



大数据网络服务器群智能伸缩机制与架构研究

鞠洪尧

(浙江纺织服装职业技术学院 宁波 315211)

摘要:阐述了大数据网络服务器群智能伸缩原理及架构的构建过程,对大数据网络服务器群智能伸缩涉及的关键技术做了深入的探讨。围绕服务器群工作负载监测、服务器数量智能增减控制以及访问负载调度等关键技术展开研究,给出了大数据网络服务器群智能伸缩机制和实现方法,为大数据网络智能伸缩及有效节能提供了技术支持。依据所提的构建方法和关键技术,给出了大数据网络服务器群智能伸缩模型。

关键词:大数据网络;服务器群;智能伸缩;负载调度

doi: 10.11959/j.issn.1000-0801.2015073

Research on Smart Scaling Mechanism and Structure for Big Data Network Server Groups

Ju Hongyao

(Zhejiang Textile & Fashion College, Ningbo 315211, China)

Abstract: The principle and construction of the structure of smart telescopic on big data network server groups were elaborated. The key technologies involved in the smart scaling of big data network server groups were discussed. After investigating key technologies, including the monitoring of the workload of the server group, the smart control of the number of servers, and the scheduling of access loads, the principle and realization method for the smart scaling of big data network server groups were proposed and technical supports were provided for the smart scaling and effective energy conservation of big data network server groups. According to the proposed construction method and key technologies, a smart scaling model for big data network server groups was provided as well.

Key words: big data network, server group, smart scaling, load scheduling

1 概述

随着计算机网络技术和信息化技术的快速发展,各行各业正在快速生成海量的业务数据,这些海量的数据^[1~4]由结构化、半结构化和非结构化的数据构成,其中含有

大量对行业发展和决策有价值的信息^[5,6],各行业都希望快速获得这些有价值的目标信息,因此,对目标信息的快速查找、分析以及高质量集成越来越受到各行各业的高度关注,且正在逐步形成一种新型的信息服务产业^[7,8]。通过对大数据的概念、构成及现状的综合分析,对大数据

收稿日期:2014-09-15;修回日期:2015-03-09

基金项目:科技部创新基金资助项目(No. 14C26213311599),宁波市智团创业基金资助项目(No. 2013B70071)

Foundation Items: The Innovation Fund for Technology Based Firms of the Chinese Ministry of Science and Technology(No.14C26213311599), Ningbo Intelligence Group Entrepreneurship Project (No.2013B70071)

论文引用格式:鞠洪尧. 大数据网络服务器群智能伸缩机制与架构研究. 电信科学,2015073

Ju H Y. Research on Smart Scaling Mechanism and Structure for Big Data Network Server Groups. Telecommunications Science, 2015073

网络涉及的关键技术进行详细探讨,奠定了推进大数据产业化研究的理论基础,形成了大数据挖掘与处理的技术指导路线。

大数据挖掘与处理技术的关键在于如何有效提升查询、分析和集成效率,大幅度缩短大数据网络服务的响应时间,使终端用户在较短的时间内获得高质量的大数据分析结果。在承载大数据处理的信息网络中,服务器群是提供查询、分析和集成服务的核心部分,因此,服务器群结构优化技术、服务器群中各成员服务器的协同技术和自适应伸缩技术就成为解决大数据挖掘和处理问题的关键技术。

目前,大数据网络服务器群智能伸缩主要通过群集来实现,其存在如下缺陷:群集对多个服务器进行统一管理实现负载均衡的过程中,同一个集群中服务器的个数不能超过 32 个,且空载的服务器必须实时在线。上述缺陷造成系统重载或超载时增加成员服务器有技术困难,系统空载时,开启的服务器数量无法减少,形成能源的大量浪费。

通过对大数据网络特征和关键技术的研究,在参考文献[9~21]的基础上,集成了服务器负载监测、服务器群智能伸缩控制和负载调度等技术,给出了大数据网络服务器群智能伸缩的原理和实现方法,消除了同一个服务器群集对成员服务器数量的限制,同时使得空载的成员服务器能够及时关闭,并在系统重载或超载时自动开启,解决大数据网络服务器群中各成员服务器高效协同和智能拓展问题。

2 大数据网络服务器群智能伸缩结构及原理

2.1 构建原则及结构模型

根据大数据数量庞大、类型繁杂、处理和集成所需时间长的特点,构建大数据网络服务器群结构^[9-12]时,就必须保证服务器群具备多级并行处理能力。这种多级并行处理包括两个部分:第一部分是实现单个服务器及其所承载的数据处理软件的并行处理,这种并行处理能力对于当前的服务器硬件来说,都能够顺利实现,数据处理软件的并行处理能力也可以通过实时多副本的工作模式顺利实现;第二个部分是实现服务器群中各成员服务器间的并行协同,并建立自适应协同机制,在服务器群并行协同过程中,来自用户终端的大量并发访问是服务器群的主要负载,这些负载被即时调度到成员服务器上。当服务器群工作负载不超过临界值时,成员服务器按照实时工作负载大小被分配新的访问负载。当服务器群工作负载超过临界值时,服务器群的结构将进行自适应伸缩,以适应访问负载的需求。第二个部分是本文研究的关键。

根据上述原则,本文给出了如图 1 所示的大数据网络服务器群智能伸缩结构模型^[13-16]。

2.2 伸缩原理

服务器群^[17,18]采用多台成员服务器构成冗余结构,各成员服务器彼此互联并与负载监测器、智能伸缩控制器、调度策略控制器和负载调度器建立通信通道。服务器群中的成员服务器按指定逻辑编号排序,并保持部分成员服务

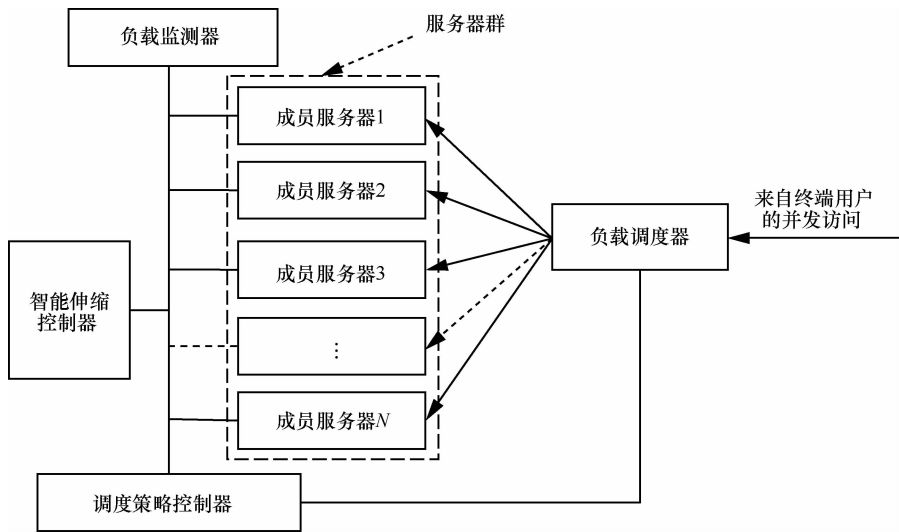


图 1 大数据网络服务器群智能伸缩结构模型



器处于开启状态,其他部分的成员服务器处于关闭状态,所有成员服务器均支持网络唤醒功能。

来自终端用户的大量并发访问通过负载调度器依次调度到服务器群中不同的成员服务器上,每个成员服务器承担的负载比例由负载调度器中与该成员服务器对应的IP地址组中IP地址的数量决定,若干个IP地址组共同构成了负载调度器的调度列表,负载调度器由应用指针和调度列表构成,终端用户的每次访问都通过应用指针对调度列表的依次查询来实现。

负载监测器实时监测成员服务器的工作负载,并对监测结果进行统计分析,统计和分析的结果实时发送给调度策略控制器和智能伸缩控制器。

调度策略控制器根据负载监测器对各成员服务器工作负载的统计分析结果,确定增减调度列表中的IP地址组,以匹配服务器群增减成员服务器的过程;或者增减某个IP地址组中IP地址的数量,以匹配服务器群的负载均衡策略。

当服务器群的总工作负载超载时,智能伸缩控制器启动服务器群智能拓展策略,向服务器群中增加新的成员服务器,同时通知调度策略控制器向调度列表中增加IP地址组,使新增的成员服务器分担访问负载。当服务器群的总工作负载空载时,智能伸缩控制器启动服务器群智能收缩策略,停止向最后一个成员服务器分配访问负载,延时等待一定时间后,从服务器群中删除最后一个成员服务器。

服务器群的智能伸缩控制过程如图2所示。

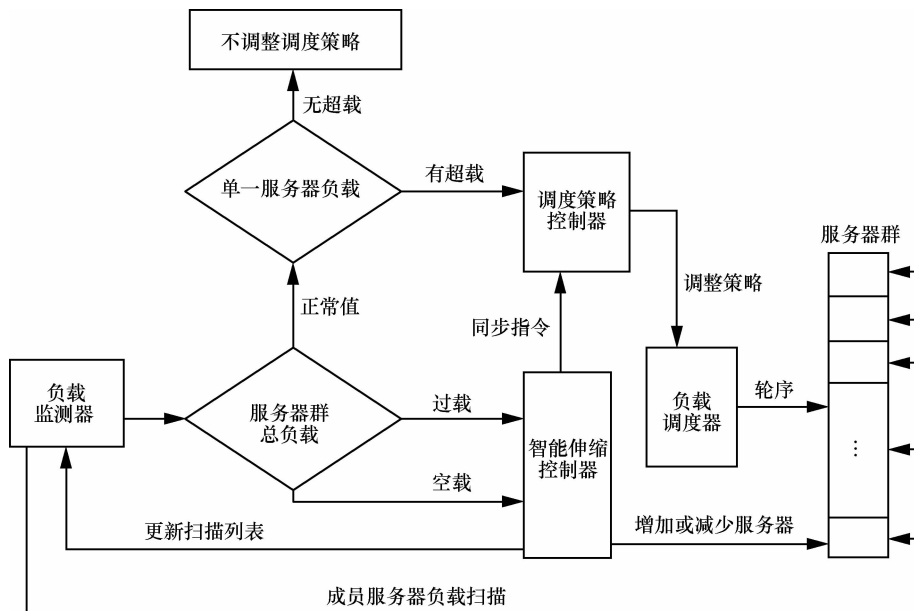


图2 服务器群智能伸缩控制原理

3 大数据网络服务器群智能伸缩关键技术

3.1 负载监测技术

负载监测^[19,20]技术集成在负载监测器中,负载监测器由扫描列表、扫描程序、分析统计程序和统计结果输出程序4个部分构成。扫描列表由服务器群中处于工作状态的成员服务器的IP地址构成,每个成员服务器使用的IP地址组中无论包含多少个IP地址,只有IP地址组中的第一个IP地址进入扫描列表。扫描程序根据扫描列表依次对成员服务器进行工作负载扫描,扫描结果由分析统计程序进行处理,结果输出程序将统计分析结果发送给调度策略控制器或智能伸缩控制器。

成员服务器的工作负载由CPU工作负载、内存工作负载和网卡流量负载构成,其中,CPU和内存工作负载由指令 `systeminfo -s IPi -u user name -p password` 获得,这里的IP_i是服务器群中第*i*个成员服务器使用的第一个IP地址。网卡工作流量由 `Netstat -e -s` 获得。

成员服务器的工作负载 $F(t)$ 和服务器群中处于工作状态的成员服务器的总工作负载 $G(t)$ 统计方法如下:

$$F_i(t) = a \times x_{\text{CPU}}^i(t) + b \times y_{\text{DRAM}}^i(t) + c \times z_{\text{NETCARD}}^i(t) \quad (1)$$

$$G(t) = \sum_{i=1}^n \frac{F_i(t)}{n} \quad (2)$$

其中, $1 \leq i \leq n$, $1 \leq n \leq N$, $F_i(t)$ 代表第*i*个成员服务器在时刻*t*的工作负载值, $G(t)$ 代表*t*时刻服务器群的工作总负载。 $x_{\text{CPU}}^i(t)$ 、 $y_{\text{DRAM}}^i(t)$ 、 $z_{\text{NETCARD}}^i(t)$ 分别代表第*i*个成员服务器

CPU、内存以及网卡接口在 t 时刻的工作负载值, a, b, c 分别代表各项的权重系数。 n 代表服务器群中处于工作状态的成员服务器的数量, N 代表服务器群中成员服务器总数。

3.2 负载调度技术

负载调度^[2]技术由调度策略控制器和负载调度器两个部分共同完成。

3.2.1 调度策略控制器

调度策略控制器的主要功能是根据负载监测器的输出结果判定服务器群工作负载是否处于正常值, 若处于正常值, 继续判定服务器群中处于工作状态的成员服务器的工作负载均衡情况, 并根据第 i 个成员服务器的工作负载 $F_i(t)$ 与该成员服务器工作负载的超载临界值 F_{imax} 和空载临界值 F_{imin} 进行比较, 决定是否向负载调度器发出负载调度策略调整指令, 具体判定关系见表 1。

调度策略控制器由备用 IP 地址列表和 IP 地址增减控制程序构成。当 $F_i(t) \geq F_{imax}$ 时, 调度策略控制程序删除服务器群中第 i 个成员服务器网卡绑定的 IP 地址组中的最后一个 IP 地址, 并同步删除负载调度器中 IP 地址列表中相应 IP 地址组的最后一个 IP 地址, 被删除的 IP 地址追加进入备用 IP 地址列表。当 $F_i(t) \leq F_{imin}$ 时, 调度策略控制程序增加服务器群中第 i 个成员服务器网卡绑定的 IP 地址, 这个 IP 地址从备用 IP 地址列表中第一个 IP 地址获得, 并同步追加负载调度器 IP 地址列表中相应 IP 地址组的 IP 地址, 同时, 将这个新增加的 IP 地址从备用 IP 地址列表中

删除。调度策略控制器工作过程如图 3 所示。

3.2.2 负载调度器

负载调度器包括应用指针和调度列表两个部分。

应用指针由点运算符和多个字符串构成, 这些字符串分别代表应用程序所在的网络区域、域和主机标识, 应用指针是终端用户访问网络应用程序时使用的域名。应用指针与调度列表中的所有 IP 地址形成映射关系, 在大量终端用户并发访问同一个网络应用时, 应用指针依次轮序访问调度列表中的 IP 地址。

调度列表由多个 IP 地址组构成。每个 IP 地址组对应服务器群中一个处于工作状态的成员服务器, 其中的 IP 地址的数量及排序与对应的成员服务器网卡绑定的 IP 地址数量及排序相同; IP 地址组中 IP 地址的数量与对应成员服务器的并行处理能力成正比; IP 地址组在调度列表中的排列顺序与服务器群中处于工作状态的成员服务器的排序相同。

当终端用户对大数据网络进行数据查询、分析等操作时, 访问请求首先通过应用域名对负载调度器进行应用地址查询, 负载调度器中的应用指针查询调度列表, 并将查询获得的应用程序 IP 地址发送给终端查询用户。终端用户使用接收到的应用 IP 地址定位运行指定应用软件工具的成员服务器, 并打开指定成员服务器中应用程序, 完成对大数据查询和分析等的处理, 并将处理结果返回给指定的终端用户。

表 1 服务器群负载调度策略调整判定关系

$F(t)$ 与阈值关系	负载监测器输出值	执行的操作情况
$F_{imax} > F_i(t) > F_{imin}$	0	服务器 i 工作负载正常, 不调整调度策略
$F_i(t) \geq F_{imax}$	1	服务器 i 超负荷工作, 通知负载调度控制器减少后续访问负载进入该成员服务器
$F_i(t) \leq F_{imin}$	2	服务器 i 负载过轻, 通知负载调度控制器增加后续访问负载进入该成员服务器

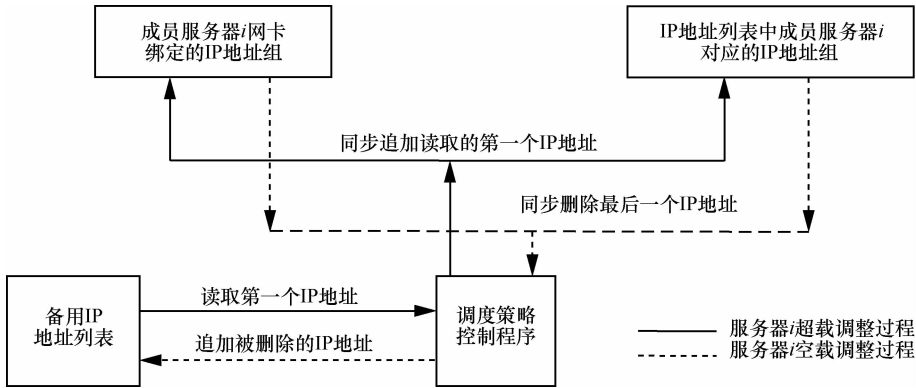


图 3 调度策略控制器工作过程



假设服务器群中含有 n 台处于工作状态的成员服务器, 第一台成员服务器网卡绑定 i 个 IP 地址, 第二台成员服务器网卡绑定 j 个 IP 地址, 第 N 台成员服务器网卡绑定 k 个 IP 地址。于是, 调度列表中含有 n 个顺序排列的 IP 地址组, 第一个 IP 地址组含有 i 个 IP 地址, 第二个 IP 地址组含有 j 个 IP 地址, $\dots$, 第 n 个 IP 地址组含有 k 个 IP 地址。在大量终端用户并发访问时, 应用指针依次顺序返回调度列表中的 IP 地址给终端用户, 在包含 $i+j+\dots+k$ 个访问负载的一个调度周期内, 服务器群的第 1 个成员服务器被分配 i 个访问负载, 第 2 个成员服务器被分配 j 个访问负载, 第 n 个成员服务器被分配 k 个访问负载。负载调度器工作过程如图 4 所示。

应用指针每次到达调度列表末端时, 会自动返回调度列表首部, 重新进行下一轮查询。

3.3 智能伸缩控制技术

智能伸缩控制由智能伸缩控制器与调度策略控制器两部分协同完成。智能伸缩控制器的主要功能是根据负载监测器输出结果将服务器群的工作总负载 $G(t)$ 与过载临界值 $G_{\max}$ 和轻载临界值 $G_{\min}$ 进行比较, 确定是否进行服务器群伸缩操作, 具体判定情况见表 2。

智能伸缩控制器由伸缩判定程序和伸缩控制程序构成。伸缩判定程序根据负载监测器的监测结果, 当服务器群总负载值达到或超过临界值时启动伸缩控制程序, 增加服务器群中成员服务器数量或关闭服务器群中工作负载最轻的成员服务器, 同步通知调度策略控制器增减负载调

度列表中 IP 地址组的数量, 同时更新负载监测器的监测列表, 并分别更新服务器群处于工作状态成员服务器和备用服务器的顺序编号。

当服务器群总负载值达到或超过超载临界值时, 即 $G(t) \geq G_{\max}$, 智能伸缩控制程序向服务器群中的第一个备用成员服务器发出开机指令, 待该备用成员服务器启动成功后, 智能伸缩控制程序读取调度策略控制器中备用 IP 地址列表中的第一个 IP 地址, 并将读取的 IP 地址及掩码写入成员服务器的参数配置文件, 同时刷新, 然后通知调度策略控制器将其备用 IP 地址列表中第一个 IP 地址追加到负载调度器的调度列表中, 并标记这个新建 IP 地址组, 同时删除备用 IP 地址列表中的第一个 IP 地址。最后, 智能伸缩控制程序将读取的 IP 地址追加到负载监测器的监测列表中, 使得负载监测器能够对新加入服务器群中的成员服务器进行工作负载实时监测。服务器群智能扩展控制过程如图 5 所示。

当服务器群总负载达到或小于空载临界值时, 即 $G(t) \leq G_{\min}$, 智能伸缩控制程序首先检测服务器群中最后一个成员服务器网卡绑定的第一个 IP 地址并记录, 将该 IP 地址从负载监测器的监测列表中删除, 同时通知负载调度策略控制器将负载调度器中对应的 IP 地址组删除, 并将删除的 IP 地址组中的全部 IP 地址追加到备用 IP 地址列表中备用。此后, 服务器群的访问负载不会被调度到最后一个成员服务器上, 待最后一个成员服务器的作业完成

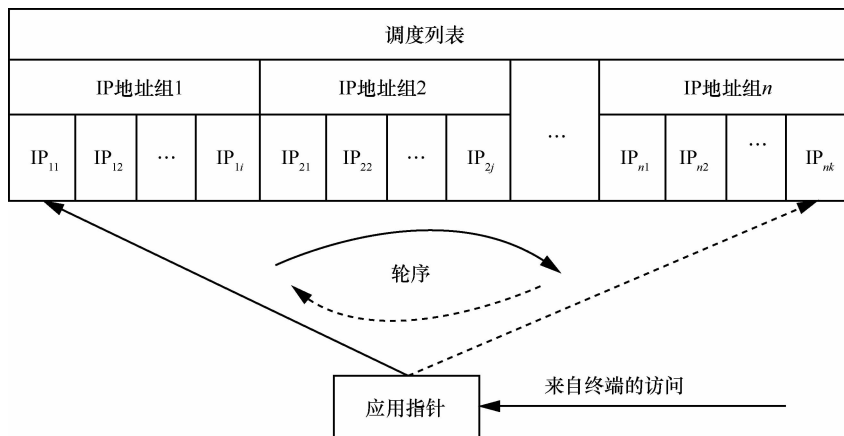


图 4 负载调度器工作过程

表 2 服务器群伸缩控制判定关系

$G(t)$ 与 阈值关系	负载监测器输出值	执行的操作情况
$G_{\max} > G(t) > G_{\min}$	0	服务器群负载正常, 进入单台服务器工作负载判定过程
$G(t) \geq G_{\max}$	1	服务器群超负荷工作, 通知伸缩管控器增加成员服务器数量
$G(t) \leq G_{\min}$	2	服务器群负载过轻, 通知伸缩管控器关闭最后一个处于工作状态的成员服务器

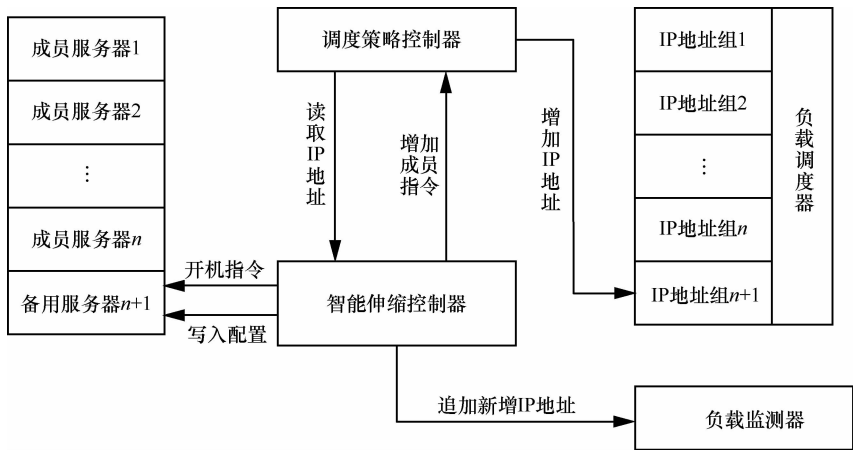


图 5 服务器群扩展控制过程

后，智能伸缩控制程序向服务器群中即将被删除的成员服务器发出清空配置参数指令，防止该成员服务器再次启动时产生 IP 地址冲突。智能伸缩控制程序发出关机指令，同时刷新服务器群中处于工作状态的成员服务器的排列序号。当服务器群中只有 1 个成员服务器处于工作状态时，则不执行服务器群收缩控制，保证服务器群至少有 1 台成员服务器为终端用户提供服务。服务器群智能收缩控制过程如图 6 所示。

4 测试环境及测试结果

4.1 测试环境

测试环境中，在服务器群中设置了 3 台成员服务器且安装相同的软件工具，操作系统均为 Windows Server 2008，负载监测器、负载策略控制器、负载调度器和伸缩控制器的操作系统均为 Windows Server 2008。测试过程中，

考虑到负载的连续和稳定性等因素，测试采用了高频度的 FTP 终端访问服务器群中的数据，得出了服务器群访问负载增加和减少两个过程中成员服务器负载、服务器群工作负载的变化数据，其中，服务器群中成员服务器工作负载过轻和过重的临界值分别设定为 $F_{\min}=0.4$ 和 $F_{\max}=2$ ，负载监测时间间隔为 2 min，访问负载减少时，去除成员服务器的关机延续时间为 120 s。

4.2 测试结果与分析

在服务器群访问负载连续增加时，测得服务器群、成员服务器的工作负载随时间的变化关系如图 7 所示。

从图 7 可以看出，服务器群工作负载不断增加的过程中，在开始阶段，服务器群工作负载随时间的变化曲线与成员服务器 1 工作负载随时间的变化曲线重合，此时，服务器群中只有成员服务器 1 在工作， $G(t)=F_1(t)$ ， $G_{\max}=F_{1\max}$ ，全部负载由成员服务器 1 承担。

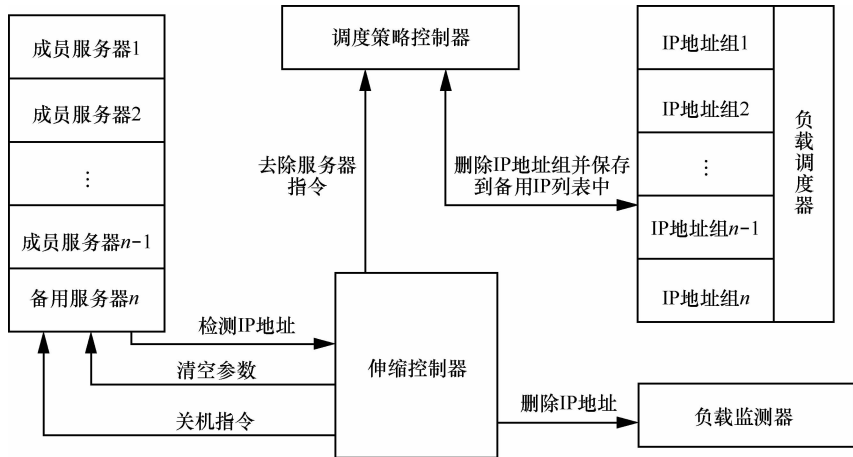


图 6 服务器群智能收缩控制过程

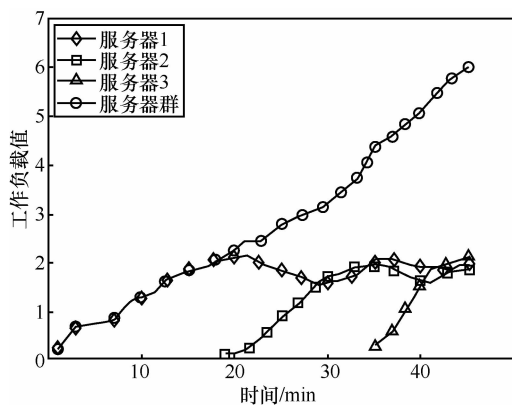


图7 访问负载增加时工作负载变化曲线

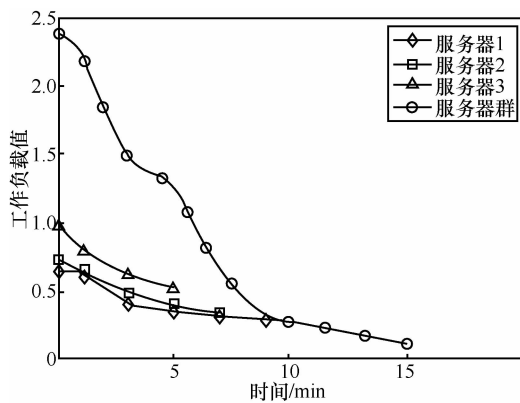


图8 访问负载减少时工作负载变化曲线

当服务器群工作负载增加到 $F_{1\max}=2$ 后,伸缩控制器报警,即 $G(t) \geq G_{\max}$,成员服务器2被检测到有负载出现,说明成员服务器2已经开始工作,且被分配了工作负载,此时服务器群总负载由成员服务器1、2共同承担,服务器群总工作负载临界值为 $G_{\max}=F_{1\max}+F_{2\max}=4$ 。随着终端用户访问量的持续增加,成员服务器1工作负载在稍有延时后开始略有下降,成员服务器2工作负载开始上升,在负载调度策略的控制下,成员服务器2快速承担更多的访问负载,成员服务器1负载下降趋于缓慢。当成员服务器1、2的工作负载曲线出现交叉后,成员服务器1的工作负载开始上升,且成员服务器1、2的工作负载曲线非常接近,表明此时服务器群进入负载均衡状态。在负载调度策略的控制下,工作负载相对较小的成员服务器1的工作负载上升加速,逐渐超过服务器2。

当成员服务器1、2的工作负载之和超过4时,伸缩控制器又开始报警,即 $G(t) \geq G_{\max}=F_{1\max}+F_{2\max}=4$,成员服务器3被检测到有负载出现,说明成员服务器3已经开始工作,此时 $G_{\max}=F_{1\max}+F_{2\max}+F_{3\max}=6$ 。随着终端用户访问负载的继续增加,负载调度策略继续发挥作用,成员服务器1、2工作负载均呈现略微下降趋势,成员服务器3的工作负载快速增加,一段时间后成员服务器3的工作负载曲线与成员服务器1、2的工作负载曲线出现交叉,然后3条曲线呈现相同的变化趋势,说明服务器群再次进入负载均衡状态。

从图7中各成员服务器的工作负载曲线的变化可以看出,后续加入服务器群中的成员服务器承担的工作负载相对前期进入服务器群的成员服务器要多一些,可见在 $G_{\min} < G(t) < G_{\max}$ 时,策略调度控制器在负载均衡过程中给工作负载相对较轻的成员服务器对应的IP地址组中加入了

更多的IP地址,同时将这些新增的IP地址绑定到工作负载相对较轻的成员服务器的网卡上。

当成员服务器1、2、3的工作负载之和达到超载临界值时,即 $G(t) \geq G_{\max}=F_{1\max}+F_{2\max}+F_{3\max}=6$,聚合控制器开始报警,此时开始进行了服务器群访问负载逐渐减少的实验测试过程,获得服务器群、成员服务器的工作负载随时间的变化关系,如图8所示。

从图8的负载变化情况看,随着服务器群外部访问量的减少或停止,成员服务器1、2、3的负载逐渐下降。当服务器群工作负载下降至 $G(t) < G_{\max}=F_{1\min}+F_{2\min}+F_{3\min}=1.2$ 时,成员服务器3开始无检测数据,说明服务器3已经从服务器群中被去除,此时服务器群工作负载由成员服务器1、2共同承担,且服务器群工作负载临界值 $G_{\min}=F_{1\min}+F_{2\min}=0.8$ 。此后,随着部分外部访问的不断结束,服务器群工作负载再次开始下降,在服务器群工作负载下降至 $G(t) < G_{\min}=0.8$ 之后,成员服务器2开始无检测数据,表明服务器2已经从服务器群中被去除,之后服务器群的工作负载曲线与成员服务器1的负载曲线重合,因为此时服务器群工作负载全部由成员服务器1承担。此后,服务器群工作负载随着外部访问的陆续结束而继续下降,由于服务器群中只有1个成员服务器在工作,智能伸缩控制器没有发出去除成员服务器1的指令。被去除的成员服务器3和成员服务器2分别在服务器群工作负载出现拐点后2min左右被成功关机。上述结果验证了服务器群在轻负载时的自动收缩能力和节约电力能源的效能。

5 结束语

目前,基于物联网、云计算技术的大数据网络的核心服务器群规模庞大,建设费用特别巨大,持续能耗惊人,这

些服务器群一般均要求成员服务器具有相同的硬件配置,且对每个服务器群中成员服务器的数量具有明确限制,所有成员服务器即使在空载情况下也必须处于开启状态,时刻准备迎接大量终端用户的并发访问。本文提出的大数据网络服务器群智能伸缩模型及实现方法,建立了一种实用性较强的服务器群应用架构,并且在大数据处理系统理论研究与应用等方面具有指导作用,其特点如下。

- 对构建核心服务器群的成员服务器无特殊要求,可以利用现有服务器构建新的服务器群,节省投资。
- 核心服务器群可以根据系统负载的变化情况自动增减成员服务器,并同步调整负载调度策略,保证处于工作状态的成员服务器间的负载均衡,为终端用户提供高效和快速的响应。
- 在系统负载比较低时,只保留一定规模的成员服务器处于工作状态,关闭服务器群中多余的成员服务器,降低服务器群的能耗。

大数据网络服务器群智能伸缩模型及实现方法虽然突破了服务器群对成员服务器性能的具体要求和数量限制,实现了智能伸缩及负载均衡,但是在成员服务器数量增加的同时,服务器群进入负载均衡状态的时间也进一步延长,今后还需要进一步研究和完善。

参考文献

- 1 王元卓,靳小龙,程学旗. 网络大数据:现状与展望. 计算机学报, 2013, 36(6): 1125~1138
Wang Y Z, Jin X L, Cheng X Q. Network big data: present and future. Chinese Journal of Computers, 2013, 36(6): 1125~1138
- 2 王珊,王会举,覃雄派等. 架构大数据:挑战、现状与展望. 计算机学报, 2011, 34(10): 1741~1752
Wang S, Wang H J, Tan X P, et al. Architecting big data: challenges, studies and forecasts. Chinese Journal of Computers, 2011, 34(10): 1741~1752
- 3 孟小峰,慈祥. 大数据管理:概念、技术与挑战. 计算机研究与发展, 2013, 50(1): 164~169
Meng X F, Ci X. Big data management: concepts, techniques and challenges. Journal of Computer Research and Development, 2013, 50(1): 164~169
- 4 吴金红,张飞,鞠秀芳. 大数据:企业竞争情报的机遇、挑战及对策研究. 情报杂志, 2013, 32(1): 5~9
Wu J H, Zhang F, Ju X F. Big data: opportunities, challenges and strategies of enterprise. Journal of Intelligence, 2013, 32(1): 5~9
- 5 王树良,丁刚毅,钟鸣. 大数据下的空间数据挖掘思考. 中国电子科学研究院学报, 2013, 8(1): 8~17
Wang S L, Ding G Y, Zhong M. On spatial data mining under big data. Journal of CAEIT, 2013, 8(1): 8~17
- 6 马建光,姜巍. 大数据的概念、特征及其应用. 国防科技, 2013, 34(2): 10~17
Ma J G, Jiang W. The concept, characteristics and application of big data. National Defense Science & Technology, 2013, 34(2): 10~17
- 7 Manyika J, Chui M, Brown B, et al. Big data: the next frontier for innovation, competition, and productivity. http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation, May 2011
- 8 Big data, big impact: new possibilities for international development. <http://www.weforum.org/reports/bigdata-big-impact-new-possibilities-international-development>, January 2012
- 9 元开元,赵卓峰,房俊等. 针对高速数据流的大规模数据实时处理方法. 计算机学报, 2012, 35(3): 476~490
Qi K Y, Zhao Z F, Fang J, et al. Real-time processing for high speed data stream over large scale data. Chinese Journal of Computers, 2012, 35(3): 476~490
- 10 陶健,张占平,郑戊午等. 机载实时容错分布式计算机系统数据网络的设计与实现. 航空计算技术, 2008, 38(6): 102~106
Tao J, Zhang Z P, Zheng W W, et al. Design and implementation of airborne real-time fault-tolerance distributed computer system's data network. Aeronautical Computing Technique, 2008, 38(6): 102~106
- 11 黄哲学,曹付元,李俊杰等. 面向大数据的海云数据系统关键技术研究. 网络新媒体技术, 2012, 1(6): 20~26
Huang Z X, Cao F Y, Li J J, et al. Developing sea cloud data system key technologies for large data analysis and mining. Journal of Congwen.net, 2012, 1(6): 20~26
- 12 时念云,王文佳,马力. 基于普适云的大数据挖掘. 计算机系统应用, 2013, 22(11): 7~12
Shi N Y, Wang W J, Ma L. Data mining of big data based on ubicloud. Computer Systems & Applications, 2013, 22(11): 7~12
- 13 徐国虎,孙凌,许芳. 基于大数据的线上线下电商用户数据挖掘研究. 中南民族大学学报(自然科学版), 2013, 32(2): 100~105
Xu G H, Sun L, Xu F. Study on big data-based mining of O2O ec users' data. Journal of South-Central University for Nationalities (Nat. Sci. Edition), 2013, 32(2): 100~105
- 14 曾斌,姚路,陈志诚. 基于负载均衡智能体的分布式 Web 代理群集研究. 计算机工程与设计, 2011, 32(8): 2631~2635



- Zeng B, Yao L, Chen Z C. Study of distributed web proxy cluster using load balancing agents. *Computer Engineering and Design*, 2011, 32(8): 2631~2635
- 15 李帅, 李盖凡, 许胤龙. 大规模多层次分布式网络系统及其仿真. *计算机仿真*, 2012, 29(9): 139~143
- Li S, Li G F, Xu Y L. Large-scale multi-layer distributed network system and its simulation. *Computer Simulation*, 2012, 29(9): 139~143
- 16 陈斌, 白晓颖, 马博等. 分布式系统可伸缩性研究综述. *计算机科学*, 2011, 38(8): 17~24
- Chen B, Bai X Y, Ma B, *et al.* Survey on software scalability of distributed system. *Computer Science*, 2011, 38(8): 17~24
- 17 尹浩, 林闯, 文浩等. 大规模流媒体应用中关键技术的研究. *计算机学报*, 2008, 31(5): 755~774
- Yin H, Lin C, Wen H, *et al.* Research on key technologies of large-scale streaming media. *Chinese Journal of Computers*, 2008, 31(5): 755~774
- 18 姚婧, 何聚厚. 基于模糊聚类分析的云计算负载平衡策略. *计算机应用*, 2012, 32(1): 213~217
- Yao J, He J H. Load balance strategy of cloud computing based on fuzzy clustering analysis. *Journal of Computer Applications*,

2012, 32(1): 213~217

- 19 黄晓斌, 钟辉新. 基于大数据的企业竞争情报系统模型构建. *情报杂志*, 2013, 32(3): 37~43
- Huang X B, Zhong H X. Build enterprise competitive intelligence system model based on big data. *Journal of Intelligence*, 2013, 32(3): 37~43
- 20 Chaiken R, Jenkins B, Larson P, *et al.* SCOPE: easy and efficient parallel processing of massive data sets. *Proceedings of the VLDB Endowment*, Auckland, New Zealand, 2008: 1265~1276
- 21 Abadi D J, Ahmad Y, Balazinska M, *et al.* The design of the borealis stream processing engine. *Proceedings of Second Biennial Conference on Innovative Data Systems Research*, Asilomar, USA, 2005: 277~289

[作者简介]



鞠洪尧, 男, 浙江纺织服装职业技术学院副教授, 主要研究方向为制造网络、云计算、智能优化等。

2015073-9

(上接 2015064-6)

- 5 Wang Y, Pandharipande A, Polo Y L, *et al.* Distributed compressive wide-band spectrum sensing. *Proceedings of the Information Theory and Applications Workshop*, San Diego, USA, 2009
- 6 Gu B, Yang Z, Hu H F, *et al.* Cooperative compressed sensing for wide-band spectrum detection. *Proceedings of the 6th International Conference on Wireless Communications Networking and Mobile Computing*, Chengdu, China, 2011
- 7 Liang J H, Liu Y, Zhang W J, *et al.* Joint compressive sensing in wideband cognitive networks. *Proceedings of the Wireless Communications and Networking Conference*, Sydney, Australia, 2010
- 8 王韦刚, 杨震, 胡海峰. 一种宽带频谱检测的空域频域压缩感知方法. *电子与信息学报*, 2013, 35(2): 255~260
- Wang W G, Yang Z, Hu H F. A method of space-frequency compressed sensing on wideband spectrum detection. *Journal of Electronics & Information Technology*, 2013, 35(2): 255~260
- 9 Tropp J A, Laska J N, Duarte M F, *et al.* Beyond Nyquist: efficient sampling of sparse band limited signals. *IEEE Transactions on Information Theory*, 2010, 56(1): 520~544

- 10 Duarte M F, Sarvotham S, Baron D, *et al.* Distributed compressed sensing of jointly sparse signals. *Proceedings of the 2005 Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, USA, 2005
- 11 程云鹏, 张凯院, 徐仲. 矩阵论. 西安: 西北工业大学出版社, 2006
- Cheng Y P, Zhang K Y, Xu Z. *Matrix Theory*. Xi'an: Northwestern Polytechnical University Press, 2006
- 12 Zeng F Z, Li C, Tian Z. Distributed compressive spectrum sensing in cooperative multihop cognitive networks. *IEEE Journal of Selected Topics in Signal Processing*, 2011, 5(1): 37~48

[作者简介]



闫盛楠, 女, 博士, 燕山大学信息科学与工程学院讲师, 主要研究方向为认知无线电、协作通信技术、压缩感知技术等。

2015064-7