

基于云模型的负载均衡问题研究

赵 超¹, 王 晟^{2,3}

(1 河南商业高等专科学校, 河南 郑州 450052; 2 富国银行, 美国; 3 麦斯泰克公司, 美国)

摘 要: 云环境是一个复杂多变的环境, 其不同层次提供不同的服务, 在处理服务请求时其服务类型与服务器节点供给都是一个动态, 随机的变化过程, 如何高效的组织以及利用这些节点的处理能力, 是该环境中负载技术应该考虑的关键问题. 本文对云计算中私有云模型做了概括总结; 并对负载均衡调度问题在不同平台中的研究现状做了相关描述, 最后, 指出了负载均衡调度问题在私有云中未来的研究方向.

关键词: 云计算; 云模型; 私有云; 负载均衡

中图分类号: TP393

文献标识码: A

文章编号: 1000-7180(2012)03-0167-03

Load Balancing Based on Cloud Model

ZHAO Chao¹, WANG Sheng^{2,3}

(1 Henan Business College, Zhengzhou 450052, China; 2 Wells Fargo, USA; 3 Mastech Co. Ltd., USA)

Abstract: Cloud environment is a complex and changing environment, to provide different services at different levels, in dealing with the type of service when the service request and the server node is a supply of dynamic, random process of change, and how efficient organization and use of these nodes processing power, that is, the load in the environment, technology should consider. In this paper, the cloud made the private cloud model summarized; and load balancing scheduling problem on different platforms in the description of the relevant research status, and finally, that the load balance scheduling in the private cloud in the direction of future research.

Key words: cloud computing; cloud model; private clouds; load balancing

1 引言

云计算是从并行计算、网格计算和效用计算这些技术中演进而来, 目前已经有很多实际的应用, 所涉及的技术和构架可行性更强^[1]. 云环境是一个复杂多变的环境, 其不同层次提供不同的服务, 在处理服务请求时其服务类型与服务器节点供给都是一个动态, 随机的变化过程, 如何高效的组织以及利用这些节点的处理能力, 就是该环境中负载技术应该考虑的问题. 总而言之, 无论在什么领域都必须根据具体的特点, 综合考虑不同算法的策略来提高基于该领域的负载均衡度. 由于每个领域其自身的特点, 因此它们侧重提高算法的那个具体指标也不一样, 有着重提高吞吐率的, 也有着重提高服务响应时间的, 还有的是避免网络节点失效带来不可估量的损失

的. 等等这些, 都为我们研究云计算的负载均衡技术提供了大量珍贵的参考资料. 负载均衡技术在云计算这个领域中的研究甚少. 本文对云计算中私有云模型做了概括总结; 并对负载均衡调度问题在不同平台中的研究现状做了相关描述, 最后, 指出了负载均衡调度问题在私有云中未来的研究方向.

2 私有云模型

云计算可以分为公有云和私有云, 尽管公共云提供商提供规模经济和更成熟的云计算产品, 在可预测的未来, 私有云服务将大行其道. 私有云是为一个客户或一个企业单独使用而构建的, 因而提供对数据、安全性和服务质量的最有效控制. 因此, 在维持原有预算水平不变的前提下改进 IT 业务的最好方法就是搭建一个私有云架构, 其最大的好处就是

可以达到降低内部服务成本的目的^[2]。

如何可行并且高效的把企业现有的数据中心迁移到云计算环境中在最近这几年受到了广泛的关注,也有了比较成熟的案例,IBM 公司构建的 IBM Research Compute Cloud(RC2)就是其中之一。通过虚拟化技术的采用,可以对大型基础架构资源池(包括处理器、存储和网络等)实现聚合和全面管理,使其成为一种无缝、灵活、动态的操作环境,构建自己的数据中心内部的私有云架构。

私有云内部云网络问题比较复杂^[3],因为它的性能取决于有多少内部云连接与运行中的应用相连。所有的云计算组件都要依靠内部云网络来将用户与应用所需的资源相连。资源连接一旦形成,内部云网络的范围就决定了应用是否可以兼容多个系统。如果遵循 SOA 的规则,应用组件及应用和用户之间就会有数据的流通。这些连接的性能将会从整体上影响云计算的性能。私有云网络架构常常被用来支持主要数据中心站点间的数据流通。无论目标应用在哪个私有云上运行,这些数据流通必须要有足够的容量支持以确保网络延迟最小化。事实上创建私有云的效率很大程度上要取决于组成云的主要数据中心间网络连接的质量。

随着云计算技术的日益成熟,部署私有云的开源云计算产品也相继有了发布,比较典型的有 Eucalyptus,它主要是实现云计算环境的弹性需求的软件,该软件层的工具可以用来通过配置服务器集群来实现私有云,并且其接口也是与公有云相兼容,开源满足私有云与公有云混合搭建扩张的云计算环境。该云计算系统使用维护十分方便,使用 SOAP 安全的内部通信,且把可伸缩性作为主要的设计目标,具有简单易用,扩张方便的特点,其物理部署结构如图 1 所示^[4]。

3 负载均衡调度问题

私有云系统中有大量的异构资源组成,目的是实现资源的全面共享和协同工作,因此资源调度问题会变得越来越重要。文献[5]中在网格系统环境中对各类的静态调度算法和动态调度算法进行资源调度的仿真,并对各类算法的运行进行负载均衡和性能的比较。跟网格计算有所不同,云计算环境中资源种类多、规模大,对资源的实时监控和管理变得十分困难。一个云计算环境可能有成千上万的计算任务。这些都给负载均衡调度算法的复杂性和有效性提出了新的挑战。

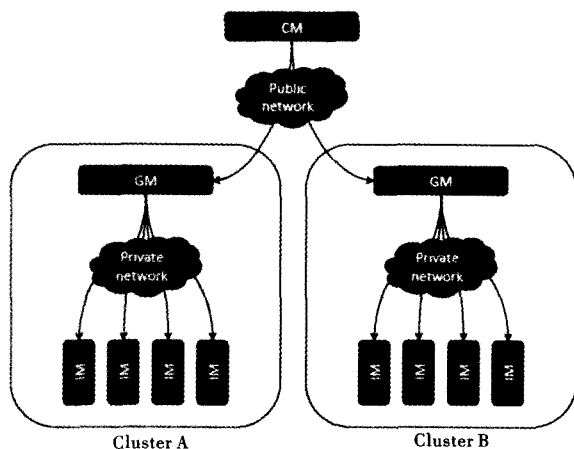


图 1 Eucalyptus 物理部署结构图

所谓负载,是指被分配到各个服务器结点上并执行的子任务。负载在并行系统的各服务器结点上分布的均衡程度称为负载均衡度。负载均衡是实现多台服务器协同工作和并行处理的手段,可以极大地提高服务器的性能,充分利用网络资源^[6]。

负载均衡策略的好坏也取决于是否达到负载均衡的技术指标,从均衡策略上划分负载均衡可分为静态负载均衡和动态负载均衡。静态负载均衡在进行负载均衡时不需要知道集群中当前各结点的负载状态,只是根据集群中各结点处理能力上的差异,按事先确定好的分发策略将任务相应地分配到各个服务器上。由于信息服务中越来越多地使用动态嵌入对象(scripts 程序等)和数据库查询任务,使得不同请求任务的工作负荷有很大差异,静态和动态的负载可以相差 10 倍或 100 倍^[7]。单靠静态算法已经不能达到负载均衡的效果。

动态负载均衡通过应用软件实时的分析数据包,获取集群中各服务器的当前负载状态及数据流量状况,把任务动态地合理地分配到各个服务器。此类算法还引进了一种反馈机制,通过监测每一个节点的实时负载和响应能力,而不断调整任务分布的比例,来尽量避免服务器利用率严重倾斜。随着访问用户和新型服务的不断增加,现有的负载均衡调度算法逐渐显露出了各种弊端。

最理想的调度算法是把负载请求调度到综合负载最轻的服务器上,还有负载均衡策略应有对网络故障、服务器系统故障、应用服务故障的监测方式和能力,以便一台负载均衡器崩溃可以把数据流引向网络连接正常的负载均衡器^[8],但在实际应用中很难实现。主要问题是:

(1) 算法复杂度增大。如果同时有若干负载请

求,调度算法到底分配多少请求到综合负载最轻的服务器上很难确定,算法实现起来相当困难。

(2) 如果把负载请求调度到综合负载最轻的服务器上,而此时某一台服务器的综合负载已经很大,如对服务器再进行数据采集(本次数据采集是获取服务器负载数据)的操作只会加重服务器的负荷,造成服务器的阻塞。

负载均衡技术发展至今,产生了很多对某一具体领域的负载均衡调度算法的研究与改进。文献[9]中通过校园分析视频点播系统面临的诸多问题,根据校园网络的特点和影响负载均衡的因素,对基于校园网络的负载均衡技术进行了分析,提出一种动态反馈机制原理的负载均衡算法,实现校园网络中视频资源的共享、均衡校园视频资源的负载。

文献[10]中基于数据复制/转移策略,提出一种在P2P网络环境中的动态平衡算法。根据节点的能力,当前节点负载状态、负载转移代价预估算,在整个系统范围内找到一组传输代价较小并且负载较轻的节点集合,从中随机选取较为适宜的节点进行负载转移或者数据复制,该算法能够有效地均衡负载的分布以及降低负载的迁移率。

文献[11]中基于单位时间开销代价矢量,提出同时考虑地形数据请求在服务器队列中的等待时间与服务器对请求的处理时间的最小总代价分布式算法。算法计算转发请求给每台服务器的概率空间,并根据地形数据请求到达时临时生成的随机数在概率空间中的落点确定转发请求的目标服务器。通过配置单位时间开销代价矢量,可以灵活地维护与升级集群服务器,具有良好的可扩展性。最小总代价算法能在大规模、高强度的地形漫游中均衡地分发请求。使集群服务器充分发挥其优势从而获得最小的漫游响应时间。在低强度地形漫游时,各种负载均衡算法表现大致相同;在高强度地形漫游时,设计负载均衡算法最好考虑服务器端队列的排队情况。

在分布式环境中要按照具体的环境(处理的服务类型的种类不同,主机的数量不同等)采用不同的算法。从以上负载均衡应用中可以得知,选择合适的负载均衡策略,才能使服务器之间能很好的共同完成任务。如何选择合适的负载均衡策略对云计算中的大量异构资源进行有效的分配是云算系统管理模块负责管理和分配所有可用的资源的核心问题。当系统面临不同的应用请求时,通过负载均衡技术快速并且有效的响应用户请求,首先对云计算中的大量异构资源进行重组,使相似的处理能力的节点聚

合到一起,从而把一个大型的复杂系统分割成若干个子系统,子系统之间以较少的边连接,而子系统内部节点则以较多的边连接,如图2所示^[12]。

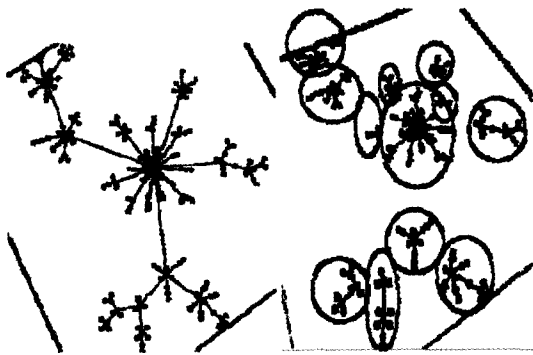


图2 异构资源重组

异构资源重组后,再对这些资源进行有效的负载均衡策略。由于云计算是一个按需提供资源给用户的平台,根据用户不同的需求,分配不同的资源,因此也可以根据这个变动提供不同的负载均衡调度策略。面对不同的资源需求时调用这两个算法更符合私有云中面临的资源负载均衡问题,在处理任务类型相对少而集群中主机数多时调用活动聚类算法,而处理任务类型和主机数都比较多时采用“蜜蜂”战略算法,利用两个算法的各自优势最终达到优化云环境中负载均衡性能的目的。

4 结束语

本文对云计算中私有云模型做了概括总结;并对负载均衡调度问题在不同平台中的研究现状做了相关描述,最后、指出了负载均衡调度问题在私有云中未来的研究方向。

参考文献:

- [1] 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.
- [2] 金海. 漫谈云计算[J]. 中国计算机学会通讯,2009,5(6):22-25.
- [3] 尹国定,卫红. 云计算——实现概念计算的方法[J]. 东南大学学报:自然科学版,2003,33(4):502-506.
- [4] Gu Yunhong, Robert L G. Sector and sphere: the design and implementation of a high-performance data cloud[J]. Philosophical Transactions of the Royal Society, 2009(367):2429-2445.

- [2] Mikic I, Cosman P, Kogut G, et al. Moving shadow and object detection in traffic scenes [C]//Proceedings of the 15th International Conference on Pattern Recognition, Barcelona, Spain;IEEE, 2000.
- [3] Haritaoglu I, Harwood D, et al. W4:real-time surveillance of people and their activities[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000,22(8):809-830.
- [4] Cucchiara R, Grana C, Neri G, et al. The sakbot system for moving object detection and tracking[J]. Video-based Surveillance Systems-Computer Vision and Distributed Processing. 2001, 145-157.
- [5] Stauder J, Mech R, Ostermann J. Detection of moving cast shadows for object segmentation [J]. IEEE Transactions on Multimedia,1999(1):65-76.
- [6] Herley G. Wavelet and recursive filter banks[J]. IEEE Trans Signal Processing,2536-2556.
- [7] Sweldens W. The lifting scheme: A construction of second generation wavelets [J]. SIAM Journal on Mathematical Analysis, 1997, 29(2): 511-546.
- [8] Daubechies I, Sweldens W. Factoring wavelet transforms into lifting step [J]. Journal of Fourier Analysis and Applications,1998, 4(3), 247-269.
- [9] Munleanu A, Cornelis J, DeMuynck P, et al. Performance evaluation of the wavelet based techniques used in the lossless and lossy compression of medical and preprint image [R]. Technical Report, ETRO Department, Free University of Brussels, 1997.
- [10] Dashan Gao, Jie Zhou. Adaptive background estimation for real-time traffic monitoring[J]. Intelligent Transportation Systems,2001:330-333.

作者简介:

卡建勇 男,(1980—),博士,讲师.研究方向为计算机视觉及嵌入式系统应用.

(上接第169页)

- [5] 王琴,曾文华,史文肿. 网络资源调度算法的负载均衡及性能分析 [J]. 微电子学与计算机,2006,23(10): 201-204.
- [6] 刘高峰. 负载均衡技术全攻略 [EB/OL]. [2011-08-10]. <http://www.yesky.com/20010626/187006.shtml>.
- [7] Iyengar A, Macnair E, Nguyen T. An analysis of Web server performance[C]//Proc of Global Telecommunications Conference. New York: [s. n.], 1997: 1943-1947.
- [8] Daniel J A. Data management in the cloud: limitations and opportunities [J]. Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, 2009,32(1):3-12.
- [9] 周莹莲,任胜兵. 校园网格中 VOD 系统动态负载均衡算法研究 [J]. 计算机与数字工程, 2009(6):29-32.
- [10] 詹晓强,胡德敏. 基于 P2P 系统的动态负载均衡算法 [J]. 计算机工程与设计,2009,30(1):58-60.
- [11] 王浩,喻占武,李锐,等. 基于开销代价的网络地理信息服务负载均衡算法研究 [J]. 测绘学报, 2009(6):242-249.
- [12] Martin Randles, David Lamb, A. Taleb-Bendiab. Experiments with Honeybee Foraging Inspired Load Balancing [C]//Proceedings of the 2009 Second International Conference on Developments in eSystems Engineering. Abu Dhabi:IEEE,2009:240-247

作者简介:



赵超 男,(1975—),硕士,讲师.研究方向为信息系统、数据管理.

王晟 男,(1972—),硕士,工程师.研究方向为计算机大型数据库的研究与开发.