

云计算环境下基于效益最优的资源调度策略

熊安萍¹, 黄 凯¹, 杨 川²

(1 重庆邮电大学 计算机科学与技术学院, 重庆 400065; 2 中国电信股份有限公司泸州分公司, 四川 泸州 646000)

摘 要: 针对云计算环境下资源调度存在调配不灵活, 资源不能合理按需配置、负载不均衡以及资源利用率低等问题, 提出了一种基于效益最优的资源调度策略, 实验结果表明, 该策略能够有效提高系统吞吐量、减少任务周转时间, 提高资源利用率, 使用户能够在云计算环境下获得更大的计算经济效益。

关键词: 云计算; 资源调度; 效益最优; 资源利用率; 经济效益

中图分类号: TP301.6

文献标识码: A

文章编号: 1000-7180(2013)11-0104-04

Resource Scheduling Strategy for Cloud Computing Environment Based on Optimal Benefits

XIONG An-ping¹, HUANG Kai¹, YANG Chuan²

(1 College of Computer Science and Technology, Chongqing University of Post and Telecommunications, Chongqing 400065, China; 2 China Telecom Corporation Ltd Luzhou Branch, Luzhou 646000, China)

Abstract: Inflexible resources scheduled deployment in the cloud computing environment, resources cannot reasonably be unbalanced, load on demand and low resource utilization, presents a resource scheduling strategy based on efficiency optimization, experimental results show that the policies to improve System throughput and resource utilization, while reduce tasks turnaround time, enabling users to achieve greater computing economic benefits in the context of cloud computing.

Key words: cloud computing; resource scheduling; benefits of optimal; resource utilization; economic benefits

1 引言

云计算是一种商业计算模型, 它将计算任务分布在大量计算机构成的资源池上, 使用户能够按需获取计算能力、存储空间和信息服务, 是并行计算、网格计算、效用计算、虚拟化技术的发展^[1]。

从经济学的角度来看, 云计算实质上是一种生产者—消费者模型, 服务被看作是有价值的经济商品, 供应商生产商品并将其分租给消费者, 云用户可以在全球范围内根据自己的需求按照一定的付费方式从供应商中购买商品^[2]。云计算的核心思想, 是将大量用网络连接起来的资源进行统一管理, 构成一个集中的资源池, 向用户提供按需式服务^[3]。云计算的根本目的是共享资源与服务, 资源的有效管理是评价资源调度是否高效的关键, 云计算环境

下, 传统的资源调度算法, 虽然能够满足用户的需求, 但有时会造成资源的浪费和资源的服务中断, 且资源利用率不高。

本文从经济学的角度^[2]研究云资源的管理和调度, 结合贪婪算法和效益函数, 提出了基于效益最优的资源调度策略, 并进行了实验验证及分析。

2 云计算环境下的资源管理模型

2.1 云计算资源的经济学模型

云计算和经济社会有很多的共性^[2]。云计算是通过集成各种资源以提供强大的计算能力, 经济社会则是通过对社会资源的优化配置从而促进更多的财富创造; 云计算资源是随着需求动态变化的, 市场中的商品交易也是随时波动的; 云计算资源调度需要解决资源的负载均衡^[5], 经济活动也要保持商品

的供求平衡;从某种意义上来说,云计算就是社会,云计算面临的所有难题,在经济社会中也同样需要解决.市场机制作为解决社会资源配置的重要手段,已被证明是非常成功的.因此,在云计算资源管理的研究中引入市场机制,具有很强的现实意义.

云计算资源的经济学模型如图1所示.主要由以下几部分组成:云用户、任务代理、云计算市场、物理资源、虚拟资源、资源代理.

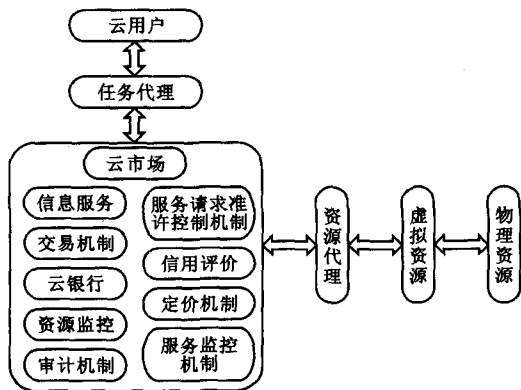


图1 云计算资源的市场经济模型

图1中所示的云用户即云计算资源的使用者,云用户从“端”通过其经纪人向云市场提出自己的资源需求和花费限制,包括所需完成任务的QoS描述,如CPU类型、CPU数目、内存大小、操作系统及其版本号、成本预算、可靠性、完成时间等.

任务代理代表用户将服务请求提交到将要被处理的数据中心和云中,查找和选择适合云计算应用需求的资源.任务代理支持多种不同的调度方案以满足用户的不同目标.

云市场资源以交易的方式实现云计算资源供需双方的匹配,并通过市场中的各种机制交互来支持云资源管理.交易机制即是市场经济模型中的资源调度器,是实现市场化配置资源的核心构件.

虚拟资源是通过虚拟化技术,在物理资源基础上构建的虚拟资源,每个虚拟机主机都可以有各自的操作系统及应用环境,用户可以快速地在虚拟资源池获得所需的应用环境,而不用考虑资源购置、维护、环境部署等问题^[4].

2.2 基于市场经济模型的资源调度算法

资源调度的主要任务就是根据任务信息采用恰当的策略把不同的任务分配到相应的资源上. Deadline and Budget Constrained Cost-Time Optimisation 算法^[6] (DBC)是由IEEE的集群计算主席、澳大利亚墨尔本 Monash 大学的 Rajkumar Buyya 博士提出的一种基于市场经济模型的算法,

支持用户的QoS的需求,用户在提交自己的任务时可以设置自己的要求,最高预算 Budget 和最晚期限 Deadline. 具体包括代价最优调度、时间最优调度、代价一时间均衡调度等.

然而,上述三种算法都是在 DBC 约束下,满足时间或代价单方面的 QoS 需求的极端情况,并没有考虑系统资源的均衡使用,有可能经常选择时间或费用最少的资源而造成资源的利用率低下.在实际应用中,计算资源往往具有动态性和多样性,大多数情况下,用户希望在时间和费用有限的情况下尽可能多地完成作业,即能够使用户获得更大的效益.

3 效益最优的资源调度策略

由于用户对系统资源有着不同的要求,有的用户要求尽可能在最短的时间内完成任务,优先选择性能高的资源,造成某些性能低的资源得不到使用;有的用户则要求费用尽量少,就会选择性能低的资源执行任务,以致于某些性能高的资源闲置,这两种情况都会造成资源负载的不均衡.任务可能在执行过程中会突然被终止,有的资源中的任务队列里也有很多的任务处于就绪状态,造成资源的负载过重,甚至还有的任务不能得到及时的执行,导致资源利用率低,执行效率也很低.因此,本文结合改进的贪婪算法和效益函数,提出了基于效益最优的云资源调度算法,实现了资源的负载均衡,提高吞吐量和资源利用率,使用户能够获得最大的效益.

3.1 基于效益最优的资源调度算法设计

我们从用户的需求出发,通过分别定义一个关于时间和费用^[6]的效益函数,以用户满意度最高为最优的资源调度目标.设 U 为用户的满意度, $U(C_i)$ 表示价格效益函数, $U(T_i)$ 为时间效益函数,最后得到综合效益,综合效益越高,用户的满意度也越高.

$$U(T_i) = (D - \sum_{i=1}^n M/N) / D \quad (1)$$

式中, $D(s)$ 为用户的最大完成时间, M (单位为 MI) 表示的是任务的指令条数, N (单位为 MIPS) 表示虚拟机每秒执行的百万指令数;

$$U(C_i) = (B - \sum_{i=1}^n P * (1 + N/\bar{N})) / B \quad (2)$$

式中, B (元) 表示用户的预算, P (元) 代表资源的基础价格, $\bar{N}$ 是所有资源的平均计算速度;

$$U(T_i, C_i) = \sum_{i=1}^n (\alpha U(T_i) + \beta U(C_i)) \quad (3)$$

式中, α 、 β 分别代表时间和费用效用的相对重要性,

根据用户的任务的不同偏好取值, $0 < \alpha, \beta < 1$, 并且 $\alpha + \beta = 1$; $i = 1, 2, 3, \dots, n$, n 为任务总数.

$$A = D/T_1$$

$$B = B/C_1$$

$$\alpha = A/A+B$$

$$\beta = B/A+B \quad (4)$$

式中, T_1 代表用户提交任务集的运行时间估计值(取所有任务的总长度除以资源平均处理能力的值与最长任务的长度除以资源的最快处理能力的值之间的最大值), C_1 代表提交任务集的运行费用估计值(取任务的总长度除以资源的平均价格).

3.2 效益最优调度算法

通过以上设计, 算法的实现步骤如下:

(1) 将任务按照优先级进行排序, 再按照 QoS 分类, 这里是按照任务对时间和费用的偏好进行分类, 以虚拟机资源计算能力的平均值为阈值, 性能好的, 用时短, 费用高; 性能差的, 用时长, 费用少. 然后将同类任务中优先级相同的按照 M 的大小进行降序排序, 再将虚拟机资源按照 N 的大小升序排列, 定义一个时间矩阵和费用矩阵, $\text{time}[i][j]$ 和 $c[i][j]$, 表示任务 i 在资源 j 上的执行时间和所需费用, 显然, $\text{time}[i][j] = M_i/N_j$, $c[i][j] = P * (1 + N_i/\bar{N})$.

(2) 初始化时, 将资源都设置为新资源, 新资源即虚拟机上没有任务, 相反, 旧资源就是虚拟机上有任务在运行.

(3) 根据任务的偏好选择资源队列(将性能好的和性能差的资源各放入一个队列中), 首先考虑优先级高的任务, 然后再考虑其他任务. 优先选择新资源, 如果没有新资源, 就将其分配给旧资源, 然后求出任务 i 在资源 j 上的效益值, 最后将任务分配给能够使其获得最大效益的资源, 如果有多种方法能够使当前结果最优, 那么就将任务分配给运行任务最少的虚拟机.

(4) 在任务的执行过程中, 如果任务被终止执行, 则重新进行分配.

(5) 在资源组合时, 计算其时间效益和费用效益, 得到将这个任务分配到该资源的效益值, 统计各个任务的效益值, 计算此用户的综合效益, 即用户的满意度.

以上步骤的流程如图 2 所示.

4 实验与结果分析

为了验证本文基于效益最优的云资源调度算法

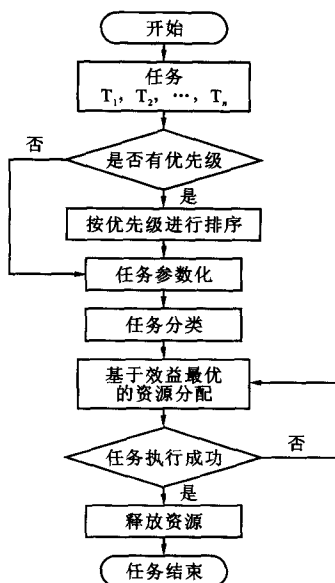


图 2 算法流程图

的性能及可行性, 本文选择 CloudSim 仿真工具对实验进行验证.

CloudSim 是提供基于数据中心的虚拟化技术、虚拟化云的建模和仿真功能, 支持云计算的资源管理和调度模拟的云计算仿真平台. 本文在 CloudSim 的基础上进行二次开发, 建立实验仿真环境. 在主机上使用 Windows, jdk 6.0-13, 将 CloudSim 2.1.1 版解压并完成 CloudSim 的配置. 这里考虑的是单用户、独立任务的情况.

实验一: 初始化 Grid-Sire 库, 创建数据中心, 创建代理 Broker, 创建虚拟机和云任务. 实验中搭建了 10 台虚拟机, 根据虚拟机实际配置分别设定各虚拟机的处理能力, 分别为 278, 289, 132, 209, 386, 450, 500, 550, 700, 1 000, 单位为 MIPS, 得出资源计算速度的平均值 $\bar{N}$. 这里设用户提交相同的计算任务, 数量分别为 20, 50, 100, 200, 500, 1 000, 时间和费用的预算分别为 10 000 ms 和 10 000 元, 比较效益最优、时间最优及时间代价最优分别在预算时间和费用内完成的任务数量, 即吞吐量.

实验一结果如图 3 所示. 基于效益最优的资源调度算法比时间最优及时间代价最优调度算法在相同时间内完成的总任务数要多, 在任务量少的时候, 三种算法之间的差距很小, 随着任务数越来越多, 几种算法之间的差距就会越来越明显. 因此, 在相同时间和预算的情况下, 基于效益最优的资源调度算法比其他两种算法具有更大的吞吐量, 且能够达到较好的负载均衡.

实验二: 保持虚拟机资源的数量不变, 任务数分

别为 20, 50, 100, 200, 500, 1 000, 比较通过效益最优、时间最优及时间代价最优调度算法完成任务所需要的时间, 结果如表 1 所示。

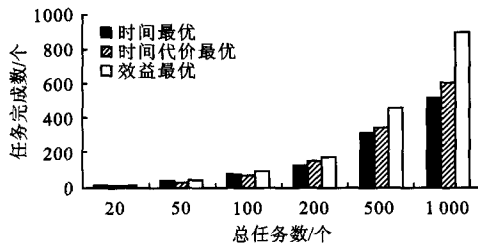


图3 任务完成数比较

表1 完成任务所需时间 ms

算法 \ 任务	20	50	100	200	500	1 000
时间最优	135.64	308.47	716.67	2 643.31	5 575.97	8 675.97
时间代价最优	167.28	402.87	798.70	3 104.33	6 101.04	8 991.04
效益最优	148.32	347.61	677.59	1 340.47	3 313.41	5 202.84

根据表 1 的结果, 得到三种策略下所有任务完成所需的时间(任务周转时间)比较, 如图 4 所示。相同任务数情况下, 在任务数较少的时候, 时间最优的调度算法相比于效益最优和时间代价最优调度算法而言, 总的周转时间最少, 但是随着任务数的增多, 基于效益最优的资源调度算法所用的时间开销优势愈加明显, 具有最少的总的任务周转时间。因此, 在相同时间和预算的限制下, 效益最优算法能提高云计算环境下资源的整体利用率。

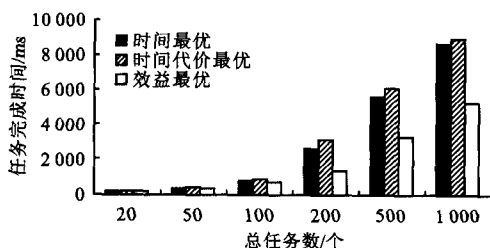


图4 任务完成周转时间比较

根据实验结果可以看出, 本文的效益最优算法能将任务合理地分配到各个资源上, 不仅能够获得较高的吞吐量, 使得资源负载更为均衡, 且有效地减少了任务周转时间。

5 结束语

本文从经济学的角度研究云资源的管理和调度, 提出了基于效益最优的资源调度策略, 能够得到较高的系统吞吐量以及较好的资源负载均衡性, 并

且有效地缩短了任务周转时间, 提高了用户满意度、资源利用率和资源调度的效率, 在基于市场经济的云环境中, 具有较高的经济效益。

参考文献:

- [1] Ian Foster, Yong Zhao, Ioan Raicu, et al. Cloud computing and grid computing 360-degree compared[C]// Grid Computing Environments Workshop. Austin, TX: IEEE, 2008: 1-10.
- [2] Grid Computing Info Center. Economy grid[EB/OL]. (2005-09-01)[2013-02-22]. <http://www.buyya.com/ecogrid/>.
- [3] Forster I, Kesselman C, Tuecke S. The anatomy of the grid: enabling scalable virtual organizations[J]. International Journal of Supercomputer Applications, 2001, 15(3): 200-222.
- [4] 袁文成, 朱怡安, 陆伟. 面向虚拟资源的云计算资源管理机制[J]. 西北工业大学学报, 2010, 28(5): 705-708.
- [5] Zhao CH, Zhang SS, Liu QF, et al. Independent tasks scheduling based on genetic algorithm in cloud computing[C]// Proceedings of the 5th International Conference on Wireless Communications, Networking and Mobile Computin. Beijing: IEEE, 2009: 1-4.
- [6] Burry R, Murshed M, Abranson D, et al. Scheduling parameter sweep applications on global grids: a deadline and budget constrained cost-time optimisation algorithm[J]. Journal Software Practice Experience, 2005; 35(5): 491-512.
- [7] 张海宾, 唐琳莎, 刘立祥. 网格调度综述[J]. 计算机工程与设计, 2009, 30(9): 2151-2154.
- [8] Song Ying, Wang Hui, Li Yaqiong, et al. Multitiered on-demand resource scheduling for VM-based data center[C]// Proceedings of the 9th IEEE/ACM International Symposium. Shanghai: IEEE Press, 2009: 148-155.
- [9] Song Ying, Li Yaqiong, Wang Hui, et al. A service-oriented priority-based resource scheduling scheme for virtualized utility computing[C]// HiPC 2008. Heidelberg: Springer-Verlag, 2008: 220-231.
- [10] 陈康, 郑伟民. 云计算: 系统实例与研究现状[J]. 软件学报, 2009, 20(5): 1337-1348.

作者简介:

熊安萍 女, (1970-), 副教授, 硕士生导师. 研究方向为操作系统、高性能计算。

黄凯 男, (1986-), 硕士研究生. 研究方向为云计算。