

云计算中基于 Session 和内容等级的数据库请求分类算法

赵涓涓 刘涛 强彦 卢军佐

(太原理工大学计算机科学与技术学院 太原 030024)

摘要 云计算是当今世界 IT 行业的研究热点,在云计算实例中存在各种各样的用户对于数据库的请求,并且这些请求是海量的,如果不对这些数据库请求进行合理的分类,必然会对系统的性能造成很大的影响。如何对这些请求进行分类以达到 Web QoS 的标准,是云计算研究的一大难点和关键点。原来基于先来先服务的请求模式不仅不能满足客户优先级的要求,不能达到利益的最大化,而且不能使服务资源得到充分利用。因此,提出了基于 Session 和内容等级的数据库请求分类算法,该算法提出了基于 Session 和内容等级的性能函数,充分考虑了客户的优先级,使得优先级高的客户的请求优先得到服务资源;而对于优先级较低请求,由于在算法中加入了时间函数,其优先级会随着等待时间的加长而变成优先级较高的请求,从而也会得到处理,不至于无限期地等待。因此,该算法既保证了优先级高的请求优先得到处理,又保证了优先级低的请求不至于由于优先级较低而得不到处理,是一种动态的数据库自适应的数据库请求分类。

关键词 云计算, Session, 内容等级, 数据库请求分类, 数据库自适应

中图分类号 TP301 **文献标识码** A

Classification Algorithm of Database Request Based on Session and Content Level in Cloud Computing

ZHAO Juan-juan LIU Tao QIANG Yan LU Jun-zuo

(Department of Computer Science and Technology, Taiyuan University of Technology, Taiyuan 030024, China)

Abstract Nowadays, many studies have been focused on cloud computing in IT industry. Mass of users' requests on database exists in instance of cloud computing. It has a significant impact on system performance if we do not classify these requests on database reasonably. How to classify these requests, in order to achieve the Web QoS standard, is a aporia and key point in the study of cloud computing. There are many defects in original request pattern, which is based on first come and first service. For instance, priority requirements from customers cannot be satisfied, and we failed to reach the aim of profit maximization. Moreover, service resources cannot be fully used. Therefore, we put forward the database request classification algorithm based on Session and content level. We also put forward a performance function based on Session and content level. This function takes customers' priority into consideration, in order to make customers' request with high priority getting service resources first. While to the others, owing to the introduction of time function in our algorithm, their priorities will get promoted and finally get processed as the waiting time goes by. The situation of endless waiting will be avoided. Therefore, this algorithm is a kind of dynamic database request classification of database adaptive. In our algorithm, not only the request with highly priority will get processed first, but also it is ensured that the request with lower priority will be processed although their priority is not enough.

Keywords Cloud computing, Session, Content level, Database request classification, Database adaptive

云计算是当今世界 IT 行业的研究热点,在云计算中存在各种各样用户的数据库请求,并且这些请求是海量的,如果不对这些请求进行合理的分类,对于不同类的请求分配不同的服务资源,势必对系统的性能造成很大影响。如何对这些请求进行分类以达到 Web QoS^[1] 的标准,是云计算研究的一大难点和关键点。在本文中,对于优先级较高的客户请求要优先分配服务资源,从而使得客户请求得到满意的保证;对于优

优先级较低的客户请求,虽然不能马上给其分配服务资源,但也不能对其置之不理,以免流失潜在的客户,不能达到效益最大化。因此,需要提出一种新的分类算法,该算法能动态地区分客户请求的优先级,使得一开始优先级较低的客户请求随着时间的推移能变成优先级较高的请求,从而得到服务资源。

原来先来先服务的请求服务模式不仅不能满足客户优先级的要求,而且不能使得服务资源得到充分利用,不能达到利

到稿日期:2012-04-22 返修日期:2012-07-13 本文受国家自然科学基金项目(61202163),山西省自然科学基金(2012011015-1),山西省科技攻关项目(20120313032-3)资助。

赵涓涓(1975-),女,博士,副教授,主要研究领域为云计算、医学图像处理、模式识别, E-mail: zh_juanjuan@126.com(通信作者);刘涛(1988-),男,硕士,主要研究领域为云计算;强彦(1969-),男,博士,副教授,主要研究领域为云计算、图像处理、并行计算;卢军佐(1986-),男,硕士,主要研究领域为云计算、医学图像处理。

益的最大化。因此,提出了基于 Session 和内容等级^[2]的请求分类算法^[3],该算法提出了基于 Session 和内容等级的性能函数^[4],充分考虑了客户的优先级,使得优先级高的客户的请求优先得到服务资源;而对于优先级较低请求,由于在算法中加入了时间函数,其优先级会随着等待时间的加长而变成优先级较高的请求,从而得到处理。因此,该算法既保证了优先级高的请求优先得到处理,又保证了优先级低的请求不至于由于优先级较低而得不到处理。

1 请求分类机制

实现 Web QoS 控制,第一个环节要做的工作是将数据库请求进行区别,然后进行分类^[9]。到目前为止,存在着多种分类机制,现存的机制大体上能够划分为两类,第一类是以客户为基础的分类,第二类是以请求目标为基础的分类^[10]。

以客户为基础的分类

以客户为基础的分类的依据包括客户的某些特殊属性及某些特征,具体包括:

(1)依据客户端 IP 地址实现分类。这种方法是根据客户端的 IP 地址来辨别不同的客户,优点是很容易实现,不足之处是代理服务器或防火墙会屏蔽客户端的 IP 地址^[12],所以其应用将会受到限制。

(2)依据 HTTP Cookie 实现分类。Web 服务器将 HTTP Cookie 返回给客户,作用是将客户信息的数据块收集起来。HTTP Cookie 标识符具有唯一性,它能够嵌入到 HTTP 的请求内,以表明客户所从属的类别。例如,持久化的 Cookie 可以用来完成对于特定商品或者服务的网上订购;此外,还可以通过 Cookie 来识别一个已经建立连接的会话(Session),从而实现依据会话的请求分类。

(3)依据浏览器 plug-ins 实现分类。浏览器 plug-ins 也可以把特定的客户标识符嵌入到每一个 HTTP 的请求体中。这些依据客户的分类方法可以用来设置哪一个客户组优先,其他客户将能够得到更好的 Web QoS 保证。

依据请求目标实现分类

依据请求目标实现分类的依据是请求目标所特有的一部分属性和一些特征,具体如下:

(1)依据请求 URL 实现分类。不同请求之间的相对重要程度可以通过 URL 请求类型或者文件名的路径来进行区分。

(2)依据目标 IP 地址或者端口号实现分类。当许多站点位于相同的 Web 服务器节点时,可以通过服务器使用请求的目的 IP 地址或端口号实现对请求的分类。依据目标实现分类的方法可以提供更好的 Web QoS,比如在重要请求的实现(信用卡的支付请求)或者为更多的虚拟站点付费时,依据目标实现分类的方法还可以对资源紧要型请求的资源消耗情况进行控制。

对于请求分类来说,通常以上两种分类机制就已经足够,当然也可以将依据客户和依据目标的特定属性相结合,从而进行分类。

在此,将同时考虑两方面的因素:访问内容等级和 Session,并且在二者之间考虑一定的公平性,然后设计出 Web

QoS 性能函数,设计访问内容等级和 Session 相结合的 Web QoS 调度算法,为用户提供对优先级进行区分的 Web QoS 服务。

2 算法原型设计

本算法研究的是在系统处于繁忙的状态时,请求队列中的各种等级不同的调度策略及其方法。请求队列中,系统将首先调度哪一个请求由该算法决定。每一个请求都满足:

$$f(R(x), P(k), T(t)) = \alpha R(x) + \beta P(k) + \delta T(t) \quad (1)$$

$(0 \leq \alpha \leq 1, 0 \leq \beta \leq 1, 0 \leq \delta \leq 1)$

式中, $f(R(x), P(k), T(t))$ 为性能函数, $R(x)$ 为内容等级函数,其值由内容的等级决定; $P(k)$ 为 Session 等级函数,其值由用户 Session 的等级决定; $T(t)$ 为等待时间函数,作用是调节时间调节数据的性能函数值。各个函数的权重因子为: α, β, δ 。当 $\alpha=0, \delta=0$ 时, $f(R(x), P(k), T(t))$ 代表的是 Session 等级的性能函数;当 $\beta=0, \delta=0$ 时, $f(R(x), P(k), T(t))$ 代表的是依据内容的性能函数。队列缓存区的内容设为 n ,则任一时刻 t ,都有:

$$\begin{cases} f_{\max}(R(x_i), P(k_i), T(t)), i \in [1, n] \\ \sum_1^n f(R(x_i), P(k_i), T(t)) \end{cases} \quad (2)$$

这样,具备 $f_{\max}(R(x_i), P(k_i), T(t))$ 的数据包将得到系统的始终优先处理,尽量使得当前队列缓存区中的 $\sum_1^n f(R(x_i), P(k_i), T(t))$ 保持最小。

这样,在性能函数中,各个相关函数值的影响权重就可以通过调节 α, β, δ 来实现。一般情况下, $(R(x_i), P(k_i))$ 综合的权重比较大。当最高优先级 Session 的用户在请求高优先级内容访问时,就能够优先得到处理,然后保证获得级别较高的 Web QoS。而那些来自于低优先级的 Session 用户在对低优先级内容请求访问时(即 $(R(x_i), P(k_i))$ 综合的权重较小时),这样的性能函数值较小,因此需要等待被处理。但是伴随等待时间 $T(t)$ 的不断增长,该性能函数值将不断变大,该数据包将在经过一定时间的等待后升级成为 $f_{\max}(R(x_i), P(k_i), T(t))$,进而得到系统资源,得到处理,从而考虑达到该算法的公平性。

3 算法模拟程序设计

算法的模拟程序部分主要是表达出算法的基本思想,因此,我们适当地简化了原性能函数,将 $(R(x), P(k))$ 合并为一项(作为性能函数的等级项)进行考虑,同时将权重因子 α, β 合并为一个等级项权重因子 γ ,则性能函数简化为:

$$f(C(x), T(t)) = \gamma C(x) + \delta T(t) \quad (3)$$

这样可以根据 $C_i(x)$ 将数据简单地分为两类:高级别数据和低级别数据。

然后,将设计程序来对队列缓存区处于最繁忙时候的状态进行模拟(一直保持全满状态),并假设当 CPU 服务能力达到极限时,系统将拒绝接受任何数据,新的空间只有在队列缓存区中有数据在进行处理或者数据被丢弃后,才会允许接收新的数据。该算法的关注点在于,如何在队列缓存区内优先处理等级不同的请求。模拟程序运行如图 1 所示。

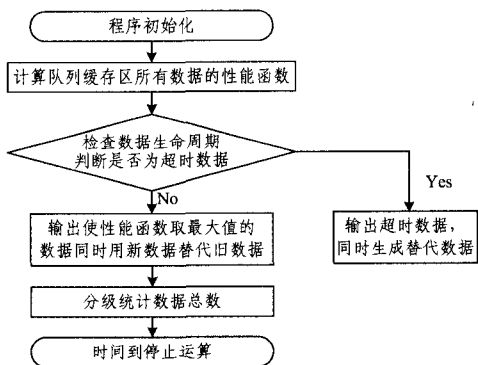


图1 模拟程序运行

其中,程序初始化阶段又包括定义队列缓存区容量、生成队列缓存区内的原始数据、定义数据生命周期、定义数据的单位处理时间、初始化等级权重因子和定义模拟时间6个部分。

模拟算法运行结果及其分析

在模拟算法中,模拟运算将按照如下默认参数进行:

- 1)处理请求的缓冲区队列设为100;
- 2)综合等级的权值设为0.6;
- 3)等待时间的权值设为0.05;
- 4)单位处理时间设为10ms;
- 5)请求等待超时设为500ms;
- 6)总模拟时间为120s。

可以得到如图2所示的模拟算法运行结果。

图2 模拟算法运行结果

模拟算法的运行结果统计如表1所列。

表1 运行结果统计

	随机生成 请求	系统处理 的请求	系统丢弃 的请求	缓存区的 遗留请求
高优先级请求	12079	9802	2228	49
低优先级请求	11869	2197	9621	51
请求总数	23948	11999	11849	100

由表1可以看出,系统将优先处理优先级别较高的请求,对低优先级请求处理的公平性由算法经过调节时间函数 $T(t)$ 来兼顾,所以,在系统优先处理高优先级请求的同时,系统也处理一定数量的低优先级请求。同理,高优先级请求由于性能函数值比较大,等待的时间比较短,因此在进行超时丢弃处理时,被丢弃的请求量也相对较少。还能够看到,当算法的运算完成时,由于随机生成了请求,因此缓存区内请求仍然基本满足随机分布。演示的结果表明,该算法能够实现按照请求的等级来进行有区分的优先级请求处理。如果将该算法应用到为请求队列提供 Web QoS 服务的调度中,就能

够以适当公平性为前提为不同的用户提供有区别的 Web QoS 服务。此外, α, β, δ 参数的值会对算法产生较大的影响,这也是为网站选择不同的 Web QoS 服务给出了一定的灵活性的依据。

结束语 本文以数据库请求分类机制以及 Web QoS 的控制方法为基础,提出了一种优先级动态分类思想,该思想能够同时兼顾 Session 和 Web 内容,还给出了兼顾公平性的性能函数(兼顾 Session 和内容之间的公平性),并依据提出的性能函数对算法原型进行了设计。最后,数据模拟了该算法,仿真实验的结果证实了该算法可以在兼顾公平性的基础上为用户提供有区分的 Web QoS 优先级服务。

参考文献

- [1] Siala F, Ghedira K. A Multi-Agent Selection of Web Service Providers Driven by Composite QoS[C]//ISCC. 2011;55-60
- [2] Mehta H, Kanungo P, Chandwani M. Decentralized Content Aware Load Balancing Algorithm for Distributed Computing Environments[C]//ICWET. 2011;370-375
- [3] Vinek E, Beran P P, Schikuta E. Classification and Composition of QoS Attributes in Distributed, Heterogeneous Systems[C]//CCGRID. 2011;424-433
- [4] Niehorster O, Keller A, Brinkmann A. An Energy-Aware SaaS Stack[C]//MASCOTS. 2011;450-453
- [5] lou D, Yoshida P-M, Nobuko. Dynamic Multirole Session Types [J]. ACM SIGPLAN Notices, 2011, 46(1): 435-446
- [6] Jeong K, Kim J, Kim Y-T. QoS-aware Network Operating System for Software Defined Networking with Generalized Openflows[C]//Network Operations and Management Symposium (NOMS). 2012
- [7] Dávid S, Bálint D, András A B. Content-based Trust and Bias Classification Via Biclustering [C] // WebQuality' 12 Proceedings of the 2nd Joint WICOW/AIRWeb Workshop on Web Quality. 2012;41-47
- [8] 李银钊, 闫怀志, 张佳, 等. 基于自适应模型的数据库入侵检测方法[J]. 北京理工大学学报, 2012, 32(3): 258-262
- [9] Mostafa, Mahmoud, Kalam E, et al. QoS-aware Firewall Session Table[C]//6th International Conference on Risks and Security of Internet and Systems(CRISIS). 2011
- [10] Chang Bao-rong, Tsai H-F, Lin Z-Y, et al. Adaptive Performance for VVoIP Implementation in Cloud Computing Environment [C]//Intelligent Information and Database Systems-4th Asian Conference. 2012, 7198(PART 3): 356-365
- [11] Chen Y-M, Wang W-C. An Adaptive Rescheduling Scheme Based Heuristic Algorithm for Cloud Services Applications[J]. International Conference on Machine Learning and Cybernetics, 2011, (3): 961-966
- [12] Chen K-Y, Lin C-Y, Hou T-W. The low-cost secure sessions of access control model for distributed applications by public personal smart cards[C]//Proceedings of the International Conference on Parallel and Distributed Systems. 2011;849-859
- [13] 姚金成, 张世栋, 史玉良, 等. 基于 Chunk Folding 的多租户数据库缓存管理机制[J]. 计算机学报, 2011, 34(12): 2320-2331
- [14] 林子雨, 赖永炫, 林琛, 等. 云数据库研究[J]. 软件学报, 2012, 23(5): 1148-1166
- [15] 尚云鹏, 刘宴兵, 杨莎莎, 等. 基于云计算的 OMS 即时通讯系统设计与实现[J]. 重庆邮电大学学报: 自然科学版, 2010, 22(4): 468-472