{\text{Mm}_j}) * T_{r_{ij}} \quad (3)$$

$$\text{Cost}_{\text{wait}_{ij}} = \text{Length}_i * \text{Time}_{\text{wait}_{ij}} \quad (4)$$

式中, P_{Cpu_j} , P_{Bw_j} , P_{Mm_j} 分别表示主频 (Cpu)、带宽 (Bw)、内存 (Mm) 的单位时间内虚拟机各个性能指标运行所耗费的成本, $T_{r_{ij}}$ 表示任务 i 在虚拟机 j 中的执行时间.

定义 5 用户的损耗满意度函数定义如下:

$$\text{Uqos}_{ij} = \text{qos}_{t_{ij}} + \text{qos}_{p_{ij}} \quad (5)$$

$$\text{qos}_{p_{ij}} = \begin{cases} \sqrt{(f(ct_{ij}) - E_{P_i})^2}, & f(ct_{ij}) \geq E_{P_i} \\ 0, & f(ct_{ij}) < E_{P_i} \end{cases} \quad (6)$$

$$\text{qos}_{t_{ij}} = \begin{cases} \sqrt{(f(T_{ij}) - E_{T_i})^2}, & f(T_{ij}) \geq E_{T_i} \\ 0, & f(T_{ij}) < E_{T_i} \end{cases} \quad (7)$$

式中, $\text{qos}_{t_{ij}}$, $\text{qos}_{p_{ij}}$ 表示任务的时间、费用的满意损耗度,用户的满意度随着损耗满意度函数的值的增大而减小.

$f(ct_{ij})$ 表示在虚拟机 j 上执行用户任务 i 所需要耗费的成本的参数归一化后的值, E_{P_i} 表示用户提供的任务 i 预期付出的费用. 同理, $f(T_{ij})$ 表示虚拟机 j 完成用户任务 i 需要的时间, E_{T_i} 表示用户提供的任务 i 预期完成时间,参数归一化方法如下.

定义 6 服务提供商的满意度函数主要由任务执行的单位成本提供的用户满意度决定,定义如下:

$$Dqos_{ij} = Eqos_{ij} / (Cost_{ij} * Time_{ij}) \tag{8}$$

式中, $Eqos_{ij}$ 表示用户对任务的满意度,本文中主要使用两个 Qos 指标,用户任务的预期满意度之和为 1. $Cost_{ij}$ 表示在虚拟机 j 上执行任务 i 所需要的成本,包含执行成本和等待成本. $Time_{ij}$ 表示虚拟机 j 上执行任务 i 所需要的完成时间.

定义 7 对于虚拟机的负载的定义为

$$Load_{ij} = Time_{wait_{ij}} / Load_j \tag{9}$$

在 vm 的等待序列中,排在该任务之前的任务的总的完成时间与 vm_j 的最大负载的比值, vm_j 的最大负载表示为

$$Load_{max_j} = Mm_j * HD_j / Cpu_j \tag{10}$$

虚拟机的负载对于云计算系统是一个重要的性能指标,系统负载不均衡或者过重,都将影响任务的执行效率以及资源的利用率.

2.3 任务优先级模型

任务的优先级决定了同时提交的一批任务中的任务调度顺序,高优先级的任务将优先进行任务调度. 本文重新定义了任务的优先级,用来提高任务的满意度.

针对同时到达的一批任务,如果用户的需求,在不同的 vm 上的满意度差别较大,则优先调度此类任务,以便提高整批任务的满意度.

3 参数归一化方法描述

本文通过将成本和时间归一到[0,1] 区间,与用户的期望值进行统一比较,确定用户的满意损耗度.

具体方法如下:

(1) 对于参数 x 来讲,当 $x_{max} \neq x_{min}$ 时:

$$f(x) = (x_i - x_{avg}) / (x_{max} - x_{min}) \tag{11}$$

(2) 当 $x_{max} = x_{min}$:

若 $x_{max} = 1$,则 $f(x)$ 的值为 1;若 $x_{max} = 0$,则 $f(x)$ 的值为 0.

4 基于 Qos 满意度的任务调度算法(UCD)描述

算法步骤如下所示.

Step1:输入:任务以及虚拟机 vm 相关信息.

目标:输出 task-vm 匹配组合序列.

Step2:根据用户的期望 Qos 序列来计算任务预分配到各个 vm 上的用户满意度损耗值,并计算各

个任务的最大满意损耗度和最小满意损耗度.

Step3:计算任务的最大满意损耗度和最小满意损耗度的差值,并以此为标准对任务进行排序,优先调度差值较大的,以便于取得更好的批任务满意度.

Step4:以任务在 vm 上的用户满意损耗度为标准,对 vm 进行顺序排序(从小到大). 若存在多个拥有最小损耗度的 vm,则对这些 vm 按执行成本顺序排序,从最终的 vm 队列中取出队首 vm,记作 V_{user} .

Step5:以单位成本时间所获取的用户满意度作为服务商对任务的满意度,计算每个任务分配到 vm 上的服务商满意度. 以服务商满意度为标准,对 vm 进行逆序排序(从大到小),若是存在多个相同服务商满意度的 vm,则对这些 vm 按获取的用户满意度逆序排序,从最终的 vm 队列中取出队首 vm,记作 V_{dealer} .

Step6:对从 Step4 和 Step5 中取出的 vm 进行参数对比. 若两步骤中取得的为同一台 vm,则任务效益最大,直接绑定即可. 若不相同,则比较两者的负载. 若 V_{user} 的负载大于 V_{dealer} ,则将任务绑在 V_{dealer} 虚拟机上,反之则绑定在 V_{user} 上.

5 实验仿真与分析

为验证基于综合满意度的云计算任务调度策略的性能,本文在开源的 Cloudsim 云仿真平台下对该策略进行仿真实验,并将实验结果与传统的顺序执行算法以及 Min-Min 算法分别进行比较.

5.1 实验环境与参数设置

实验环境参数如下.

(1) 任务的数量范围为[0,250],长度为[1, 40 000],期望的 Qos 序列为[Q_{Price}, Q_{Time}],其中 Q_{Price} , Q_{Time} 的值的范围均为[0,1],且 $Q_{Price} + Q_{Time} = 1$.

(2) 实验涉及的 vm 参数如表 1 所示.

表 1 虚拟机参数列表

VMid	Cpu	Bw	HD	MM
1	1 024	800	800 000	1024
2	512	1 000	160 000	1 024
3	1 024	1 200	320 000	2 048
4	1 024	700	600 000	512
5	512	500	100 000	1 024

5.2 仿真结果分析

在 Cloudsim 环境,通过设置相同的环境参数,对比 Min-Min、顺序执行算法以及 UCD 算法,实验结果如图 2~5 所示.

由图 2 可知:在任务数量较少的场景中,就系统平均负载而言,算法的优越性不太明显. 主要原因是

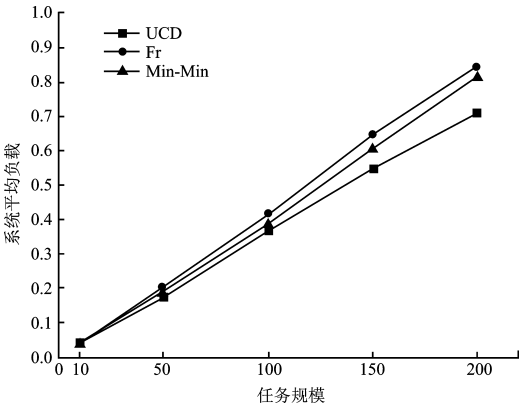


图 2 系统平均负载

处于等待序列中的任务较少,系统负载差别不大,但是算法效果还是优于其他两种算法,随着任务数量的增多,由于 UCD 算法考虑了任务的等待时间以及等待费用,考虑了虚拟机的负载情况,在任务匹配阶段,比其他两种算法更具有优越性.

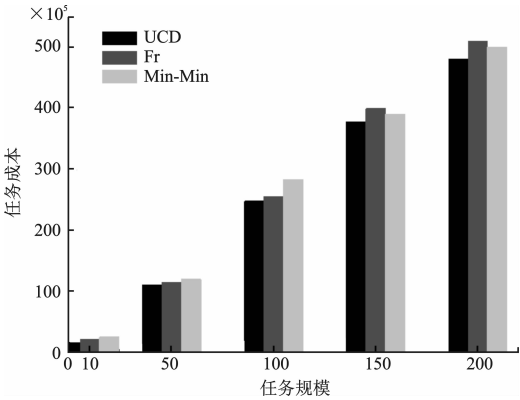


图 3 任务平均成本

由图 3 可知,在任务的平均成本方面,UCD 算法比其他两种算法有较好的实用性.

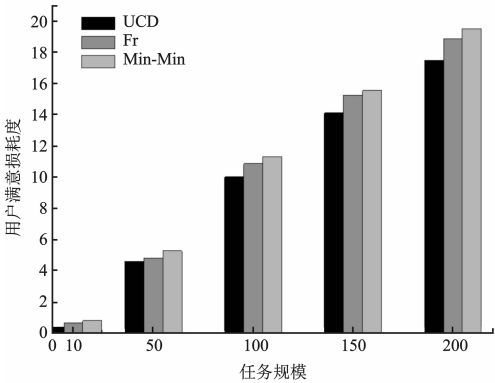


图 4 用户满意损耗度

由图 4 可知,在用户满意度方面,UCD 算法比其他两种算法有较好的实用性,损耗度越低,用户的

满意度越好.

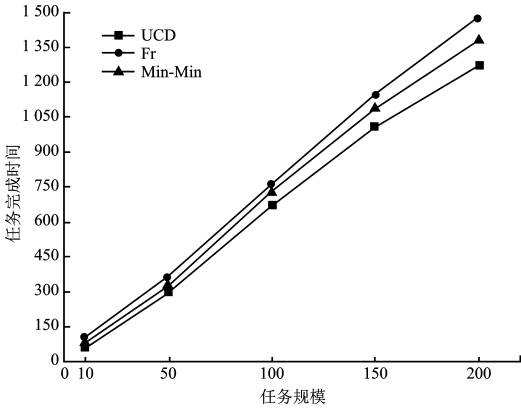


图 5 任务完成时间

图 2~5 分别代表 UCD 算法、Min-Min、顺序执行算法(Fr)在相同的实验条件下的实验结果. 从实验结果上可以看出 UCD 算法在任务完成时间、用户满意损耗度,以及任务完成成本、系统平均负载上均优于 Min-Min 算法、顺序执行算法(先来先服务算法). 由于 UCD 算法考虑了任务的用户满意损耗度,以及服务商的满意度(单位成本收益的用户满意度),所以在满意损耗度、任务成本方面,UCD 优于 Min-Min 算法和顺序执行算法. 而且由于考虑了任务的等待成本(同一任务等待时间越长,等待成本越大)以及虚拟机负载情况,UCD 算法在任务完成时间以及系统平均负载方面,比 Min-Min 算法、顺序执行算法有较好的效果.

6 结束语

针对具有商业特征的云任务调度,本文提出了一种基于 Qos 综合满意度的任务调度策略,在此策略下提出了一种具有任务调度算法,即 UCD 算法,最终的实验结果表明,该算法可以提高调度的效率以及用户及服务提供商的满意度.

在将来的研究工作中,将继续研究适合云计算商业化发展特性的任务调度策略,针对系统负载过重情况下,如何提高算法的效率进一步做研究.

参考文献:

[1] 李乔, 郑啸. 云计算研究现状综述[J]. 计算机科学, 2011, 7(4):32-37.
[2] 梁桂才. 云计算环境下任务调度的趋势分析[C]// 2014 年全国开放式分布与并行计算学术年会论文集. 南京:计算机学会,2014.

ADPSO 粒子群算法做了对比,仿真的实验结果表明,提出的以度为规则的小世界粒子群算法能够避免粒子种群的聚集,出现早熟现象,维持粒子种群的多样性.在以后的学习中,应该把小世界模型应用于更多拓扑结构中,比如四类结构、金字塔结构等等,在这些模型中应用小世界的思想来优化粒子群算法.

参考文献:

- [1] 王伟,李枚毅,彭霞丹.一种双层可变子群的动态粒子群优化算法[J].小型微型计算机系统,2012,33(1):145-150.
- [2] 黄太安,生佳根,徐红洋,等.一种改进的简化粒子群算法[J].计算机仿真,2013,30(2):327-330.
- [3] 石松,陈云.层次环形拓扑结构的动态粒子群算法[J].计算机工程与应用,2013,49(8):1-5.
- [4] Watts DJ, Strogatz S. Collective dynamics of small-world networks[J]. Nature, 1998, 393(4):440-442.
- [5] Newman MEJ, Watts DJ. Renormalization group analysis of the small-world network model[J]. Phys Lett A, 1999, 263(4):341-346.
- [6] Ming Li, Wenqiang Du, Fuzhong Nian. An adaptive

particle swarm optimization algorithm based on directed weighted complex network[J]. Mathematical Problems in Engineering, 2014, 2014(4):1-7.

- [7] Guangquan Xu, Chao Xu. PSO-TPS: An optimal trust path selection algorithm based on particle swarm optimization in small world network[C]// Proceedings - 2nd International Conference on Cloud and Green Computing and 2nd International Conference on Social Computing and Its Applications, Xiangtan: IEEE, 2012:594-600.
- [8] 龚月娇,稽智源.自适应小世界粒子群优化算法[J].计算机工程与设计,2015,36(6):1598-1606.

作者简介:

靳雁霞 女,(1973-),博士,副教授.研究方向为智能优化算法、虚拟现实技术.

银莉(通讯作者) 女,(1991-),硕士研究生.研究方向为智能优化算法. E-mail:826082049@qq.com.

张晓闻 女,(1990-),硕士研究生.研究方向为智能优化算法.

张鑫 男,(1994-),硕士研究生.研究方向为智能优化算法.

(上接第 114 页)

- [3] 吴高峰,蒋玉明,杨林,等.基于 QoS 改进的 Min-Min 网络调度算法[J].微计算机信息,2009,25(27):110-112.
- [4] Venters W, Whitley E A. A critical review of cloud computing: researching desires and realities[J]. Journal of Information Technology, 2012, 27(3): 179-197.
- [5] 邓见光,赵跃龙,袁华强.一种多 QoS 目标约束的云计算任务调度策略[J].计算机应用研究,2016,33(8):1-6.
- [6] Pallis G. Cloud computing: the new frontier of internet computing[J]. IEEE Internet Computing, 2010, 14(5): 70-73.
- [7] YAN G, YU J, YANG X. Reliability-aware workflow scheduling strategy on cloud computing platform[J]. Journal of Computing Application, 2014, 34(3): 673-677.
- [8] 邓见光,赵跃龙,袁华强,等.一种成本驱动的云计算任务调度策略[J].江苏大学学报(自然科学版),2014,35(2):214-219.
- [9] 张腊,刘淑芬,韩璐.基于负载均衡的任务调度算法[J].吉林大学学报(理学版),2014,52(4):769-772.
- [10] 周发超,王志坚,叶枫.一种新型的云任务调度算法

研究[J].中国科学技术大学学报,2014,44(7):590-598.

- [11] Lai Li Yuan Jun, Tao Fei, et al. A study of optimal allocation of computing resources in cloud manufacturing systems [J]. International Journal of Advanced Manufacturing Technology, 2012, 63(5-8): 671-690.
- [12] Hirai T, Masuyama H, Kasahara S, et al. Performance analysis of large-scale parallel-distributed processing with backup tasks for cloud computing [J]. Journal of Industrial and Management Optimization, 2014, 10(1): 113-129.
- [13] edersen J M, Riaz M T, Dubalski B, et al. Using latency as a QoS indicator for a global cloud computing services [J]. Concurrency and Computation Practice and Experience, 2013, 25(18): 2488-2500.

作者简介:

葛君伟 男,(1961-),博士,教授.研究方向为云计算以及软件体系架构. E-mail:gejw@cqupt.edu.cn.

王清玲 女,(1990-),硕士.研究方向为云计算.

方义秋 女,(1963-),博士,副教授.研究方向为云计算和软件架构.