

访问 Web GIS 服务的一种动态负载均衡方法

张喜平 汪林林²

(重庆邮电学院计算机科学与技术学院 重庆400065)¹ (重庆邮电学院软件学院 重庆400065)²

摘 要 本文分析了 Web GIS 的瓶颈问题,引入了基于请求分配器的服务器集群和代理技术,提出了访问 Web GIS 服务的一种动态负载均衡的方法,从而以较低的成本消除了 Web GIS 网络瓶颈,提高了 Web GIS 的数据传输性能。

关键词 Web GIS,负载均衡,代理,请求分配器,Web 服务器集群

The Method of Dynamic Load Balance Access to Web GIS Service

ZHANG Xi-Ping¹ WANG Lin-Lin²

(Department of Computer Science and Technology, Chongqing University of Post & Telcom, Chongqing 400065)¹

(Soft College of Chongqing University of Post & Telcom. , Chongqing 400065)²

Abstract This paper analyzes the bottleneck problem of Web GIS and in the technology of agent and Web sever cluster based on dispatcher introduces the dynamic method of load balance access to Web GIS service. This method uses cheap cost to avoid the problem of bottleneck problem in Web GIS and improve data transport performance in Web GIS.

Keywords Web GIS, Load balancing, Agent, Dispatcher, Web sever cluster

1 前言

随着 Web GIS 的广泛应用,Web GIS 用户的增多,大量的空间数据和属性数据需要在网络中传输,空间数据量大及其造成的网络传输慢是 Web GIS 系统最大的瓶颈,这使得在 Web GIS 应用中出现了客户端响应速度慢、访问不了 Web GIS 站点等一些影响 Web GIS 性能的问题,解决这些问题成为 Web GIS 发展的重要问题。如何设计一种廉价有效的方法来扩展服务器带宽、增加吞吐量以加强网络数据处理能力、提高网络传输速度和用户操作处理响应速度,从而通过提高网络的灵活性和可用性来提高 Web GIS 的数据传输性能是本文重点研究、解决的问题。

基于以上问题本文提出了访问 Web GIS 服务的一种动态负载均衡的方法,该方法采用了基于请求分配器的 Web 服务器集群和代理技术来实现动态的负载均衡。根据此方法我们设计了用动态负载均衡方法访问 Web GIS 服务的网络体系结构、负载均衡代理的结构以及请求分配器的负载均衡策略。

2 负载均衡概念

负载均衡并非传统意义上的“均衡”,一般来说,它只是把有可能拥塞于一个地方的负载交给多个地方分担。说得通俗一点,负载均衡在网络中的作用就像轮流值日制度,把任务分给大家来完成,以免让一个人累死累活。不过,这种意义上的均衡一般是静态的,也就是事先确定的“轮值”策略。

与轮流值日制度不同的是,动态负载均衡通过一些工具实时地分析数据包,掌握网络中的数据流量状况,把任务合理分配出去。结构上分为本地负载均衡和地域负载均衡(全局负载均衡),前一种是指对本地的服务器集群做负载均衡,后一种是指对分别放置在不同的地理位置、在不同的网络及服务器集群之间作负载均衡。

在现有网络结构之上,负载均衡提供了一种廉价有效的

方法来扩展服务器带宽和增加吞吐量,加强网络数据处理能力,提高网络的灵活性和可用性。它主要完成以下任务:

- ◆解决网络拥塞问题,服务就近提供,实现地理位置无关性。
- ◆为用户提供更好的访问质量。
- ◆提高服务器响应速度。
- ◆提高服务器及其他资源的利用效率。
- ◆避免了网络关键部位出现单点失效。

3 基于请求分配器的 Web 服务器集群技术

当前,在 Web GIS 中,业务量的发展都超出了过去最乐观的估计,尤其是 Web GIS 网络的核心部分,其数据流量和空间数据计算强度之大,使得单一设备根本无法承担,因此引入服务器集群来支持其 Web GIS 服务。

Web 服务器集群系统是由分布在 LAN 或 WAN 上的多台 Web 服务器主机(同构的或异构的)相互联结而成的一种服务器体系结构,它采用负载均衡策略将到达的请求分配给集群中的某台服务器进行处理。

基于请求分配器(Dispatcher)的 Web 服务器集群能够完全控制所有到来的请求并且实现精细粒度的负载均衡。在基于请求分配器的集群系统中,前端的请求分配器作为到达请求的代理,负责集中地接收所有到达的 HTTP 请求,并且按照特定的负载均衡策略将客户的请求均衡、透明地分配给集群中的后端服务器。整个集群系统具有一个单一的虚拟 IP 地址,即集群地址,因此集群中的服务器对用户端是透明的。实际上,集群地址就是请求分配器的 IP 地址,即请求分配器为集群系统提供了一个单一的虚拟接口,使得整个集群对外部而言犹如一台单一的主机。

4 访问 Web GIS 服务实现动态负载均衡的网络体系结构

为了快速响应客户端的请求,根据以上介绍的技术,我们

设计了动态负载均衡方法访问 Web GIS 服务的网络体系结构(图1)。该体系结构中客户端通过 Internet 与基于代理的请求分配器相连接,请求分配器执行基本的负载均衡操作它与服务器端的 Web 服务器集群以及地图服务器和数据库一起工作,根据请求分配器的负载均衡策略来为客户端的请求选择合适的 Web 服务器,从而实现负载均衡,减少 Web 服务器的负载。

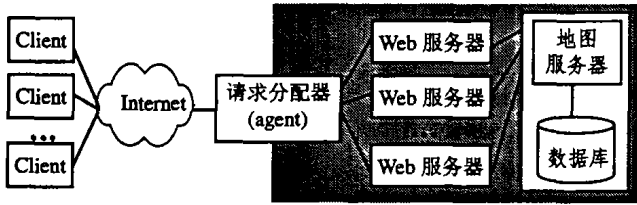


图1 动态负载均衡方法访问 Web GIS 服务的网络体系结构

图2描述了体系结构中消息的传递过程:当客户端请求地图的时候请求分配器根据分配策略选择最适合的 Web 服务器,然后 Web 服务器向地图服务器请求地图服务最后把请求的结果返回给客户端。

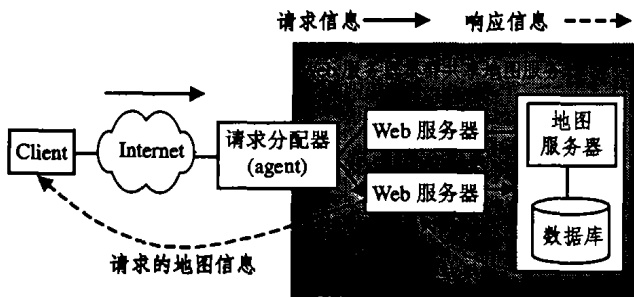


图2 消息传递过程

5 负载均衡代理的设计

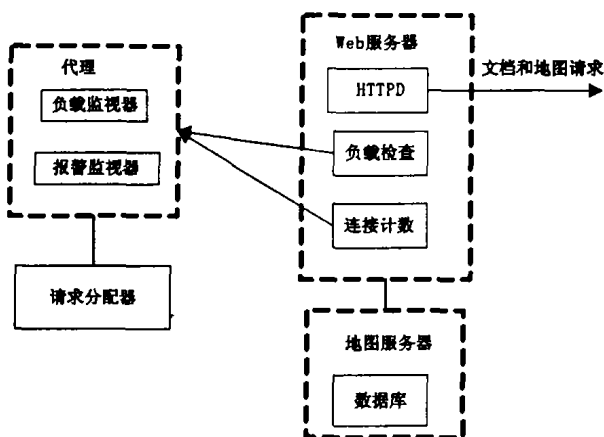


图3 Web 服务器、地图服务器和代理处理负载均衡的结构

负载均衡代理位于请求分配器之上,其作用是减少 Web 服务器的负担,提高客户端的响应速度。图3是我们设计的代理与 Web 服务器、地图服务器一起处理负载均衡的结构。该结构中,要处理负载均衡需要在代理和 Web 服务器中设计相应的组件。其中,在代理中我们设计了负载监视器和报警监视器组件,它们追踪并反馈 Web 服务器的负载状况,防止服务

器过载,并且把收集到的信息存在一个负载监控表里面。Web 服务器端我们设计了负载检查和连接计数组件,这两个组件将安装在集群中的每个 Web 服务器中。其中负载检查组件追踪服务器的 CPU 负载、存储器负载和磁盘使用状况信息。连接计数组件计算 Web 服务器和地图服务器之间的连接数量信息。这些信息周期地提供给代理软件并保存在代理的负载监控表里,请求分配器从代理的负载监控表里读取信息并根据负载均衡策略来选择负载值最小的 Web 服务器。

6 请求分配器的负载均衡策略

请求分配器的负载均衡策略如下:当客户端请求地图服务的时候,请求分配器根据服务器的负载信息选择使用一个 Web 服务器,被选中的服务器将是拥有最低负载值的服务器。为了计算的方便我们计算负载的时候只考虑了 cup 负载、存储器负载、磁盘的使用状况以及连接 Web 服务器和地图服务器数量的信息。计算 Web 服务器负载的公式如下:

$$\text{Total_Load} = w \times (\text{Cpu_Load} + \text{Memory_Load} + \text{Disk_using} + \text{Num_map_server_connection})$$

其中:

Total_Load—Web 服务器的负载值;W—权值;Cpu_Load—cpu 负载;Memory_Load—存储器负载;Disk_using—磁盘使用状况;Num_map_server_connection—连接 Web 服务器和地图服务器的数量。

权值的赋予根据每个负载的不同而不同。当集群的结点初次投入系统中使用时,系统管理员根据结点的硬件配置情况对每个结点都设定一个初始权值 W(Ni) (通常根据结点的硬件配置来定义,硬件配置越高的结点默认值越高),在请求分配器上也先使用这个权值。然后,随着结点负载的变化,请求分配器对权值进行调整。权值是一个大于0而小于等于1的参数。根据权值 Web 服务器负载值计算公式变形如下:

$$\text{Total_Load} = W_{\text{cpu}} \times \text{Cpu_Load} + W_{\text{mem}} \times \text{Memory_Load} + W_{\text{disk}} \times \text{Disk_using} + W_{\text{conn}} \times \text{Num_map_server_connection}$$

其中:W_{cpu}+W_{mem}+W_{disk}+W_{conn}=1

结论 本文提出了访问 Web GIS 服务的一种动态负载均衡的方法,该方法的提出能够让集群中的多台 Web GIS 服务器共同承担一些繁重的 Web GIS 计算或 I/O 任务,从而以较低成本消除网络瓶颈,通过提高 Web GIS 系统中网络的灵活性和可靠性,从而提高了 Web GIS 的性能,为 Web GIS 的客户端访问和用户决策提供了可靠的保证。

参考文献

- 1 骆宗阳,王澄,等. 具有高可用性的负载均衡技术的研究与实现[J]. 计算机工程与应用,2003
- 2 朱利,张兴军. Web 服务器组的负载均衡方法研究[J]. 小型微型计算机系统,2003(12)
- 3 吴信才,等. 地理信息系统原理与方法[M]. 北京:电子工业出版社,2002(3):214~216
- 4 修文群,池天河,等. 城市地理信息系统(GIS)[M]. 北京:希望电子出版社,1999(6):100~12