

开源电商物流信息系统解决方案

云 威,葛化敏,朱新龙

(南京信息工程大学 信息与控制学院,江苏 南京 210044)

摘要:为了在物流信息系统软件构建、部署、实施层面上保持国内中小电商的轻量级特性,避免其自建IT或外包从零构建部署系统带来成本、精力的大量投入,更好地专注于商务和产品领域,不同于以往的物流信息系统实现思路,文中主要提出了将开源系统 Magento 和 OpenERP 集成的低成本物流信息系统业务软件方案,该方案包括了中小电商对 OMS、WMS、ERP、CRM 等需求的完善支持。针对业务软件的双服务器部署无法应对电商业务增长所带来的承载需求的现实问题,文中从集群角度研究了 Magento 和 OpenERP 集成的开源物流信息系统业务软件方案的集群承载构建问题,使得整体电商物流信息系统解决方案更加完备,满足电商创业到扩张的所有需求。最后,为了服务器硬件层面上保持电商的轻量级特性,大为减少其进入和退出成本,文中提出了将物流信息系统业务软件承载集群部署到 IaaS 的解决思路,并做了具体实施。整个开源电商物流信息系统解决方案技术路线先进,成本低廉,具有较强的技术前瞻性和实际应用价值。

关键词:物流信息系统;OpenERP;电子商务;负载均衡;集群;云计算

中图分类号:TP393

文献标识码:A

文章编号:1673-629X(2012)12-0233-04

Open Source E-commerce LIS Solution

YUN Wei, GE Hua-min, ZHU Xin-long

(College of Information and Control, Nanjing Information Engineering University, Nanjing 210044, China)

Abstract: In order to keep lightweight characteristics of domestic medium and small electric business in logistics information system software construction, deployment, implementation level, avoid large investment of cost, energy for the self-built or outsourcing IT from zero constructing deployment system, better focus on business and products field, different from the previous logistics information system implementation approach, mainly put forward the low cost of logistics information systems business software solutions integrated open source system Magento and OpenERP, this solution can perfectly satisfy the demand for OMS, WMS, ERP and CRM. According to the problem that business software deployment of double server could not cope with the electric business business growth, from the angle of cluster it studied the clusters bearing construction problem of open source logistics information systems business software solutions integrated OpenERP and Magento, make whole electric business logistics information system solutions more complete, meet all demand of electric business venture to the expansion. Finally, in order to keep electric business lightweight characteristics, greatly reduce the entering and exiting cost, put forward the solution that logistics information system business software bearing cluster deployed to IaaS, and practice. The solutions' technical route is advanced, the cost is low, has strong technical prospective and actual application value.

Key words: logistics information system; OpenERP; e-commerce; load balance; cluster; cloud computing

0 引言

Dale S. Rogers 等人早在 90 年代就撰文^[1]指出,“在物流领域,顶尖高手和平庸之辈的差距就在于企业物流信息技术的能力。”在互联网飞速发展的今天, Dale S. Rogers 的观点得到了更好的验证。利用基于互联网的物流信息技术缩短订单周期(Order Cycle Timer),改进供应链管理,利用信息代替库存消减物流成

本^[2]对于当前处于“市场春天”而面临市场充分竞争的电商来说,显得尤其重要。

就国内电商行业现状而言,电商平台独大(如淘宝占据网上零售份额的 70% 以上),国内平台与独立电商之间的比例严重失调,这与国外 B2C 独立电商占主流形成鲜明对比。国内平台电商的火爆说明了中小型电商缺乏有效电商商务解决方案,无论是自建 IT 构建商务站点还是选择服务外包,初期开发和后期维护成本都难以承担,而对于加强供应链管理,实施 ERP 系统,费用更是难以承受。这样,提供技术先进、低成本、高可靠、完善具体的物流信息系统解决方案对于中小型电商摆脱平台的束缚,实施独立的电子商务,加强供应链管理都具有很强的实用参考价值。

收稿日期:2012-03-29;修回日期:2012-07-02

基金项目:国家自然科学基金资助项目(10903003)

作者简介:云 威(1988-),男,硕士研究生,研究方向为系统工程、软件工程、集群与云计算;葛化敏,副教授,研究方向为系统工程、智能仪表与集成。

1 方案的双服务器部署实施

Magento 是当下流行的开源电子商务系统,采用 PHP 开发,曾获得过 sourceforge 的 2008 年最佳新项目奖,并在 2011 年 8 月 18 日被 ebay 全盘收购。Magento 在方案中主要用作 OMS。在 GNU/Linux 服务器上部署 Magento 系统,首先需要构建 LAMP 或 LNMP 环境。Nginx 的网络 IO 模型使用 Linux 2.6 内核的 epoll,较 Apache 而言,能够在支撑高连接并发的同时,对 CPU、内存的消耗较低,这里承载 Magento 系统的服务器选择部署 LNMP 环境。服务器选择的 GNU/Linux 发行版为 Ubuntu Server 11.04 LTS,都安装了 openssh-server 以便于在工作台上使用 putty 远程操作,MySQL 数据库的管理则使用 PhpMyAdmin。使用 putty 远程登录服务器后,使用 wget 命令从 Magento 官方网站下载 1.6 版本安装包,解压到站点目录并使用 chown 命令更改目录权限设置,如果需要多台服务器承载 Magento 系统组建集群,可以在工作台上下载安装包后使用 winscp 远程推送到服务器上即可。安装 Magento 之前需使用 PhpMyAdmin 新建数据库供其使用。安装过程中设置好语言、时区和货币等本地化配置,指定好数据库,并设置好加密机要信息的 Encryption Key 即可。

OpenERP 是当下非常流行的开源 ERP/CRM 系统,由比利时程序员 Fabien Pinckaers 所创建,遵循 GPL 协议发布。OpenERP 使用 Python 语言开发,所以其源码的可读性非常强,而且其灵活开放的模块架构使得其扩展变得非常方便,程序员掌握 python 和 xml 即可进行一些业务功能的扩展。目前使用 OpenERP 知名组织主要有 ENA(法国行政学校),CCI(比利时商会)等。

OpenERP6.1 于 2012 年 2 月 20 日正式发布,速度稳定性和界面相对 2011 年 8 月正式发布的 6.03 版本有很大改进,并借助 JQuery Mobile 提供 Mobile UI 以支持移动商务。OpenERP6.1 发布时将 web-client(B/S 架构中的 web 服务器端)和 openerpserver 整合在了一起,不再分开发布,但仍然提供各平台的 GTKClient(C/S 架构中的客户端)。

在 GNU/Linux 服务器上安装部署 OpenERP6.1 系统之前需要安装 PostgreSQL,并新建一个数据库供 OE 使用。同样,可以使用 Putty 远程登录服务器后使用 wget 命令从官方下载安装包,如果需要多台服务器承载 OE 系统组建集群,可以在工作台上下载安装包后使用 winscp 远程推送到各个服务器上。Openerpserver 安装相对繁琐,开始安装 openerpserver 之前也需要安装依赖的一些 Python 库,限于篇幅,这里不详加描述,读者可参考官方文档。

OpenERP 提供的模块功能主要包括销售、采购、

仓库管理、运送管理、人力资源管理、CRM/SRM、日常工作(wiki、calendars、dashboards、webmail)、MRP、项目管理、业务流程管理、适用于 IT 企业的 Scrum 和 Bug 追踪模块、EDI(Electronic Data Interchange)、BI、WMS 等。诸如 WMS、EDI 等扩展模块需要从 OpenERP 的 APP 站点下载后推送到服务器 openerpserver 安装目录下 bin/addons 中,再在 OpenERP 模块界面中安装,安装时要注意依赖关系。

Magento 系统与 OpenERP 系统之间的集成方案这里采用 openlabs 提供的开源的 OpenERPConnector(Magento 插件)与 magentoerpconnect(OE 模块)。Magento 系统安装插件需要通过 magento connect 2.0 获取插件的 key 粘贴到 magento 连接管理器中安装。Openlabs 的 OE 模块暂时都还没有推出 6.1 版本,这里暂以 6.03 版本来安装模块集成 Magento。安装 magentoerpconnect 模块需要先安装其依赖模块:

```
product_m2mcategories
base_external_reententials
base_sale_multichannels
product_images_olbs
product_links
```

配置完成后在 OpenERP 系统中会显示 Magento Connection 模块,使用 GTK Client 登录后新建一个 Magento Instance 输入 Magento 系统 URL 和在 Magento 后台配置的 API 账号即可使用。

方案使用双服务器承载这种部署方式比较简单,服务器硬件成本相对较低,整个系统能够满足微小型电商的业务承载需求。但是也不难看出,系统存在单点问题,单服务器的承载能力有限,随着业务的发展,即时单台服务器性能再高,单服务器的方案会出现性能瓶颈,满足不了业务承载需求的增长^[3]。

2 方案的集群构建

最简单的扩展方法是将数据库服务器独立部署。OpenERP 系统需要修改配置文件/etc/openerp-server.conf 指定 PostgreSQL 服务器的 IP 地址,也需要远程到 PostgreSQL 服务器上配置/etc/pgpool.conf 和 pool_hba.conf 使其接受网段内其它主机的 TCP/IP 连接。Magento 系统需要修改 magento 目录下/app/etc/local.xml 来指定 MySQL 服务器的 IP 地址,也需要远程到 MySQL 服务器上配置/etc/mysql/my.cnf 来使其接受网段内其它主机的 TCP/IP 连接。

这种部署方案虽然能降低风险,稍微提升系统性能,但是仍然存在单点问题,也无法承载大量的业务,系统架构不具备良好的可伸缩性。构建高可用、可伸缩的集群来承载系统能很好地解决这个问题,运用负

载均衡和集群技术使得电商可以根据业务量方便地调整承载电商物流信息系统业务软件方案的服务器节点数量^[4]。Magento 系统和 OpenERP 系统可以用 LVS^[5]和 Keepalived^[6] 分别构建集群承载,实现高可用(HA)和负载均衡(LB)。为了解决承载数据库服务器的单点和性能问题,可以使用 MySQL Proxy^[7](实现 RW Splitting 和 R 的负载均衡)和 MySQL Replication(实现 master-slaves 同步)构建 MySQL 集群支撑 Magento 系统集群,使用 pgpool-II^[8]来构建 PostgreSQL 集群支撑 OpenERP 系统集群。承载与支撑 Magento 系统的集群系统如图 1 所示。

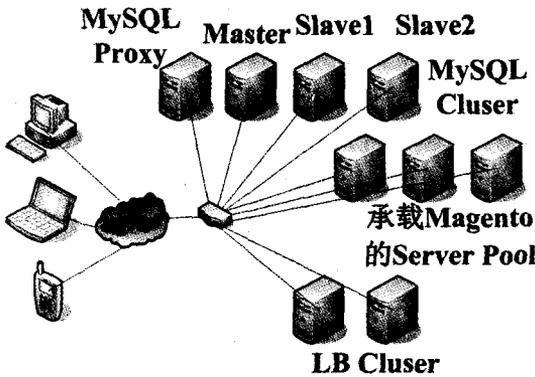


图 1 承载与支撑 Magento 系统的集群

放置 2 台 LB(load balancer)的目的是为了避免单点故障(single point of failure),防止集群裂脑(split-brain)发生。负载均衡采用基于 IP 负载均衡的 IPVS 中常用的 VS/DR 方式,IPVS 另两种方式分别是 VS/NAT 和 VS/TUN。VS/NAT 方式中,server pool 中节点对于请求的响应需要从 LB 送出,当系统规模膨胀后,调度器 LB 往往成为性能瓶颈。VS/TUN 采用 IP 隧道(IP tunneling)技术,LB 需要将客户的请求 IP 报文封装一次,源地址为分配的 server pool 中的节点,这带来了额外的开销。

IPVS 现在已经是 Linux 的标准组件,在 LB1 和 LB2 上只需要安装管理软件 ipvsadm 和开源软件包 keepalived。keepalived 主要用于 server pool 中节点的失效检查(由 Checkers 负责)和 LB1 与 LB2 之间的 failover(失效转移,通过 VRRP Stack)。在 LB1 和 LB2 上需要配置 keepalived.conf,将 HTTP_GET 换为 TCP_CHECK,并设置正确的 VIP, virtual_server 和 real_server。需要选择合适的调度算法和调度方式,这里设置 lb_algo 为 wrr,lb_kind 为 DR,并为各个 real_server 根据其实际物理性能设置合适的权重 weight。将 LB1 的 keepalived.conf 中 vrrp_instance 中 state 设置为 MASTER,而在 LB2 中将 state 设置为 BACKUP,并将 LB1 的 keepalived.conf 文件中 priority 值设置得比 LB2 中高。

支撑 Magento 系统的 db Cluster 基于 MySQL Proxy

和 MySQL Replication 技术构建。配置 MySQL 主从复制,需要远程到各 MySQL 服务器上配置 my.cnf 文件指定各服务器 id。在 Master 节点上 binlog-do-db 字段指定需要同步的数据库,并进入 MySQL monitor 来为各 Slave 节点创建账户。在各个 Slave 节点上需配置 my.cnf 指定 Master 节点的 IP、端口,获取日志时连接 Master 节点的用户名和密码。MySQL Proxy 用来实现读写分离和读 R 操作的负载均衡。MySQL Proxy 默认会监听 2 个端口,port 4040 接受外部 SQL,port 4041 为其自身的管理端口。MySQL Proxy 实现的读写分离是先判断 SQL 是 SELECT 还是非 SELECT 后再分发到后端节点,这借助 rw-splitting.lua 脚本实现,需要在该节点服务器上需要安装 Lua 环境。使用 mysql-proxy 命令启动服务时使用参数 --proxy-backend-addresses 指定 Master 节点 IP 与端口,参数 --proxy-read-only-backend-addresses 指定各 Salve 节点 IP 与端口,--proxy-lua-script 指定 rw-splitting.lua 脚本路径即可。

使用 LVS 和 keepalived 构建承载 OE 的服务器集群以及使用 pgpool-II 来构建 PostgreSQL 集群支撑 OpenERP 系统集群,如图 2 所示。

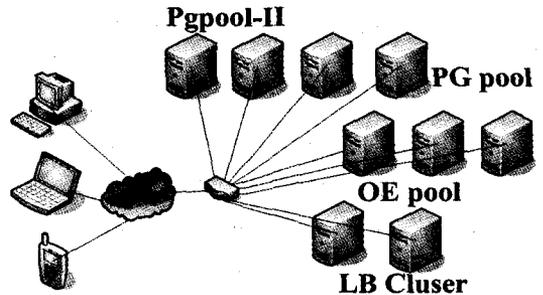


图 2 承载与支撑 OpenERP 系统的集群

与前面 Nginx 默认 80 端口不同,OpenERP 6.1 中默认 Web Client(B/S 架构中的 S 端)的默认监听端口为 8069,配置 keepalived.conf 时注意即可。PG cluster 基于 pgpool-II^[8]构建。pgpool-II 实际上是一个中间件,对外而言,它充当一个 PostgreSQL 服务端,对于 pg pool 中的节点而言,它充当客户端,构成单入口,多数数据节点的架构(SIMO)。pgpool-II 提供一个 pcp_port(默认 9898)来作为管理端口,执行 PCP 命令,pcp.conf 用来配置管理的用户名和密码的 md5 哈希值(可以由 pg_md5 命令生成)。要让 pgpool-II 节点接受其它主机的 TCP/IP 连接需要配置 pgpool.conf 与 pool_hba.conf。pgpool.conf 中通过 backend_hostnamex、backend_portx、backend_weightx 三个字段指定 PG pool 中的节点的 IP、端口和负载均衡权值,末尾的 x 是 pool 中节点的序号(0,1,2,...)。设置 replication_mode = true 和 load_balance_mode = true 分别用来开启 replication 和 LB(主要是读 R 的负载均衡)功能。前者告知 pgpool-II

将接收到的查询发送一份拷贝到所有的数据库节点,后者对 R 操作根据各 PG 节点的权重进行负载均衡。

图 1 和图 2 所构建的集群本身具有很好的可伸缩性,但是为了承载波动性的需求在硬件层面扩展集群会造成硬件资源的浪费(波谷需求时),更极端的例子是对集群承载能力的需求存在尖峰性需求(Lumpy)或者虽然是规律性需求(Regular)但是存在很大波动,硬件层面扩展集群存在利用率和成本问题。而借助 IaaS 来部署方案可以很好地避免这个问题。

3 降低方案成本:将方案集群部署到云中

浙江大学电子商务与信息安全重点实验室吴吉义等人在 2009 年恰当地^[9]将云计算理解为由虚拟化(virtualization)、效用计算(Utility Computing)、IaaS、PaaS、SaaS 等概念混合演进并跃升的产物,并认为云计算是分布式计算(Distributed Computing)、并行计算(Parallel Computing)继续发展的商业实现。NIST 定义了云计算主要的三种服务模式^[10]: SaaS、PaaS、IaaS。云计算有能力将传统的零碎的 B/S 架构迁入云中,形成 B/C(Cloud)架构,因为对于企业来说,这具有明显的成本优势。借助云计算,对于服务器硬件资源的获取不再需要一次性投入巨大成本,客户在使用云计算时只需要按使用行为进行付费,几乎没有退出成本。将承载物流信息系统的集群部署到 IaaS 中,能够在服务器硬件层面上保持电商的轻量级特性,大为减少其进入和退出成本。这样,中小型电商就能够将精力专注到商务和市场上去,不再支付原先高额的服务器运行管理费用(电力,空间,空调设备,人事费用等),这也更加符合电子商务的本质。

国外知名 IaaS 主要是 Amazon 的 AWS,服务主要包括^[11]弹性云计算 EC2、简单存储服务 S3、Simple DB 等。国内 IaaS 刚起步,处于快速发展中。知名的有阿里云(目前仍需按月付费)、百度开放云平台(公测中)、CloudEx Inc.(云快线)、盛大云等。相比较而言,盛大云对按需付费支撑的粒度更细(小时),价格也比较具有诱惑力。盛大云采用的是 Amazon 的 AWS 模式,提供弹性扩展的云硬盘(弹性块存储 EBS)并按容量付费。用户申请的云主机(C2 实例)是在一个局域网段内,可以 ping 通,能够实现方案中的四层负载均衡。盛大云中,C2 实例和与公网 IP 捆绑的网络带宽是分离的,可以将网络带宽绑定到任意申请的 C2 实例上。针对方案集群而言,LVS-DR 集群的负载均衡需要配置一个网络带宽,服务器节点需要分别配置网络带宽。db cluster 中各节点使用内网线路即可。对于绑定网络带宽的 C2 实例可以使用 Putty 远程。对于内网线路的 C2 实例,可以由盛大云登录后网页的“控制

台”进入。

C2 实例正常关机数据是不会丢失的,但一旦删除 C2 实例,资源将被回收,运营数据需要申请云硬盘(EBS)挂载到 C2 实例上使用。以 MySQL 节点为例,将挂载到 C2 实例 EBS 卷使用 fdisk 命令分区、使用 mount 命令挂载后,配置 my.cnf 中 datadir 字段指向挂载分区即可。需要注意的采用 AWS 模式的盛大云也采用安全组的概念,新建的 C2 实例只默认开放 22 端口(ssh)。对于方案集群而言,需要根据节点开放端口的类别规划和创建若干安全组,比如 mysql 服务器安全组需要开放 3306 端口、OE6.1 应用服务器节点需要开放 8069 端口(web-client)和 8070 端口(供客户 GT-Kclient 连接使用)、pg 节点需要开放 5432 端口、mysql-proxy 节点需要开放 4040 和 4041 端口、pgpool-II 节点需要开放 5433 端口和 9898 端口(pcp_port)。方案中所有节点都开启 PING(ICMP Echo),以便进行 ICMP 数据包探测。

在为了谨慎起见,将数据部署在云中,需要充分考虑到信息安全,采用强加密措施,著名的信息安全国际会议 RSA2010 将云计算安全列为焦点问题^[12]。另外,作为容灾的一项重要措施,将方案部署到 IaaS 时要做好异地备份,以免 IaaS 的云服务中断造成损失。最后,要做好在云之间进行迁移的准备。因为 IaaS 服务商的收费策略和服务质量都可能变动,作为客户而言,总是青睐于性价比高的 IaaS 服务。国内外 IaaS 发展十分迅速,竞争激烈,服务品质不断推陈出新,客户需要在 IaaS 之间进行迁移做好事先的考虑和准备,敏捷的应对变化。

4 结束语

方案能够在技术层面上保持国内中小电商的轻量级特性。一方面方案使得中小电商能够避免自建 IT 或外包从零构建物流信息系统带来的资本和精力的大量投入,另一方面将方案部署到 IaaS 中能避免在服务器硬件上的大笔投入,大为减少电商行业的进入和退出成本。

就目前国内电商行业情况而言,使得中小电商进入和退出成本高的另一个重要的环节是仓储物流。国内客户对用户体验(核心是订单周期)的高要求带给电商的成本压力是巨大的。国内大型电商大多采用地方仓分区域发货的方式缩短物流距离,缩短订单周期并降低配送成本,但这种方式成本投入巨大,电商要投入巨大金钱和精力进行管理维持运转,这造成电商由轻量型企业变得庞大臃肿,而且退出成本非常高。中小电商实力有限,往往采用统仓的方式发货,商品配

(下转第 240 页)

所示:

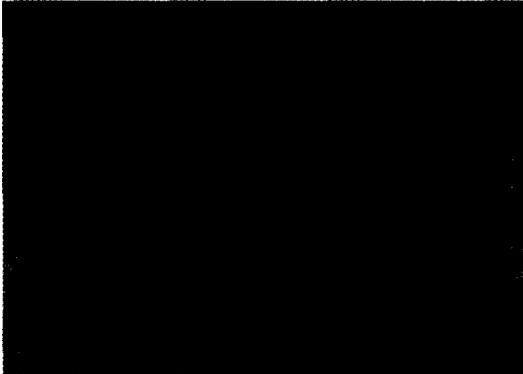


图 6 输出三维模型

经过与测试图纸比对,表明本方案可大幅加快建模速度,构件识别率高,输出模型准确。

4 结束语

人工构建建筑物三维模型费时费力。由建筑平面图自动生成三维模型则可以解决这一问题。基于建筑平面图 DXF 数据,文中完成了文件读取、构件识别与三维重建等工作。该方案首先识别各类建筑构件,将它们统一归为闭合轮廓,并由闭合轮廓获取拓扑信息。然后对闭合轮廓进行自底向上的拉伸操作得到三维模型数据,并导出 WRL 文件。实验结果表明该方案能加快建模速度,输出模型准确实用。

参考文献:

[1] Ah-Soon C, Tombre K. Architectural Symbol Recognition U-

sing a Network of Constraints[J]. Recognition Letters, 2001, 22(2):231-248.

[2] Tombre K, Ah-Soon C, Dosch P, et al. Robust and Off-the-Shelf Methods for Graphics Recognition[J]. ICPR, 1998(1): 406-408.

[3] Lewist R, Sequin C. Generation of 3D building models from 2D architectural plans [J]. Computer - Aided Design, 1998, 30(10):765-799.

[4] MIT Building Model Generation Project [EB/OL]. 2002. <http://city.csail.mit.edu>.

[5] Kashlev D. Efficient 3D Building Model Generation from 2D Floor Plans [D]. Massachusetts: Massachusetts Institute of Technology, 2008.

[6] Horna S, Meneveaux D, Damiand G, et al. Consistency constraints and 3D building reconstruction [J]. Computer-Aided Design, 2009, 41(1):13-27.

[7] 杨华飞, 杨若瑜, 路通, 等. 建筑工程图的解释识别方法[J]. 计算机应用, 2007, 27(9):2242-2248.

[8] 路通, 杨若瑜, 杨华飞, 等. 三维结构构件渐进式整合与重组方法[J]. 计算机辅助设计与图形学学报, 2007, 19(4):491-495.

[9] 胡筋, 杨若瑜, 曹阳, 等. 基于图形理解的建筑结构三维重建技术[J]. 软件学报, 2002, 13(9):1873-1880.

[10] 张赐, 吴健平. 基于 AE 的校园房产三维 GIS 系统开发与研究[J]. 计算机技术与发展, 2011, 21(2):215-218.

[11] 袁庆萍. 建筑图三维重建技术研究[D]. 南京:南京工业大学, 2005.

[12] 曹家. 建筑图的三维重建系统[D]. 哈尔滨:工业大学, 2010.

(上接第 236 页)

送距离越远,在配送途中被 3PL 转包的几率和次数越大,配送的及时性无法保证,而且物流成本高,且普遍缺失与 3PL 的议价权。对于如何解决这方面的问题从而在仓储物流角度保持中小电商的轻量型特质,大幅减少其进入和退出行业成本,文中受限于选题和精力没有涉及,但这方面意义重大,是未来的主要工作方向。

参考文献:

[1] Richard D S, Guerra P. Information Technology: Logistics Innovations for the 1990s [C]//Annual Conference Proceedings. [s.l.]:[s.n.], 1991:247-248.

[2] Ballou R H. 企业物流管理:供应链的规划、组织和控制[M]. 北京:机械工业出版社, 2006.

[3] Azar Y, Feder M, Lubetzky E. The Multicast Bandwidth Advantage in Serving a Web Site[J]. Computer Science, 2003, 2233:180-188.

[4] 喻莉, 阮文涛. 负载均衡技术的研究与实现[J]. 计算机技术与发展, 2007, 17(8):120-122.

[5] 章文嵩. 可伸缩网络服务的研究与实现[D]. 长沙:国防科学技术大学, 2000.

[6] Alexandre Cassen [EB/OL]. 2012. <http://www.keepalived.org>.

[7] Oracle Corporation [EB/OL]. 2012. <http://dev.mysql.com/downloads/mysql-proxy>.

[8] PgFoundry [EB/OL]. 2012. <http://www.pgpool.net>.

[9] 吴吉义, 平玲娣, 潘雪增, 等. 云计算:从概念到平台[J]. 电信科学, 2009(12):23-29.

[10] De S K, Krishna P R. Clustering web transactions using rough approximation[J]. Fuzzy Sets and Systems, 2004(3):131-138.

[11] 张建成, 宋丽华, 鹿全礼, 等. 云计算方案分析研究[J]. 计算机技术与发展, 2012, 22(1):165-167.

[12] 冯登国, 张敏, 张妍, 等. 云计算安全研究[J]. 软件学报, 2011(1):71-83.